UNIVERSITE GASTON BERGER DE SAINT LOUIS

Ecole **D**octorale des **S**ciences et des **T**echnologies
UFR de **S**ciences **A**ppliquées et de **T**echnologie

THESE DE DOCTORAT UNIQUE

Mention : Informatique
Spécialité : Réseaux Complexes

PROPOSITION D'UNE NOUVELLE STRUCTURE MÉSOSCOPIQUE ET DE SES APPLICATIONS DANS UN RÉSEAU COMPLEXE

Présentée et soutenue publiquement le 10 Décembre 2022 par

Issa Moussa DIOP

Directeur de Thèse **: Cherif DIALLO**

Composition du Jury

M. Moussa LO	Professeur Titulaire, Université Virtuelle du Sénégal	Président
M. Ibrahima NIANG	Professeur Titulaire, Université Cheikh Anta Diop de Dakar, Sénégal	Rapporteur
M. Cheikh SARR	Professeur Titulaire, Université Iba Der Thiam de Thiès, Sénégal	Rapporteur
M. Youssou FAYE	Maître de Conférences, Université Assane Seck de Ziguinchor, Sénégal	Rapporteur
M. Hocine CHERIFI	Professeur, Université Franche-Comté, France	Examinateur
M. Alassane DIOP	Maître de Conférences, Université Virtuelle du Sénégal	Examinateur
Mme. Chantal BONNER-CHERIFI	Maître de Conférences, Université Lyon 2, France	Invitée
M. Cherif DIALLO	Maître de Conférences, Université Gaston Berger de Saint-Louis, Sénégal	Directeur

Thèse Numéro **EDST 017-2022**

Préparée au sein du Laboratoire d'Algèbre, Cryptographie, Codes et Applications (LACCA)

UNIVERSITE GASTON BERGER

L'excellence au service du développement

THÈSE DE DOCTORAT

DE L'UNIVERSITE GASTON BERGER

Préparée à l'Université Gaston Berger
Laboratoire d'Algèbre, Cryptographie, Codes et Applications (LACCA), UFR SAT

Proposal of a new mesoscopic structure and its applications in a complex network

Soutenue par

Issa Moussa DIOP

Le 10 Décembre 2022

École doctorale n°xxx

Ecole Doctorale Sciences et Technologies

Spécialité

Informatique

Préparée au

Laboratoire LACCA, UFR SAT, UGB

Composition du jury :

Pr. Moussa LO, Professeur Titulaire Université Virtuelle du Sénégal	*Président du jury* *Président*
Pr. Ibrahima NIANG, Professeur Titulaire Université Cheikh Anta Diop, Dakar, Sénégal	*Rapporteur*
Pr. Cheikh SARR, Professeur Titulaire Université Iba Der Thiam, Thiès, Sénégal	*Rapporteur*
Dr. Youssou FAYE, Maître de Conférences université Assane Seck, Ziguinchor, Sénégal	*Rapporteur*
Pr. Hocine CHERIFI, Professeur Université Franche-Comté, Dijon, France	*Examinateur*
Dr. Alassane DIOP, Maître de Conférences Université Virtuelle du Sénégal	*Examinateur*
Dr. Chantal BONNER-CHERIFI, Maître de Conférences Université Lyon 2, France	*Invitée*
Dr. Cherif DIALLO, Maître de Conférences Université Gaston Berger, Saint-Louis, Sénégal	*Directeur de thèse*

UNIVERSITE GASTON BERGER

L'excellence au service du développement

Remerciements

Mes premiers remerciement s'adressent à mon Directeur de Thèse, le Professeur Cherif Diallo par qui tout a commencé. Je vous remercie vivement pour m'avoir donné l'opportunité de vivre cette expérience. Votre encadrement et vos conseils m'ont permis d'atteindre mes objectifs. Merci vraiment pour votre soutien et votre engagement qui m'ont été d'une grande aide.

Je vous remercie vous tous pour votre patience, votre rigueur. Je n'oublierai jamais nos réunions, ces moment d'échanges fructueux, où vous corrigez mes erreurs, enlevez mes doutes, répondez à mes questions, et donnez de nouvelles directions à mes recherches. J'ai beaucoup appris avec vous. Je ne vous remercierai jamais assez. Je vous dois beaucoup. Encore une fois, merci beaucoup pour tout.

Avec les mêmes mots, je remercie également, autant, très sincèrement et très chaleureusement les Professeurs Hocine Chérifi de l'Université Franche-Comté et Chantal Bonner-Chérifi de l'Université Lyon 2.

Je remercie très particulièrement le Professeur Moussa Lo de m'avoir fait l'honneur de présider mon jury de thèse, mais aussi et surtout d'avoir participé à ma formation.

Je remercie également le Professeur Ibrahima Niang de l'Université Cheikh Anta Diop de Dakar, le Professeur Cheikh Sarr de l'Université Iba Der Thiam de Thiès et le Professeur Youssou Faye de l'Université Assane Seck de Ziguinchor pour avoir accepté d'être les rapporteurs de ce travail.

Mes remerciements vont également à l'endroit du Professeur Alassane Diop de l'Université Virtuelle du Sénégal pour avoir accepté d'évaluer ce travail en tant que membre du jury.

Je remercie le Centre d'Excellence Africain en Mathématiques, Informatique et TIC (CEAMITIC) de m'avoir donné un bon cadre de travail.

Je remercie tous les Professeurs de l'UFR SAT, particulièrement les Professeurs du département Informatique. Vous avez joué un rôle important dans ma formation.

Merci à mes collègues doctorants qui sont au CEAMITIC. Je vous remercie tous, autant que vous êtes.

A mes amis, je vous remercie tous, j'espère que vous vous reconnaîtrez, je ne vais pas citer de nom.

A toutes les personnes qui m'ont apporté leur soutien de près ou de loin, je vous adresse mes remerciement les plus chaleureux.

A mes frères et sœurs, je vous remercie pour vos encouragements, votre assistance et votre soutien. Enfin, mention spécial à mes parents qui n'ont ménagé aucun effort pour me mettre dans de très bonnes conditions. Vous êtes les premiers à m'avoir encouragé à faire la thèse et je sais que vous n'avez jamais cessé de prier pour ma réussite.

Résumé

Les réseaux complexes constituent un domaine scientifique qui permet de modéliser des systèmes complexes en se basant sur les graphes qui représentent le mieux les interactions de différents objets, personnes, sites, entités, etc. L'étude des réseaux complexes permet ainsi de mieux comprendre plusieurs phénomènes dans différents systèmes complexes tels que les systèmes financiers, la propagation d'épidémie, les réseaux informatiques, les réseaux de transports, etc. L'analyse d'un réseau complexe peut se faire sur trois niveaux: le niveau macroscopique, le niveau microscopique et le niveau mésoscopique. Au niveau macroscopique, différentes mesures globales sont utilisées pour caractériser le réseau. Au niveau microscopique, on s'intéresse à la propriété d'un nœud ou d'un lien. Au niveau mésoscopique, on met l'accent sur l'émergence de groupes de nœuds qui ont des similarités. Les deux structures mésoscopiques les plus étudiées sont la structure communautaire et la structure noyau-périphérie. En effet, elles sont présentes dans beaucoup de réseaux réels et trouvent plusieurs applications. Cependant, aucun des deux ne permet de dissocier les interactions dans les groupes et les interactions entre les groupes. Ainsi, dans cette thèse, nous proposons une nouvelle structure mésoscopique appelée la structure en composante qui permet de démêler les interactions locales et globales des structures communautaires et noyau-périphérie. Ainsi, un réseau peut être vu comme un agrégat de sous-réseaux qu'on peut analyser séparément. Extensible également aux réseaux pondérés, cette nouvelle structure mésoscopique offre de nombreux champs d'application telles que la recherche de noeuds critiques, l'extraction de k-core, la robustesse, etc. Ainsi, nous l'avons appliqué au réseau mondial de transport aérien dont on a aussi étudié la résilience en proposant de nouvelles stratégies d'attaques plus efficaces comparées aux attaques classiques.

Mots clés : Réseaux complexes, Structure mésoscopique, Structure communautaire, Structure noyau-périphéreie, Structure en composante, Mesure de centralité, k-core, s-core, Robustesse

Abstract

Complex networks are a scientific field that makes it possible to model complex systems based on the graphs that best represent the interactions of different objects, people, sites, entities, etc. the study of complex networks thus makes it possible to better understand several phenomena in different complex systems such as financial systems, the spread of epidemics, computer networks, transport networks, etc. The analysis of a complex network can be done on three levels: the macroscopic level, the microscopic level and the mesoscopic level. At the macroscopic level, different global measurements are used to characterize the network. At the microscopic level, we are interested in the property of a node or a link. At the mesoscopic level, emphasis is placed on the emergence of groups of nodes that have similarities. The two most studied mesoscopic structures are the community structure and the core-periphery structure. Indeed, they are present in many real networks and find several applications. However, neither of the two makes it possible to dissociate the interactions in the groups and the interactions between the groups. Thus, in this thesis, we propose a new mesoscopic structure called the component structure which allows to disentangle the local and global interactions of community and core-periphery structures. Thus, a network can be seen as an aggregate of sub-networks that can be analyzed separately. Extendable also to weighted networks, this new mesoscopic structure offers many fields of application such as the search for critical nodes, k-core extraction, robustness, etc. Thus, we have applied it to the global air transport network, whose resilience has also been studied by proposing new attack strategies that are more effective compared to conventional attacks.

Keywords : Complex networks, Mesoscopic structure, Community structure, Core-periphery structure, Component structure, Centrality measure, k-core, s-core, Robustness

Table des matières

Remerciements i

Résumé ii

Abstract iii

Liste of figures ix

Liste of tables xv

1 Introduction Générale 1
1 Les réseaux réels . . . 2
1.1 Les réseaux technologiques . . . 2
1.2 Les réseaux sociaux . . . 3
1.3 Les réseaux d'information . . . 3
1.4 Les réseaux biologiques . . . 3
2 Les graphes . . . 4
2.1 Définition . . . 4
2.2 Matrice d'adjacence . . . 5
2.3 Degré d'un noeud . . . 5
2.4 Chaine . . . 5
2.5 Chemin . . . 6
2.6 Composantes . . . 6
2.7 Distance . . . 6
3 Niveau macroscopique . . . 6
3.1 Densité . . . 6
3.2 Diamètre . . . 7
3.3 La moyenne des plus court chemin . . . 7
3.4 Coefficient de clustering . . . 7
3.4.1 Coefficient de clustering local . . . 7
3.4.2 Coefficient de clustering global . . . 7
3.5 Assortativité . . . 8
3.6 Distribution des degrés . . . 8
3.6.1 Le test de K-S . . . 8
3.7 Les modèles de réseaux complexes . . . 8
3.8 Le modèle aléatoire de Erdös-Rényi . . . 9
3.9 Le modèle petit-monde de Watts et Strogatz . . . 10
3.10 Le modèle sans-échelle de Barabasi-Albert . . . 10

4 Niveau mesoscopique . . . 11
4.1 La structure communautaire . . . 11
4.1.1 Algorithmes de détection de communautés . . . 12
4.1.2 Évaluation . . . 13
4.2 La structure noyau-périphérie . . . 14
4.3 La structure hiérarchique . . . 14
4.4 Le k-core . . . 15
5 Niveau microscopique . . . 15
5.1 Centralités . . . 15
5.1.1 Centralité de degré . . . 15
5.1.2 Centralité de force pondérée . . . 15
5.1.3 Centralité d'intermédiaire . . . 16
5.1.4 Centralité d'interdépendance pondérée . . . 16
5.1.5 Centralité du PageRank . . . 16
5.1.6 Centralité de proximité . . . 16
5.1.7 Centralité de vecteur propre . . . 17
6 La robustesse . . . 17
6.1 Les attaques . . . 17
6.2 Évaluations . . . 17
7 Contexte, contributions et organisation de la thèse . . . 18
7.1 Contexte . . . 18
7.2 Contributions . . . 18
7.3 Organisation de la thèse . . . 19

2 Context, contributions and organization of the manuscript 21
1 Thesis context . . . 22
2 Main contributions . . . 22
3 Organization of the manuscript . . . 23

3 Revealing the component structure of the world air transportation network 25
1 Introduction . . . 26
2 Litterature review . . . 27
3 Component structure of a network . . . 31
3.1 Definition . . . 31
3.2 Component structure detection algorithm . . . 31
4 Uncovering the dense parts of the world air transportation network . . . 32
4.1 Data . . . 32
4.2 Community detection . . . 33
4.3 Community structure analysis . . . 33
5 Local component structure of the world air transportation network . . . 36
5.1 Analysis of the large local components . . . 37
5.1.1 Basic macroscopic topological properties . . . 37
5.1.2 Degree Distribution . . . 40
5.1.3 Degree-degree distance distribution of the local components . . . 41
5.1.4 Distribution of airports by country . . . 43
5.2 Analysis of the small local components . . . 44
5.2.1 Basic macroscopic topological properties . . . 46
6 Global component structure of the world air transportation network . . . 49
6.1 Analysis of the large global component . . . 49
6.1.1 Basic macroscopic topological properties . . . 49

6.1.2 Degree and degree-degree distance distribution . . . 50
6.1.3 Distribution of the airports between regions . . . 50
7 Comparison of the world air transportation network with the large components . 51
7.1 Basic macroscopic topological properties . . . 51
7.2 Degree and degree-degree distance distribution . . . 52
8 Degree Centrality Analysis . . . 53
8.1 Exploring the regional hubs in the large local components . . . 53
8.2 Exploring the inter-regional hubs in the large global component . . . 57
8.3 Comparison of the regional and inter-regional hubs with the hubs of the world air transportation network . . . 59
9 Core Analysis . . . 62
9.1 Exploring the max k-core in the large local components . . . 62
9.2 Exploring the max k-core in the large global components . . . 64
9.3 Comparison of max k-core of the components with the max k-core of the world air transportation network . . . 66
10 Conclusion . . . 66
11 Appendix . . . 68

4 Exploring the Regional Structure of the Worldwide Air Traffic and Route Networks 82
1 Introduction . . . 83
2 Literature review . . . 84
3 Data and Tools . . . 85
3.1 Data . . . 85
3.2 Tools . . . 85
3.2.1 Extracting the component structure . . . 86
3.2.2 Extracting the communities . . . 86
3.2.3 Quality metrics . . . 86
4 Community structure analysis . . . 88
4.1 Comparing the weighted community structure uncovered by Louvain and Combo . . . 88
4.2 Comparing weighted and unweighted community structures uncovered by Louvain . . . 89
5 Component structure . . . 90
5.1 Local components . . . 91
5.1.1 Large local components . . . 91
5.1.2 Small local components . . . 92
5.2 Global components . . . 93
6 Topological properties of the large components . . . 94
6.1 Clustering coefficient . . . 94
6.1.1 Large Local components . . . 94
6.1.2 Large Global component . . . 97
6.2 Degree-degree correlation . . . 97
6.2.1 Large Local components . . . 98
6.2.2 Large Global component . . . 98
6.3 Strength Distribution . . . 98
6.3.1 Large Local components . . . 98
6.3.2 Large Global component . . . 100
7 Topological properties of the world air transportation . . . 101

7.1 Clustering coefficient . . . 101
7.2 Degree-degree correlation . . . 101
7.3 Strength distribution . . . 101
8 Centrality Analysis . . . 103
8.1 Top five Strength Analysis . . . 103
8.1.1 Regional Analysis . . . 103
8.1.2 Inter-Regional Analysis . . . 105
8.2 Comparison with the World Transportation network . . . 106
8.3 RBO Analysis . . . 107
9 S-core Analysis . . . 113
9.1 Regional Analysis . . . 113
9.2 Interregional Analysis . . . 117
9.3 Comparison with the world air transportation network . . . 120
10 Discussion and Conclusion . . . 120

5 Targeted attacks on the world air transportation network : impact on its regional structure 123
1 Introduction . . . 124
2 Background . . . 125
2.1 Component structure . . . 125
2.2 Targeted Attack . . . 126
2.3 Evaluation measures . . . 126
3 Data and Methods . . . 126
3.1 Data . . . 126
3.2 Methods . . . 126
4 Experimental results . . . 127
4.1 Component structure . . . 127
4.2 Degree targeted attack . . . 127
4.3 Betweenness targeted attack . . . 130
5 Discussion and Conclusion . . . 132

6 Robustness of the Weighted Air Transport Network 133
1 Introduction . . . 134
2 Background . . . 135
2.1 Component structure . . . 135
2.2 Targeted Attack . . . 136
2.2.1 Weighted Strength Centrality . . . 137
2.2.2 Weighted Betweenness Centrality . . . 137
2.2.3 Weighted PageRank centrality . . . 137
2.3 Evaluation Measures . . . 137
2.3.1 Largest Connected Component . . . 137
2.3.2 Weighted Efficiency . . . 137
2.4 Jaccard Index . . . 138
3 Data and Methods . . . 138
3.1 Data . . . 138
3.2 Methods . . . 139
4 Component structure . . . 140
5 Strength targeted attack . . . 140
5.1 Evolution of the Largest Connected Component . . . 140
5.2 Weighted Efficiency . . . 142

6 Weighted Betweenness targeted attack . . . 143
6.1 Evolution of the Largest Connected Component . . . 143
6.2 Weighted Efficiency . . . 146
7 Weighted PageRank Targeted Attack . . . 147
7.1 Evolution of the Largest connected component . . . 147
7.2 Weighted Efficiency . . . 148
8 Comparing the attacks . . . 149
8.1 Evolution of the Largest Connected Component . . . 149
8.2 Weighted Efficiency . . . 154
9 Conclusion . . . 154

7 Targeted Attack of the Air Transportation Network Global Component 156
1 Introduction . . . 157
2 Data and methods . . . 157
2.1 Data . . . 157
2.2 Component Structure . . . 157
2.3 Targeted Attack on the Largest Global Component . . . 158
2.4 Performance Measure . . . 159
3 Experimental results . . . 159
3.1 Component structure . . . 159
3.2 Degree attacks . . . 159
3.3 Betweenness attacks . . . 161
4 Discussion and Conclusion . . . 164

Conclusion and future works 166

List of publications 168

Bibliography 169

Liste of figures

3.1 The component structure detection algorithm. First, the community structure of the network is uncovered. Second, the local components are obtained by eliminating inter-community links from the community structure.Third, the global components are obtained by eliminating intra-community links and subsequent isolated nodes from the community structure. 32

3.2 The community detection algorithms uncover similar communities. (A) Louvain reveals 3 communities in North and Central America-Caribbean (airports in red), East-Southeast Asia (airport in black), and in Russia-Central Asia-Transcaucasia (airports in blue). (B) LPA finds similar communities in these areas. 35

3.3 Communities uncovered by Louvain are grouped in a single community by LPA. (A) Louvain reveals two communities in Europe (airports in brown) and Africa-Middle East-Southern Asia (airports in lavender). (B) LPA finds a unique community covering the same geographical area. 35

3.4 Large communities uncovered by Louvain split into multiple communities by LPA. (A) Louvain reveals 3 communities. The indigo points represent the airports located in Africa-Middle East-Southern Asia. The orange points represent the airports located in Oceania. The cyan points represent the airports in South America. (B) LPA uncovers 57 communities in the same geographical areas. There are 16 communities in Africa-Middle East-Southern Asia, 28 in Oceania, and 13 in South America. 36

3.5 The seven large local components. 1) Airports in the North and Central America-Caribbean are in red. 2) Airports in the European component are in black. 3) Airports in the East and Southeast Asia component are in blue.4) Airports in the Africa-Middle East and India component are in green. 5) Airports in the Oceania component are in orange. 6) Airports in the South America component are in brown. 7) Airports in the Russia-Central-Asia-Transcaucasia component are in indigo. 37

3.6 Basic topological properties of the large components. Each bar plot concerns a different property (Diameter, Average shortest path, Density, Transitivity, Assortativity, and Hub dominance). The blue bar corresponds to the large local component. The red bar corresponds to the large global component. For assortativity, absolute values are plotted, but in reality, the values are negative. 38

3.7 Degree distributions of the large local components. Empirical distribution (**dot**), estimates (**line**).The distributions under test are : power law (PL), truncated power law (TPL), log-normal (LN) and stretched exponential (S-EXP). The values in bold are the parameters of the best fit according to the Kolmogorov-Smirnov test. 42

3.8 Degree-degree distance distributions of the components. Empirical distribution (**dot**), estimates (**line**).The distributions under test are : power law, truncated power law, log-normal and stretched exponential. The values in bold are the parameters of the best fit according to the Kolmogorov-Smirnov test. 43
3.9 Figure A represents the proportion of airports in each main local component of the world air transportation network. Figures B to H reports the distribution of airports in the various countries of the seven large local components. Figure I presents the distribution of airports per country in the large global component. . 45
3.10 The twenty small local components. The blue circles show the significant small components (size greater than five airports). The green circles show the small components with sizes up to five airports. The geographical areas not appearing in the figure do not contain small components. 46
3.11 Typical small local components structures. The left image presents an example of string-based topology. It is the largest small local component in Canada. The right image gives an example of star-based topology. It is the second largest small local component in Alaska. 48
3.12 (A) The airports in the large global component. (B) The 8 small global components which are circled. The size range are between 2 and 3 airports. 50
3.13 The degree (**left**) and degree-degree distance (**right**) distributions of the large global component. Empirical distribution (**dot**), estimates (**line**).The distribution under test are : power law (PL), truncated power law (TPL), log-normal (LN) and stretched exponential (S-EXP). The values in bold are the parameters of the best fit according to the Kolmogorov-Smirnov test. 51
3.14 The degree (**left**) and degree-degree distance (**right**) distributions of the world air transportation network. Empirical distribution (**dot**), estimates (**line**). The distribution under test are : power law (PL), truncated power law (TPL), log-normal (LN) and stretched exponential (S-EXP). The values in bold are the parameters of the best fit according to the Kolmogorov-Smirnov test. 53
3.15 Core of the large local components. The dots represent the airports, and K is the maximum k-core. The core of the North and Central America-Caribbean component contains 42 airports. The core of Europe contains 65 airports. The core of East and Southeast Asia contains 38 airports. The core of Africa-Middle East-Southern Asia contains 26 airports. The core of Oceania contains 8 airports. The core of South America contains 16 airports. The core of Russia-Central Asia-Transcaucasia includes 19 airports. 65
3.16 Figure A presents the 79 airports in the maximum k-core of the world air transportation network. Figure B presents the 24 airports in the maximum k-core of the large global component. The red dots are the airports and K is the maximum k-core. 66
3.17 Small local components (size greater than 5 nodes). (A) Small components located in Alaska (USA). (B) small components in Ontario (Canada). (C) Small components in Green-Land and Norway (Europe). (D) Small component in Algeria and Kenya (Africa). 68

4.1 A) represents the communities identified by the Louvain community detection algorithm. It includes seventeen communities. B) presents the communities uncovered by the Combo community detection algorithm. It contains seven communities. For both, each color represents a community. Similar communities are similar colors. 88

4.2 A represents the airports in the large components of the weighted network. B represents the airports in the large components of the unweighted network. Each color is associated with a component. The North and central America-Caribbean component is red (1). The Africa-Middle East-India component is green (2). The South American component is brown (3). The Europe-Russia-Central Asia-Central Asia component is black (4). It splits in Europe in black (4) and Russia in purple (7) in the unweighted network. The East and Southeast Asia-Oceania component (5) breaks in East and Southeast Asia in blue (5) and Oceania in orange (6) in the unweighted network. 90
4.3 Small local components in the weighted (A) and unweighted (B) world air transportation network. The red circles are the components disappearing in the weighted network. The blue circles are similar small components in both networks. The green circles are the components that do not appear in the unweighted network. The orange circles are the components splitting in the weighted network. Geographical areas outside the figure do not contain small components. 91
4.4 A) The airports in the large global component. B) The 11 small global components are circled. The size range are between 2 and 3 airports. 93
4.5 Average clustering coefficient versus degree of the large components. The blue dots represent the average clustering coefficient for the unweighted network. The red dots represent the average clustering coefficient for the weighted network. The blue curve is the estimated $k^{-\beta}$ function. 95
4.6 Average degree-degree correlation versus degree of the large components. The blue dots represent the average degree-degree correlation for the unweighted network. The red dots represent the average degree-degree correlation for the weighted network. 96
4.7 Strength distributions of the large local components and the largest global component. Empirical distribution (**dot**), Estimates (**line**).The distributions under test are : Power Law (PL), Truncated Power Law (TPL), Log-Normal (LN), and Stretched Exponential (S-EXP). The values in bold are the parameters of the best fit according to the Kolmogorov-Smirnov test. 97
4.8 Strength as a function of degree of the large local components and the largest global component. Empirical distribution (**dot**), Estimates (**line**). 99
4.9 A) represents the average clustering coefficient as a function of k. B) represents the average degree-correlation of degree k. For both figures, the red point is the values of the weighted world transportation network, while the blue points are the unweighted world transportation network values. 100
4.10 The strength distribution (**left**) of the world air transportation network. Empirical distribution (**dot**), Estimates (**line**).The distribution under test are Power Law (**PL**), Truncated Power Law (**TPL**), Log-Normal (**LN**), and Stretched Exponential (**S-EXP**). The values in bold are the best fit parameters according to the Kolmogorov-Smirnov test. The strength as a function of degree (**right**) of the large global component. 100
4.11 The RBO of the large components (A) and the world air transportation network (B). The top 50 hubs of the unweighted (degree) and weighted (strength) are compared by step of 5. 107

4.12 The core of the weighted large local components. S is the maximum s-core value, and K is the maximum k-core value. Blue points represent the airports belonging to the maximum s-core and the maximum k-core, the red points are the airports exclusively in the maximum s-core, and the yellow points are the airports belonging only to the maximum k-core. The maximum s-core of the North and central America-Caribbean component contains 26 airports. The maximum s-core of Europe-Russia-Central Asia-Central Asia contains 27 airports. The maximum s-core of East and Southeast Asia-Oceania includes 18 airports. The maximum s-core of Africa-Middle East-Southern Asia contains 9 airports. The maximum s-core of South America contains 10 airports. 108
4.13 A) The 33 airports in the maximum s-core of the weighted world air transportation network. B) The 22 airports in the maximum s-core of the weighted large global component. S is the maximum s-core value, and K is the maximum k-core value. Blue points represent the airports belonging to the maximum s-core and the maximum k-core, the red points are the airports exclusively in the maximum s-core, and the yellow points are the airports belonging only to the maximum k-core. 118

5.1 (A) Process to uncover the component structure. The community structure is used for example, to extract the dense parts (B) Attack on the network. 127
5.2 Left figure :The airports in the seven large local components. Each color represent a component. Right figure : The largest global component. 128
5.3 The size of the giant component of the world air transportation network, and the large components as a function of removed nodes under targeted attacks on the world air transportation network. 129

6.1 This figure illustrates the process of uncovering the component structure using a community detection algorithm to extract the dense parts forming the local components. After revealing the local components, one builds the global components with the nodes and links shared by the local components. In this example, the network has three communities colored green, light blue, and dark blue. Therefore, it has three local components. The links and nodes joining these local components form three global components. 136
6.2 This toy example illustrates the influence of an attack on the network component structure. Removing node A in the original network impacts the dark blue local component (One node and three links are removed). It also affects the largest global component, which splits into two components with two nodes and an isolated node. Indeed, A has a global and local role. 138
6.3 The figure at the left represents the airports in the uncovered large local components. They are in well-delimited geographic areas. Each color represents a component. 1) North America-Caribbean, 2) Europe-Russia, 3) East and Southeast Asia-Oceania, 4) Africa-Middle East-Southern Asia, 5) South America. The figure at the right represents the airports of the largest global component. In contrast to local components, they are scattered all over the world. 139

6.4 The figure at the left represents the fraction of nodes in the LCC of the large local and global components as a function of the fraction of removed nodes under a Strength attack on the world air transportation network. We also represent the LCC of the world air transportation network in this situation. The Europe-Russia component and the global component are less sensitive to the attack. The figure at the right shows the weighted efficiency of the large local and global components as a function of the fraction of removed nodes under a Strength attack on the world air transportation network. We also represent the weighted efficiency of the world air transportation network in this situation. Efficiency decreases at a high rate as the fraction of top Strength nodes removed increases. Furthermore, The North America-Caribbean and Africa-Middle East-Southern Asia components appear to be the most vulnerable. 141
6.5 The figure at the left represents the fraction of nodes in the LCC of the large local and global components as a function of the fraction of removed nodes under a Weighted Betweenness attack on the world air transportation network. We also represent the LCC of the world air transportation network in this situation. The Europe-Russia component and the global component are less sensitive to the attack. The figure at the right shows the weighted efficiency of the large local and global components as a function of the fraction of removed nodes under a Strength attack on the world air transportation network. We also represent the weighted efficiency of the world air transportation network in this situation. The weighted efficiency of the world air transportation network and the large components are very sensitive to attacks based on the betweenness centrality. 144
6.6 The figure at the left represents the fraction of nodes in the LCC of the large local and global components as a function of the fraction of removed nodes under a Weighted PageRank attack on the world air transportation network. We also represent the LCC of the world air transportation network in this situation. The Europe-Russia component and the global component are less sensitive to the attack. The figure at the right shows the weighted efficiency of the large local and global components as a function of the fraction of removed nodes under a Strength attack on the world air transportation network. We also represent the weighted efficiency of the world air transportation network in this situation. The world air transportation network appears as the most vulnerable. 147
6.7 LCC of the North America-Caribbean component after its isolation from the network. The red dots are the remaining airports after attacks on the world air transportation network (A. Strength attack, B. Weighted Betweenness attack. C. Weighted PageRank attack). The airports in the LCC isolated by the Strength and Weighted PageRank attacks are concentrated in the US and a small part of Canada. In contrast, airports are scattered in all the areas covered by the initial component after the Weighted Betweenness attack. 150
6.8 LCC of the Europe-Russia component after its isolation from the network. The black dots are the remaining airports after attacks on the world air transportation network (A. Strength attack, B. Weighted Betweenness attack. C. Weighted PageRank attack). The LCC isolated by the Strength and Weighted PageRank attacks covers almost all countries sparsely. The LCC is well distributed in all the initial component areas for the Weighted Betweenness attack. 150

6.9 LCC of the East and Southeast Asia-Oceania component after isolation from the network. The blue dots are the remaining airports after attacks on the world air transportation network (A. Strength attack, B. Weighted Betweenness attack. C. Weighted PageRank attack). The LCC isolated by the Strength attack covers the countries in Oceania. The LCC separated by Weighted PageRank attack, covers East and Southeast Asia major countries. In contrast, the LCC isolated by the Weighted Betweenness attack covers all the regions of the initial component. . . . 151
6.10 LCC of the Africa-Middle East-Southern Asia component after isolation from the network. The green dots are the remaining airports after attacks on the world air transportation network (A. Strength attack, B. Weighted Betweenness attack. C. Weighted PageRank attack). The LCC isolated by the Strength attack covers five countries (Nepal, Bhutan, United Arab Emirates, India, and Bangladesh). The LCC isolated by Weighted PageRank attack covers East Africa and Madagascar countries. The LCC separated by the Weighted Betweenness attack, covers mainly the Middle East and Southern Asia countries. 151
6.11 LCC of the South America component after isolation from the network. The brown dots represent the remaining airports after attacks on the world air transportation network (A. Strength attack, B. Weighted Betweenness attack. C. Weighted PageRank attack). The LCC isolated by the Strength and Weighted PageRank attacks includes Brazil and Venezuela. The LCC separated by the Weighted Betweenness attack, covers all the initial component regions. 153

7.1 A)The component structure B) Attack on the global component. 158
7.2 A) Attack from the largest global component.B) Attack from the world air transportation network. 160
7.3 A) Attack from the largest global component.B) Attack from the world air transportation network. 162

Liste of tables

3.1 Basic topological properties of the world air transportation network. N is the network size. $|E|$ is the number of edges. d is the diameter. L is the average shortest path length. μ is the density. ζ is the transitivity also called global clustering coefficient. λ is the assortativity also called degree correlation coefficient. η is the hub dominance. 33

3.2 Quality metrics of the community structures uncovered by Louvain and LPA community detection algorithms : Modularity, Mixing parameter, NMI 34

3.3 Small-world property of the components. C and C_{rand} are respectively the clustering coefficients of the local components and the random network with same number of nodes and links. L and L_{rand} are respectively the average shortest path length of the components and their corresponding random network. 40

3.4 Kolmogorov-Smirnov goodness of fit values (KS-test) for the degree distribution of the components. Distributions under test are power law, truncated power law, log-normal, and stretched exponential. The smallest value (in bold) corresponds to the best fit. 41

3.5 Kolmogorov-Smirnov goodness of fit values (KS-test) for the degree-degree distance distribution. Distributions under test are power law, truncated power law, log-normal, and stretched exponential. The smallest value (in bold) corresponds to the best fit. 44

3.6 Basic topological properties of the significant small local components (size greater than five airports). N is the network size. $|E|$ is the number of edges. d is the diameter. L is the average shortest path length. L_{rand} and C_{rand} are respectively the average shortest path and the average clustering coefficient of the corresponding random network. μ is the density. ζ is the transitivity. λ is the assortativity also called degree correlation coefficient. η is the hub dominance. 48

3.7 Degree Centrality in the seven large local components. The top five hubs for each large local component are reported. ID is the internal degree. It measures the number of connections of an airport with airports located in its large local component. LR is the airport local rank in its component by decreasing internal degree. CP is the cumulative fraction of connected airports represents the proportion of airports that the top x local hubs can reach in their components. . 54

3.8 Degree centrality in the large global component. The top five inter-regional hubs in each region are reported. ED is the external degree. It measures the number of connections of an airport with airports located in the large global component. GRR is the global rank of the airport in its region. GRC is the global rank in the global component is the airport rank in the global component by decreasing external degree. 58

3.9 The top 36 hubs of the world air transportation network. WR is the worldwide rank in the world air transportation network. It is computed using the total degree (internal and external). LR is the local rank computed in the large local components. GR is the global rank in the global component. No indication is given when an airport does not belong to the global component. Top inter-regional airports are in bold. Top regional and inter-regional airports are in bold and italic. Others are top regional airports. 61
3.10 Core airports of the North and Central America-Caribbean component (42). Airports not in the top 42 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 42 regional hubs and not in the k-core are in red. 69
3.10 Core airports of the North and Central America-Caribbean component (42). Airports not in the top 42 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 42 regional hubs and not in the k-core are in red. 70
3.11 Core airports of the European component (65). Airports not in the top 65 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 65 regional hubs and not in the k-core are in red. 71
3.11 Core airports of the European component (65). Airports not in the top 65 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 65 regional hubs and not in the k-core are in red. 72
3.12 Core airports of the East and Southeast Asia component (38). Airports not in the top 38 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 38 regional hubs and not in the k-core are in red. 73
3.12 Core airports of the East and Southeast Asia component (38). Airports not in the top 38 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 38 regional hubs and not in the k-core are in red. 74
3.13 Core airports of the Africa-Middle East-Southern Asia component (26). Airports not in the top 26 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 26 regional hubs and not in the k-core are in red. 75
3.14 Core airports of the Oceania component (8). Airports not in the top 8 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 8 regional hubs and not in the k-core are in red. 76

3.15 Core airports of the South America component (16). Airports not in the top 16 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 16 regional hubs and not in the k-core are in red. 77
3.16 Core airports of the Russia-Central Asia- Transcaucasia component (19). Airports not in the top 16 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 19 regional hubs and not in the k-core are in red. 78
3.17 Core airports of the global component (24). Airports not in the top 24 inter-regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Global rank is the rank in the large local component. Airports in the top 24 inter-regional hubs and not in the k-core are in red. 79
3.18 These airports are not in the core of the European local component, but they belong to the core of the world air transportation network. Airports in the European local component are in black. Airports from other local components are in red. The core degree is the number of links in the maximum k-core network. Worldwide rank is the rank in the world air transportation network. Local rank is the rank in the large local components. Global rank is the rank in the global component. 80
3.19 Small global components. The Local component number corresponds to the number in Figure 4.3. 81

4.1 Basic topological properties of the world air transportation network and the global components. N is the network size. $|E|$ is the number of edges. d is the diameter. L is the average shortest path length. μ is the density. ζ and ζ^w are, respectively, the unweighted and weighted average clustering coefficients. λ and λ^w are respectively the unweighted and weighted assortativity, also called the degree-degree correlation coefficient. η is the hub dominance. 84
4.2 Quality metrics of the community structures uncovered by Louvain and Combo community detection algorithms : Modularity, Mixing parameter, NMI 87
4.3 The Jaccard Index of the five similar communities uncovered by Louvain and Combo. 87
4.4 The Jaccard Index of the communities uncovered by the weighted and unweighted Louvain algorithm. Three are similar (North and Central America-Caribbean, Africa-Middle East-Southern Asia, and South America), and two are regrouped by the weighted algorithm. The East and Southeast Asia-Oceania regroup the East and Southeast Asia and the Oceania components. Europe-Russia-Central Asia-Central Asia regroups the Europe and Russia-Central Asia components. . . 90
4.5 Quality metrics of the community structures uncovered by the unweighted and weighted Louvain community detection algorithms : Modularity, Mixing parameter, NMI 90
4.6 Kolmogorov-Smirnov goodness of fit values (KS-test) for the degree distribution. Distributions under test are Power-law, Truncated power law, Log-Normal, and Stretched exponential. The smallest value (bold) corresponds to the best fit. . . . 99

4.7 Strength Centrality in the five large local components. We report the top five hubs for each large local component. The internal Strength of an airport is the number of flights with airports in its large local component. Local Rank is the airport rank in its component by decreasing internal strength. The cumulative fraction of connected airports represents the proportion of airports that the top x local hubs can reach in their components. 102
4.8 Strength Centrality in the large global component. We report the top five inter-regional hubs in each region. External Strength is the number of connections of an airport with airports located in the large global component. GRR is the Global Rank in the Region is the airport Rank in its region by decreasing external Strength. GRC is the Global Rank in the Component is the airport rank in the global component by decreasing external Strength. 104
4.9 The top 25 nodes in the world air transportation network are in descending order of strength centrality. Worldwide strength rank is the strength centrality rank in the world air transportation network. It is computed using the total strength (Internal and external). WSR is the Worldwide Strength Rank. LSR is the Local Strength rank is the strength centrality rank in the large local components. It is calculated using the Internal strength. GSR is the Global Strength Rank is the strength centrality rank in the global component. It is computed using the External strength. No indication is given when an airport does not belong to the global component. Top inter-regional airports are in bold. Top regional and inter-regional airports are in bold and italic. Others are top regional airports. WDR is the Worlwide Degree Rank. 109
4.10 26 Max s-core airports of the North and central America-Caribbean component. Airports not in the top 26 regional airports with the highest Strength airports are in bold. The Core degree of a node is its number of flights in the maximum s-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports are ranked from largest to smallest strength. 110
4.11 27 Max s-core airports of the Europe-Russia-Central Asia component. Airports not in the top 27 regional airports with highest Strength are in bold. The Core strength of a node is its number of flights in the maximum s-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports are ranked from largest to smallest strength. 111
4.12 18 Max s-core airports of the East and Southeast Asia-Oceania component. Airports not in the top 18 regional airports with the highest Strength are bold. Core strength of a node is its number of flights in the maximum s-core. Local rank is the airport rank in the large local component. National rank is the airport rank in its country. Airports are ranked from largest to smallest strength. 112
4.13 9 Max s-core airports of the Africa-Middle East-Southern Asia component. Airports not in the top 9 regional airports with the highest Strength are in bold. The Core strength of a node is its number of flights in the maximum s-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports are ranked from largest to smallest strength. 113
4.14 10 Max s-core airports of the South America component. Airports not in the top 10 regional airports with the highest Strength are in bold. The Core strength of a node is its number of flights in the maximum s-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 10 regional hubs and not in the k-core are in red. Airports are ranked from largest to smallest strength. 114

4.15 22 Max s-core airports of the global component. Airports not in the top 22 inter-regional airports with the highest Strength are in bold. The Core strength of a node is its number of links in the maximum s-core. Global rank is the rank in the largest global component. Airports are ranked from largest to smallest strength. . 115
4.16 11 airports of the weighted world air transportation network which are not in the Max s-core of the North and Central America-Caribbean component. Airports not in the top 33 worldwide airports with the highest Strength are in bold. The Core Strength of a node is its number of flights in the maximum s-core. Global rank is the rank in the large local component. 116
4.17 Summary table of the main results 118
4.17 Summary table of the main results 119
4.17 Summary table of the main results 120

5.1 Basic topological properties of the world air transportation network, the 7 large local components, and the largest global component. N is the network size. $|E|$ is the number of edges. *diam* is the diameter. l is the average shortest path length. ν is the density. ζ is the transitivity also called global clustering coefficient. $k_{nn}(k)$ is the assortativity also called Degree correlation coefficient. η is the hub dominance. 128
5.2 Basic topological properties of the 7 large isolated local components after the attack based on the Degree. *LCC* is the Largest Connected Component. $|E|$ is the number of edges. *diam* is the diameter. l is the average shortest path length. ν is the density. ζ is the transitivity also called global clustering coefficient. $k_{nn}(k)$ is the assortativity also called Degree correlation coefficient. 129
5.3 Basic topological properties of the 7 large isolated local components after the attack based on the Betweenness. *LCC* is the Largest Connected Component. $|E|$ is the number of edges. *diam* is the diameter. l is the average shortest path length. ν is the density. ζ is the transitivity also called global clustering coefficient. $k_{nn}(k)$ is the assortativity also called Degree correlation coefficient. 131

6.1 Basic topological properties of the world air transportation network, the five large local components, and the largest global component. N is the network size. $|E|$ is the number of edges. δ is the diameter. l is the average shortest path length. μ is the density. ζ is the transitivity, also called the global clustering coefficient. $k_{nn}(k)$ is the assortativity, also called the Degree correlation coefficient. 139
6.2 Basic topological properties of the five local components isolated by the Strength attack. λ is the size of the isolated component. ρ is the fraction of top Strength nodes removed from the world air transportation network to isolate the component. δ is the diameter. l is the average shortest path length. μ is the density. . 141
6.3 Evolution of the weighted efficiency, the LCC of the large local components after their isolation due to attacks based on Strength. The order of isolation of the components is also presented. ω_i is the Weighted Efficiency before attacks based on Strength. ω is the weighted efficiency after the separation. λ is the Largest Connected Component size after the separation. ρ is the fraction of top Strength nodes removed from the world air transportation network to isolate the component. γ is the order in which the components are isolated. After their isolation, the components lose much of their efficiency. 143
6.4 Basic topological properties of the five local components isolated by the Betweenness attack. λ is the Largest Connected Component size. ρ is the fraction of top Strength nodes removed from the world air transportation network to isolate the component. δ is the diameter. l is the average shortest path length. μ is the density. 145

6.5 Evolution of the weighted efficiency, the LCC of the large local components after their isolation due to attacks based on Weighted Betweenness. We also report the order of isolation of the components. ω_i is the Weighted Efficiency before attack based on Weighted Betweenness. ω is the weighted efficiency after the separation. λ is the Largest Connected Component size after the split. ρ is the fraction of top Weighted Betweenness nodes removed from the world air transportation network to isolate the component. $\gamma$is the order in which the components are isolated. Globally, after their isolation, the efficiency of the LCC does not change much. Some LCCs are even more efficient. 146
6.6 Basic topological properties of the local components isolated by the Weighted PageRank attack. λ is the Largest Connected Component size. ρ is the fraction of top Strength nodes removed from the world air transportation network to isolate the component. δ is the diameter. l is the average shortest path length. μ is the density. 147
6.7 Evolution of the weighted efficiency, the LCC of the large local components after their isolation due to attacks based on Weighted PageRank. We also report the order of isolation of the components. ω_i is the Weighted Efficiency before attack based on PageRank. ω is the weighted efficiency after the separation. λ is the Largest Connected Component size after the split. ρ is the fraction of top PageRank nodes removed from the world air transportation network to isolate the component. γ is the order in which the components are isolated. Isolating the components reduces their efficiency significantly. 149
6.8 This table summarizes the results for the attacks under evaluation (Strength, Weighted Betweenness, and Weighted PageRank). We report the order of isolation, the proportion of airports removed, and the LCC of the local component. The Jaccard Index between the LCCs for various couples of attacks is reported for each large component. γ_s, γ_b, γ_p are respectively the order of isolation of the LCC from Strength, Weighted Betweenness and Weighted PageRank. ρ_s, ρ_b, ρ_p are respectively the proportion of attacked airports from the world air transportation network according to the Strength, Weighted Betweenness and Weighted PageRank. λ_s, λ_b, λ_p are respectively the LCC from Strength, Weighted Betweenness and Weighted PageRank. J_{sb}, J_{sp}, J_{bp} represent respectively the Jaccard index of the LCCs isolated by Strength and Weighted Betweenness, Strength and Weighted PageRank, Weighted Betweenness and Weighted PageRank. 152
6.9 Variation of the Weighted Efficiency of the different attacks once the component splits from the overall network. ω_s, ω_b, ω_p, are respectively the fraction of the Weighted Efficiency of the LCCs for Strength, Weighted Betweenness, and Weighted PageRank attacks. The Weighted Betweenness attack does not impact much the efficiency of the components compared to the other attacks. 152

7.1 Basic topological properties of the world air transportation network, the 7 large local components, and the largest global component. N is the network size. $|E|$ is the number of edges. $diam$ is the diameter. l is the average shortest path length. ν is the density. ζ is the transitivity, also called the global clustering coefficient. $k_{nn}(k)$ is the assortativity, also called the Degree correlation coefficient. η is the hub dominance. 158

7.2 **Basic topological properties of the five local components isolated by the Degree attack.** λ is the Largest Connected Component size. ρ is the fraction of top Strength nodes removed from the world air transportation network to isolate the component. δ is the diameter. l is the average shortest path length. μ is the density. 161
7.3 **Basic topological properties of the five local components isolated by the Betweenness attack.** λ is the Largest Connected Component size. ρ is the fraction of top Strength nodes removed from the world air transportation network to isolate the component. δ is the diameter. l is the average shortest path length. μ is the density. 163
7.4 γ, ρ, λ are the order of isolation of local components, the proportion of airports removed from the global air network before isolation, and LCCs of the local components after the separation. 164

Chapter 1

Introduction Générale [1]

Summary

Ce chapitre introduit la thèse en donnant les définitions de base des réseaux complexes et les prérequis qui seront nécessaires à la bonne compréhension des résultats. De plus, il énumère le contexte, les contributions et l'organisation de la thèse.

1. Cette thèse (i.e. les chapitres suivants) est rédigée en anglais à l'exception de cette introduction générale qui est en français conformément aux exigences du règlement intérieur de l'école doctorale.

This thesis (i.e. the following chapters) is written in English with the exception of this general introduction which is in French in accordance with the requirements of the internal regulations of the doctoral school.

Les réseaux complexes sont un domaine de la science qui permettent de modéliser des systèmes complexes. Ils se sont développés au début du 21e siècle.

Cette science repose fortement sur les graphes qui représentent le mieux les interactions de différents objets, personnes, sites, entités, etc. Les réseaux complexes sont transversaux peuvent être utilisé en informatique, mathématiques, biologie, économie, science sociale. Ils permettent de comprendre plusieurs phénomènes dans différents systèmes complexes tels que le cerveau, les réseaux informatiques, la propagation d'épidémie, les systèmes financiers, les systèmes de transports, les entreprises, etc.

1 Les réseaux réels

Selon Newman, les réseaux réels peuvent être classés en 4 groupes : les réseaux technologique, les réseaux sociaux, les réseaux d'information et les réseaux biologiques [Newman, 2018].

1.1 Les réseaux technologiques

Les réseaux technologiques ou réseaux d'infrastructure sont créés par l'homme pour relier ou connecter des entités qui sont dans des zones géographiques éloignées. Ces types de réseaux permettent de répartir des services. Parmi les réseaux technologiques, on peut citer les réseaux de transports, Internet, les réseaux électriques, les réseaux téléphoniques, les réseaux de distributions.

Les réseaux de transports participent beaucoup à la mobilité des personnes et des marchandises. Ils jouent un rôle important dans l'économie à différentes échelles : ville, pays, monde. Ils regroupent les réseaux routiers, les réseaux aériens, les réseaux ferroviaires, les réseaux maritimes, etc. Dans ces réseaux, les noeuds sont des positions géographiques et les liens sont définis par le déplacement (origine-destination) entre ces positions géographiques. Ces réseaux peuvent être non orientés, non pondérés, orientés, pondérés (nombre de passagers, nombre de voyages, prix d'un voyage).

Internet est un réseau physique, par opposition au WWW, constitué de matériels (routeur, serveur, ordinateur, téléphones) reliés par des câbles, fibre ou l'air. Il s'est rapidement développé dans le monde depuis le début des années 2000. Son évolution a fait qu'il soit très étudié, car devenant de plus en plus vital et complexe. Ainsi, la structure de ce réseau peut être représentée en 2 niveaux [Menczer, 2020]. Le premier niveau est constitué de routeurs qui sont connectés entre eux dans un réseau local. Le second niveau regroupe les Systèmes Autonomes (Autonomous System-AS) reliés entre eux. Chaque AS est un ensemble de réseaux du premier niveau.

Les réseaux électriques sont composés d'une partie de production, de distribution et de transport. Dans ces types de réseaux, les nœuds peuvent être les lieux de production et de consommation et les liens sont les lignes de transmission. Ces réseaux peuvent être pondérés par exemple par l'impédance.

Les réseaux de transport et de distribution électriques sont vulnérables. Une défaillance d'un nœud ou d'un lien peut se propager sur le réseau et peuvent causer d'énormes pertes économiques

et impacter la vie quotidienne de beaucoup de personnes [Wilkinson, 2012]. La résilience de ces réseaux est souvent étudiée ainsi que de nouvelles approches pour atténuer le risque de pannes.

1.2 Les réseaux sociaux

Un réseau social peut être défini comme un groupe de personnes (association, entreprise, gouvernement, etc) qui ont des relations particulières (connaissances, collaboration, hiérarchie, etc). Les personnes représentent les nœuds et les relations sont les liens. Ces types de réseaux ont été étudiés depuis longtemps par les sociologues [Milgram, 1967]. Ces derniers utilisent le terme acteur pour désigner un nœud d'un réseau social. Il y a plusieurs types de réseaux sociaux. On peut en citer les réseaux de co-auteurs dans lesquels les acteurs sont reliés s'ils ont écrit au moins un article ensemble, les réseaux de collaboration des acteurs de films qui regroupent les acteurs qui sont apparus ensemble dans un film, les réseaux de contact qui sont définis par les individus qui sont physiquement en contact, etc. À côté de ces types de réseaux sociaux, il y a les médias sociaux tels que Facebook, Twitter, Instagram, etc.

L'analyse des réseaux sociaux est devenue un domaine qui intéresse beaucoup de chercher, surtout avec la disponibilité des données en ligne. L'analyse des réseaux sociaux est l'ensemble des méthodes et techniques qui permettent de comprendre et de découvrir les modèles d'interactions entre les acteurs d'un réseau social. Comme exemple d'application [Carrington, 2005 ; Serrat, 2017] de l'analyse des réseaux sociaux, on peut citer l'identification des acteurs les plus influents, la détection de communauté, la propagation d'épidémie, la diffusion d'information, la prédiction de liens, etc.

1.3 Les réseaux d'information

Les réseaux d'information intègrent les connaissances sur la structure du réseau. Cette catégorie de réseau inclus les réseaux de citations, de cocitation et le WWW.

Dans les réseaux de citation, les nœuds du réseau sont les articles. Ce sont des réseaux orientés, dans lesquels on établit un lien de A vers B si l'article A cite l'article B. Les articles qui sont les plus cités sont les plus influents et les plus lus. Une alternative aux réseaux de citation sont les réseaux de cocitation. Dans ces derniers, les nœuds sont les articles. On établit un lien entre deux articles s'ils sont cités dans un même article. Les réseaux de cocitation sont non orientés.

Dans le réseau WWW, les nœuds sont les pages web et les liens sont les hyperliens qui permettent de naviguer d'une page à un autre. Ainsi, c'est un réseau orienté et c'est le plus grand réseau d'information avec des milliards de nœuds et de liens. Ces réseaux sont très étudiés et beaucoup de robots d'indexation (web crawler) sont mis au point pour extraire une partie de ce réseau.

1.4 Les réseaux biologiques

Les réseaux biologiques représentent les interactions complexes entre différentes espèces animales ou des composantes cellulaires telles que les protéines, les métabolites, les cellules à l'in-

térieur du cerveau, etc. On peut en citer à titre d'exemples les réseaux biochimique, les réseaux cérébraux, et les réseaux écologiques.

Les réseaux biochimiques représentent les modèles d'interaction au niveau des molécules dans une cellule biologiques. On peut en citer les réseaux métaboliques, les réseaux de protéines, les réseaux de régulation génétiques.

Il y a deux catégories de réseaux cérébraux. La première concerne les réseaux microscopiques de connexions qui représentent les interactions des cellules cérébrales individuelles. La seconde catégorie concerne les réseaux macroscopiques de connexions qui relient différentes régions du cerveau. Les réseaux cérébraux incluent les réseaux de neurones, les réseaux de connectivité fonctionnelle dans le cerveau.

Les réseaux écologiques décrivent les interactions entre les espèces. En effet, ces dernières se mangent entre eux, partagent des ressources, etc. Le réseau écologique le plus étudié est le réseau des chaines alimentaires qui illustre la chasse entre les animaux (proie-prédateur).

2 Les graphes

2.1 Définition

Un graphe G est un objet mathématique défini par le couple (V, E) où V est un ensemble non vide de nœuds et E un ensemble de liens ou arrêtes qui connectent les noeuds ($E = (v_1, v_2) \mid v_1, v_2 \in V$). S'il existe un noeud qui peut être lié à lui-même, le graphe contient une boucle. Un graphe dans que peut avoir plusieurs liens entre deux noeuds est appelé un multigraphe. Un graphe qui n'a pas de boucle et qui n'est pas un multigraphe est un graphe simple. Un graphe peut être non orienté ou orienté, pondéré ou non pondéré.

— Dans un graphe non orienté, les liens n'ont pas de direction, le lien $(v_1, v_2) = (v_2, v_1)$. Par exemple, Facebook peut être défini par un graphe non orienté. Les noeuds sont les personnes et les liens sont les amitiés.

— Dans un graphe orienté, la direction des liens a un sens, le lien $(v_1, v_2) \neq (v_2, v_1)$. Si (v_1, v_2) est un lien dans un graphe orienté, v_1 est appelé l'origine, v_2 est la destination. Par exemple, le réseau world wide web peut être défini par un graphe orienté. Les noeuds sont les pages web. Les liens sont les hyperliens. Si une page web B est joignable à partir d'une page web A, le lien sera dirigé de A vers B.

— Dans un graphe pondéré, chaque lien est associé à une valeur, un poids ou une intensité. Par exemple, le réseau de transport aérien dans lequel les noeuds sont les aéroports et les liens sont les vols, on peut définir un poids tel que le nombre de passagers ou le nombre de vols entre deux aéroports.

— Dans un graphe non pondéré tous les liens sont de poids 1.

On peut combiner ces différents types de graphes simples. Néanmoins, les graphes non orientés et non pondérés sont les plus utilisés.

Les graphes énumérés en haut sont différenciés par la nature des liens. Les graphes bipartis sont des graphes où on a deux types de noeuds. Un graphes biparti peut être défini par deux

ensembles disjoints de noeuds connectés par des liens. C'est un triplet (U, V, E) tel que $(u, v) \subset E \Rightarrow u \in U, v \in V$. Avec les acteurs et les films comme deux types de noeuds, on peut définir un graphe biparti dans lequel un acteur est lié à un film où il apparait.

2.2 Matrice d'adjacence

Soit $G = (V, E)$ un graphe de N noeuds, non orienté et non pondéré, sa matrice d'adjacence $A = (a_{ij})$ de taille $N * N$ est défini comme suit :

$$a_{ij} = \begin{cases} 1 & (i,j) \in E \\ 0 & sinon \end{cases} \tag{1.1}$$

Un noeud i est adjacent à noeud j s'il existe un lien entre ces deux noeuds. i et j sont des voisins.

2.3 Degré d'un noeud

Le degré d'un noeud i est le nombre de noeuds adjacent à i. On le note k_i et il est défini comme suit :

$$k_i = \sum_{j=1}^{N} a_{ij} \tag{1.2}$$

N est le nombre de noeuds du graphe, a_{ij} est un élément de la matrice d'adjacence A.

Dans le cas d'un graphe orienté, on a le **degré entrant** et le **degré sortant** pour un noeud. Le degré entrant d'un noeud i est le nombre de noeuds qui pour destination i. Le degré sortant i est le nombre de noeuds qui ont pour origine i. Ils sont exprimés comme suit :

$$k_i^{entrant} = \sum_{j=1}^{N} a_{ji} \tag{1.3}$$

$$k_i^{sortant} = \sum_{j=1}^{N} a_{ij} \tag{1.4}$$

2.4 Chaine

Soit $G = (V, E)$ un graphe de taille N, une **chaine** est une suite ordonnée de liens :

$$\{(v_1, v_2), (v_2, v_3)...(v_{N-1}, v_N)\} \, tel \, que \, \forall i \in 1, 2, ..., N, (v_{i-1}, v_i) \in E. \tag{1.5}$$

Les extrémités de la chaine v_1 et v_2 sont respectivement appelés l'origine et la destination. Un noeud ou un lien peut être visité plus d'une fois dans une chaine. La chaine est dite **élémentaire** si tous ses noeuds sont distincts. La chaine est dite **fermée** si l'origine et la destination sont les mêmes. Si tous les liens sont distincts, la chaine est dite **simple**. Un **circuit** est une chaine simple et fermée.

2.5 Chemin

Un **chemin** est une chaine qui contient plusieurs noeuds(au moins 2 noeuds) distincts (sauf l'origine et la destination). Le chemin est appelé **cycle** si l'origine et la destination sont confondues. La *longueur* d'un chemin est le nombre de liens constituant le chemin si le graphe n'est pas pondérée. Si le graphe est pondéré, c'est la somme des poids des liens constituant le chemin.

2.6 Composantes

Deux noeuds i et j sont joignables s'il existe un chemin de i vers j. Un graphe est **connexe** si chaque pair de noeud est joignable. Un graphe qui n'est pas connexe contient des parties disjointes appelées **composantes**. Il n'y a pas de chemin entre les noeuds de différentes composantes. Les composantes sont des sous-graphes connexes. Un graphe connexe contient une et une seule composante connexe. Si un graphe présente plusieurs composantes, la plus grande d'entre elles, c'est-à-dire celle qui contient le plus de noeuds, est appelé la **composante géante**. Généralement, toutes les expérimentations se font sur un graphe connexe. Si le graphe n'est pas connexe, on utilise sa composante géante. Une **clique** est un sous-graphe ou graphe dans lequel, chaque pair de noeuds est connecté. C'est un **graphe complet**.

2.7 Distance

Le plus court chemin d'un i vers un noeud j d'un graphe est la longueur minimale entre ces deux noeuds. On définit la **distance** entre deux noeuds i et j d'un graphe comme étant le plus court chemin entre ces deux noeuds. On le note d_{ij}. S'il n'y a pas de chemin entre i et j, $d_{ij} = \infty$. Aussi, $d_{ii} = 0$.

3 Niveau macroscopique

Il y a trois niveaux, pour examiner un réseau complexe : le niveau macroscopique, le niveau microscopique et le niveau mésoscopique. L'analyse macroscopique caractérise l'ensemble de la topologie du réseau par un ensemble de mesures globales. Les études microscopiques examinent les propriétés du réseau au niveau des nœuds ou des liens. L'analyse mésoscopique concerne les groupes de nœuds ou de liens partageant des caractéristiques similaires.

3.1 Densité

Pour un graphe de taille N, le nombre de noeuds possibles L est donnée par la formule suivante :

$$L = \frac{1}{2} N * (N - 1) \tag{1.6}$$

La densité d'un graphe indique la probabilité que deux noeuds soient connectés. C'est aussi le nombre de liens sur le nombre de liens possible. La densité μ d'un graphe de taille N est défini par la formule suivante :

$$\mu = \frac{2 * |E|}{N(N-1)} \tag{1.7}$$

Un graphe est épars si $\mu \ll 1$ ou $|E| \ll L$. Il est dense si $\mu \simeq 1$.

3.2 Diamètre

Le **diamètre** d'un graphe, *diam*, est le maximum de tous les plus courts chemins.

$$diam = \quad max(\{d_{ij}\}) \quad i, j \in V \tag{1.8}$$

3.3 La moyenne des plus court chemin

La moyenne des plus courts chemins d'un graphe l_G est la moyenne des distances entre tous les noeuds du graphe. Il est défini comme suit :

$$l_G = \frac{1}{N(N-1)} \sum_{i,j \in V, i \neq j} d_{ij} \tag{1.9}$$

3.4 Coefficient de clustering

Le coefficient de clustering mesure le degré de connectivité des voisins d'un noeud. Il vient du concept qui est étudié dans les réseaux sociaux : l'ami de mon ami est mon ami. Il peut être défini comme la probabilité que des voisins soient connectés. Il y a deux versions du coefficient de clustering : le Coefficient de clustering local et le coefficient de clustering global.

3.4.1 Coefficient de clustering local

Le coefficient de clustering local permet de mesurer le degré de connectivité des voisins au tour d'un nœud. Pour un nœud i, c'est le nombre de paires de voisins connectés sur le nombre de paires de voisins. Il est compris entre 0 et 1. Pour un nœud i, Watts et Strogatz proposent la formule suivante

$$C_i = \frac{2 * L_i}{k_i(k_i - 1)} \tag{1.10}$$

L_i est le nombre de liens entre les voisins des nœuds i. k_i est le degré de i. À partir de cette formule, on déduit le coefficient de clustering moyen comme étant la moyenne des coefficients de clustering des nœuds.

$$\bar{C} = \frac{1}{N} \sum_{i \in V} C_i \tag{1.11}$$

$\bar{C}$ est compris entre 0 et 1.

3.4.2 Coefficient de clustering global

Le coefficient de clustering global appelé aussi **transitivité** permet de caractériser si dans un réseau il y a des groupements de trois nœuds. Il définit la proportion de triangles dans un réseau. Sa formule est donnée par :

$$\zeta = \frac{3 * \text{nombre de triangles}}{\text{nombre de triplets}} \tag{1.12}$$

Un triplet est formé par trois nœuds dont au moins deux entre eux sont connectés. Le 3 du numérateur est dû au fait que chaque triangle est composé par 3 nœuds.

3.5 Assortativité

L'assortativité ou le **degré de corrélation** permet de savoir si les nœuds de même nature ont tendance à se connecter entre eux ou pas. Par nœud de même nature, on fait allusion au degré d'un nœud. Ainsi, un réseau est assortatif si les nœuds de même nature (nœud de forte degré ou nœud de faible degré) ont des liens entre eux. Dans le cas contraire, on dit que le réseau est non-assortatif, c'est-à-dire les nœuds de forte degré ont tendance à se lier aux nœuds de faible degré.

$$r = \frac{\sum_{jk} jk(e_{jk} - q_j q_k)}{\sigma_q^2} \tag{1.13}$$

q_k est le nombre de liens quittant le noeud, autres que celles qui relie la paire. C'est la distribution du degré restant. σ_q est l'écart type de la distribution q_k.

3.6 Distribution des degrés

Le degré d'un nœud est le nombre de voisins de ce nœud. Le degré de distribution $P(k)$ d'un graphe est définie comme étant le pourcentage de nœuds de degré k. Il dénote la probabilité qu'un nœud choisi au hasard soit de degré k. Le degré de distribution est défini comme suit :

$$P(k) = \frac{N(k)}{N} \tag{1.14}$$

$N(k)$ est le nombre de nœuds de degré k, N est le nombre de nœuds total.

Pour avoir une information globale sur le réseau, il faut le nuage de point de la distribution des degrés. On met en abscisse les degrés et en ordonnée, on met les degrés de distribution. Dans la plupart des distributions des degrés des réseaux du monde réel, on voit qu'il y a beaucoup de nœuds de faibles degrés et peu de nœuds de forts degrés. Plusieurs distributions de degrés du monde réel suivent une loi puissance ($P(k) = k^{-\lambda}$) ou une loi qui l'est similaire. Pour savoir quelle fonction est plus proche de la distribution des degrés, on utilise les testes de KS.

3.6.1 Le test de K-S

Le test de Kolmogorov-Smirnov (test KS) est utilisé pour évaluer l'adéquation d'un ensemble de données à une distribution théorique. Le test KS se base sur une comparaison point à point. On mesure la distance entre un point de la distribution et la fonction de répartition d'une loi donnée. Si s'il y a plusieurs fonctions de répartitions en compétition, celle qui a le plus KS avec la distribution est la plus proche de la distribution.

3.7 Les modèles de réseaux complexes

La modélisation des réseaux complexes est un domaine qui a très tôt intéressé les chercheurs. Les premiers travaux dans ce sens avaient pour objectif simple de modéliser un graphe. Ces types

de modèle ont une importance capitale puisque au début de la science des réseaux, il y avait un manque de données. Pour tester certaines propriétés des réseaux (graphes), il fallait générer des graphes à partir de modèles. Avec l'avènement de données réelles, on a vu que les réseaux ont des propriétés particulières. De nouveaux modèles qui tentent de capturer ces propriétés sont créés.

3.8 Le modèle aléatoire de Erdös-Rényi

Les premières modèles de graphes ont été introduits par Paul Erdös et Alfred Rényi[Erd6s, 1959]. Ils sont souvent appelés **graphes aléatoires**. Il y a deux catégories pour former les graphes de Erdös-Rényi : $G_{N,L}$ et $G_{N,p}$.

— $G_{N,L}$ est un modèle qui permet de générer des graphes aléatoires à N sommets et L liens. On considère d'abord un graphe à N noeuds qui n'a aucun lien. Ensuite, on connecte chaque fois deux noeuds pris manière uniforme et aléatoire jusqu'à atteindre L liens.

— $G_{N,p}$ permet cette fois-ci la génération d'un graphe à N noeuds dont les liens sont connectés aléatoirement. Chaque lien est présent indépendamment des autres avec une probabilité p.

Ces deux modèles étant mathématiquement équivalents, néanmoins pour générer un graphe $G_{N,p}$ la complexité est (N^2), alors que pour générer un graphe $G_{N,L}$ la complexité est L. Avec un même nombre de noeuds N, il existe une relation entre p et L qui est.

$$p * \frac{N(N-1)}{2} = L \tag{1.15}$$

En appliquant l'équation 1.6 de la densité, on obtient que la densité d'un graphe $G_{N,p} = p$. Plus p est grand, plus le nombre de liens augmente. Dans ce cas, à défaut d'avoir un graphe connexe, on peut avoir une composante géante. Si p est petit, il y aura plusieurs composantes.

On peut déduire le degré moyen du graphe.

$$\bar{k} = \frac{pN(N-1)}{N} = (N-1)p \simeq Np \tag{1.16}$$

Pour un graphe $G_{N,p}$, la distribution des degrés k suit une loi Binomiale de paramètre $N-1$ et p

$$P(k = k_i) = \binom{k}{N-1} p^k (1-p)^{N-1-k} \tag{1.17}$$

Lorsque que N est très grand ($N \to \infty$), la distribution des degrés tend vers une loi de Poisson de paramètre z.

$$P(k = k_i) = \frac{z^k exp(-z)}{k!} \tag{1.18}$$

Le paramètre $z = pN$ correspond au degré moyen. La loi de Poisson est plus utilisée pour étudier la distribution des degrés du graphe aléatoire. Cette loi dépendant d'un seul paramètre est plus facile à manipuler.

Les modèles aléatoires capturent certaines propriétés des réseaux réels tels que la composante

géante et la densité. Cependant, leur coefficient de clustering est faible, et la distribution des degrés est homogène, il n'y a pas de hubs.

3.9 Le modèle petit-monde de Watts et Strogatz

Ce modèle s'inspire de l'expérience de Milgram [Milgram, 1967] réalisé en 1967 aux États-Unies. Cette expérience, connu sous le nom de "six degrés de séparation", stipule que deux individus pris au hasard puissent correspondre par lettre par une chaine de 6 relations en moyenne. Elle décrit un phénomène appelé "petit-monde". Le modèle de Watts et Strogatz se base sur deux propriétés des réseaux réels : les diamètres sont petits (ce que montre l'expérience de Milgram), et les coefficients de clustering sont grands.

La structure du modèle petit-monde se situe entre celle d'un graphe régulier et d'un graphe aléatoire. On considère en premier un graphe régulier de N noeuds dont chacun est lié à ses k ($k =$) plus proches voisins. Ensuite, pour chaque lien, on le supprime et on le reconnecte vers un nouveau noeud. Cette action se fait avec une probabilité p. Si $p = 0$, le réseau demeure régulier. Si $p = 1$, le résultat est un graphe aléatoire.

Pour caractériser l'effet petit-monde d'un graphe réel, on compare le coefficient de clustering et la moyenne des plus court chemin du graphe réel(G_r) et d'un graphe aléatoire équivalent (C_{G_a}). Si $C_{G_r} \gg C_{G_a}$ et $L_{G_a} \leq L_{G_r}$ alors ce graphe réel est un graphe petit-monde.

Comme le modèle Erdos-Rényi, la distribution des degrés du modèle de Watts-Strogatz est homogène.

3.10 Le modèle sans-échelle de Barabasi-Albert

Beaucoup de réseaux réels tels qu'internet, le WWW, les réseaux de citation contiennent quelques noeuds de forts degrés appelés hubs, et plusieurs noeuds de faibles degrés. En effet, dans ces réseaux, un nouveau noeud, par exemple une nouvelle page web, a tendance à se connecter aux noeuds de fortes degrés qu'aux noeuds de faibles degrés. Le modèle de Barabasi-Albert [Barabási, 1999] permet de générer ces types de réseaux en se basant sur deux critères : la croissance et l'attachement préférentiel.

La croissance signifie que ces réseaux s'agrandissent au fil du temps. L'attachement préférentiel indique que lorsqu'il y a un nouveau noeud, il est plus probable qu'il se lie aux noeuds les plus connectés qu'aux noeuds de faibles degrés. Ainsi, progressivement, quelques noeuds se transformeront en hubs. Le phénomène "les riches deviennent plus riches" se matérialise. La distribution des degrés de ces types de réseaux suivent une loi puissance ($P(k) = k^{-\lambda}$). Ces réseaux réels appartiennent à la catégorie des réseaux dits réseaux sans-échelle (scale-free).

L'algorithme de génération d'un modèle de Barabasi-Albert se déroule comme suit :

Initialement, on forme un graphe avec m_0 noeuds. Chaque noeud a au moins un lien. Le réseau s'agrandi comme suit :

Chaque nouveau noeud ajouté est lié à m ($m \leq m_0$) noeuds avec une probabilité proportionnelle aux nombres de liens existants. Le nouveau noeud est connecté à un noeud i avec la

probabilité p_i :

$$p_i = \frac{k_i}{\sum_j k_j} \tag{1.19}$$

Les nouveaux noeuds "préfèrent" les noeuds fortement liés.

A chaque étape t, le réseau a un nombre de noeud $N = m_0 + t$ et $L = m_0 + mt$ comme nombre de liens. Ce modèle génère un réseau dont la distribution suit une loi puissance avec $\lambda = 3$.

La nature sans-échelle des réseaux réels est un peu controversée. En effet, il est établi qu'il y a plusieurs réseaux réels dont la distribution des degrés ne suit pas une loi puissance [Broido, 2019]. Ainsi, pour mieux caractériser les réseaux réels, les auteurs de [Zhou, 2020] utilisent la distance entre les degrés. Ils définissent pour chaque lien :

$$\eta(i,j) = max\left\{k_i, k_j\right\} / min\left\{k_i, k_j\right\} \tag{1.20}$$

La loi puissance est mieux représentée par la distribution de la distance degré-dégradé que la distribution des degrés. Cette caractéristique est plus prononcée pour les réseaux denses. La principale limite du modèle de Barabasi-Albert est l'absence de structure communautaire.

4 Niveau mesoscopique

4.1 La structure communautaire

Les réseaux réels, malgré leurs différences, on a déjà vu qu'ils ont des propriétés et des structures similaires telles que la faible densité, l'effet petit-monde, la propriété sans-échelle. Une autre propriété commune des réseaux réels est la structure communautaire. En effet, beaucoup de réseaux réels sont organisés en communautés [Fortunato, 2016a]. Par exemple, dans un média social, les individus partageant le même centre d'intérêt ont tendance à se connecter entre eux. Dans une entreprise, les gens qui sont dans un même service ou département interagissent plus entre eux. Dans le transport aérien, les aéroports qui sont dans un même espace géographiques ont beaucoup de trafic commun. Des pages web peuvent traiter des sujets qui ont la même thématique, des protéines peuvent avoir la même fonction biologique.

La définition d'une structure communautaire n'est pas unanime. Il y a des définitions basées sur densité, la similarité, la diffusion, etc. Cependant, la définition basée sur la densité est le plus utilisée. On dit qu'un réseau a une structure communautaire s'il a des groupes de nœuds denses faiblement liées. Cette définition repose sur deux aspects : la cohésion et la séparation. Les communautés ont une forte cohésion, c'est-à-dire que les communautés ont beaucoup de liens internes. De même, les communautés ont une forte séparation, c'est-à-dire que les communautés sont connectées entre eux par peu de liens.

La structure communautaire est une structure mésoscopique très importante. Elle permet une d'avoir meilleure compréhension des réseaux, de mieux visualiser les réseaux, de faire des calculs parallèles, etc. Ainsi, beaucoup d'algorithmes de détection de communautés ont été mis au point.

La détection de communauté dans les réseaux est un champ très actif dans le domaine des réseaux complexes. Comme il y a plusieurs définitions d'une communauté, il existe différentes approches pour extraire les communautés. Il y a des algorithmes qui se basent sur la modularité, sur les chemins aléatoires, sur la propagation de label, sur la diffusion d'information, etc [Coscia, 2011].

4.1.1 Algorithmes de détection de communautés

Louvain [Blondel, 2008] est un algorithme basé sur la modularité. Son principe fondamental est de maximiser la modularité à chaque appel de la procédure. Deux étapes sont répétées de manière itérative. Initialement, chaque nœud représente une communauté. Ensuite, on calcule le gain en modularité obtenu en déplaçant chaque nœud dans la communauté de ses voisins. Le processus s'arrête lorsque aucune amélioration de modularité n'est possible. La deuxième étape consiste à créer un nouveau réseau à partir des communautés identifiées lors de la première étape.

LPA [Raghavan, 2007] est basé sur la méthode de propagation des étiquettes. Elle fonctionne comme suit. À l'initialisation, on attribue une étiquette au hasard à chaque nœud. Les étiquettes changent à chaque itération. En effet, les nœuds adoptent l'étiquette la plus commune de leurs voisins jusqu'à ce que le changement d'étiquette ne soit plus possible. Les groupes de nœuds partageant la même étiquette forment les communautés découvertes par LPA.

Walktrap [Pons, 2005] extrait des communautés en se basant sur les marches aléatoires. Une marche aléatoire est un processus constitué d'une séquence d'étapes aléatoires. Dans un graphe, à partir d'un nœud, on répète un processus dans lequel un marcheur se déplace aléatoirement et uniformément vers les nœuds voisins. Le principe de l'algorithme est qu'un marcheur qui se trouve dans une communauté la quitte difficilement. Les auteurs introduisent une distance entre deux nœuds basée sur la marche aléatoire pour partitionner un réseau en communautés. Ainsi, deux nœuds appartiennent à une même communauté si leur distance est petite. Dans le cas contraire, leur distance est grande. La structure communautaire est formée en fusionnant deux à deux les communautés.

Infomap [Rosvall, 2008] combine les techniques de théorie de l'information et de marches aléatoires. Les auteurs considèrent que dans une communauté, l'information circule rapidement et facilement à cause de la forte connexion. Décrire le flux d'information correspond à un problème de codage. L'objectif est de trouver un code efficace à partir d'une partition de manière à minimiser la longueur de description estimée d'une marche aléatoire. Pour cela, un nom est donné à chaque module dans lequel un nœud est nommé en utilisant le codage de Huffman[Huffman, 1952]. La meilleure partition du réseau donnera la description la plus courte pour toutes les marches.

4.1.2 Évaluation

Modularité La modularité, introduit par Girvan et Newman [Girvan, 2002] est une mesure qui est très utilisée pour évaluer la qualité d'une structure communautaire. Le principe se base sur la définition courante d'une communauté. Un bon partitionnement en communautés signifie qu'il y a plus de lien inter-communautaires que de liens intra-communautaires. Ainsi, on compare la structure communautaire réelle avec un modèle nul sans structure communautaire. Ses valeurs sont comprises entre -1 et 1. Une modularité est négative ou nulle si le modèle aléatoire est meilleur que les communautés réelles. On dit qu'un réseau à une bonne partition en communauté si la modularité est supérieure à 0.3. Plus la modularité est élevée, plus la structure communautaire est significative. Entre deux partitions, celle qui a la plus grande modularité est considérée comme la meilleure. La modularité est définie comme suit :

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{1.21}$$

A_{ij} est la matrice d'adjacence entre i et j ; k_i est la force du nœud i ; c_i est la communauté à laquelle appartient i ; le delta de Kronecker $\delta(a, b)$ vaut 1 si $a = b$ et 0 sinon ; et $m = \frac{1}{2} \sum_{ij} A_{ij}$.

Cette définition n'utilise que la connectivité du réseau. C'est l'une des avantages majeurs de cette mesure. Elle utilise peu d'information. Plusieurs algorithmes de détection de communauté se basent sur la modularité. Ces algorithmes cherchent à optimiser la modularité. L'optimisation de la modularité a des limites. En effet, il est difficile de détecter les petites communautés dans les grands réseaux [Karrer, 2008 ; Brandes, 2007]. Deux partitions différentes peuvent avoir une même modularité. De plus, il arrive de trouver une grande modularité pour des graphes aléatoires[Montgolfier, 2011]. La modularité est étendue pour les réseaux pondérés [Blondel, 2008]. Il suffit de considérer les poids dans la matrice d'adjacence.

Paramètre de mixage (Mixing Parameter) Pour un réseau pondéré, le paramètre de mixage μ^w d'un nœud i est la proportion de poids pointant hors de sa communauté. Le paramètre de mélange du réseau est la moyenne des paramètres de mélange des nœuds. Les communautés sont bien séparées lorsque le paramètre de mélange moyen est proche de 0. Plus le paramètre de mélange est petit, plus la structure communautaire est forte. Le paramètre de mélange d'un réseau est défini comme suit :

$$\mu^w = \frac{1}{n} \sum \mu_i^w \tag{1.22}$$

où $\mu_i^w = \frac{w^{ext}}{w_i}$, w^{ext} est la somme des poids du nœud est i aux communautés qui ne contiennent pas i, et w_i est la force du nœud i.

Information mutuelle normalisée L'information mutuelle normalisée mesure la similarité de deux partitions. Les partitions sont similaires lorsque le NMI est proche de un et presque indépendantes lorsque proche de 0. Le NMI de deux partitions P_1 et P_2 est défini comme suit :

$$NMI(P_1, P_2) = 2 * \frac{I(P_1, P_2)}{H(P_1) + H(P_2)} \tag{1.23}$$

$I(P_1, P_2)$ est l'information mutuelle entre P_1 et P_2. $H(P)$ est l'entropie de P. Ici, la somme d'entropie normalise l'information mutuelle.

4.2 La structure noyau-périphérie

La structure noyau-périphérie est une structure mésoscopique qui intéresse de plus en plus les chercheurs. La structure noyau-périphérie considère qu'il existe deux ensembles de nœuds, le noyau et la périphérie. Les nœuds du cœur forment un groupe de nœuds densément connectés. En revanche, les nœuds de la périphérie sont très épars. De plus, les liens entre les nœuds du cœur et de la périphérie sont denses, mais pas aussi denses que les liens du cœur : [Borgatti, 2000]. La structure cœur-périphérie est présente dans de nombreux réseaux complexes tels que le réseau mondial de transport aérien, les réseaux sociaux, les réseaux financiers [Boyd, 2006 ; Zhang, 2015 ; Cucuringu, 2016 ; Xiang, 2018 ; Jeude, 2019].

Ensuite, deux méthodes sont élaborées. Le modèle discret dans lequel un nœud appartient soit au cœur, soit à la périphérie. Le modèle continu dans lequel un nœud a une valeur de cordialité. A partir de cette valeur, un noeud peut appartenir au noyau, à la semi-périphérie ou à la périphérie. Il existe deux types de noyau-périphérie : [Lee, 2014]. Le noyau-périphérie basé sur la densité stipule que les nœuds de degré au sein du groupe central sont élevés. Le noyau-périphérie basé sur le transport, qui se base sur le plus court chemin. Cet article n'exclut pas la présence de noyaux-périphéries multiples, qui a également intéressé les chercheurs [Rombach, 2017 ; Kojaku, 2017a].

Différentes approches sont appliquées pour vérifier une structure noyau-périphérie : [Holme, 2005 ; Da Silva, 2008]. Et plusieurs algorithmes sont proposés pour détecter ce type de structure. Ensuite, la structure cœur-périphérie est présente dans de nombreux réseaux complexes [Boyd, 2006 ; Zhang, 2015 ; Cucuringu, 2016 ; Xiang, 2018 ; Jeude, 2019].

4.3 La structure hiérarchique

Depuis plusieurs années, des travaux sont menés sur la structure hiérarchique d'un système complexe. La définition de la structure hiérarchique peut être ambiguë. En effet, les algorithmes permettant de détecter les communautés qui fonctionnent de manière hiérarchique [Blondel, 2008]. On peut définir l'organisation hiérarchique d'un réseau si celui-ci possède des communautés et à l'intérieur de ces communautés, il y a des communautés et ainsi de suite. La hiérarchie est représentée par un dendrogramme ou un arbre, avec le réseau entier en tête. Des modèles sont construits, et des méthodes sont proposées pour détecter de telles organisations. En effet, plusieurs réseaux réels présentent une structure hiérarchique [Ravasz, 2003] selon le critère que le coefficient de regroupement des nœuds est une échelle du degré d'un nœud, $C(k) \sim k^{-1}$. Dans [Sales-Pardo, 2007] un modèle de structure hiérarchique est développé pour évaluer si un réseau a une telle organisation. En s'appuyant sur une nouvelle mesure d'affinité et sur le paysage de modularité, cette méthode extrait l'organisation hiérarchique hétérogène du réseau. L'algorithme révèle la structure hiérarchique de certains réseaux réels, notamment celle du réseau mondial de transport aérien, de haut en bas.

4.4 Le k-core

Le k-core d'un réseau est un sous-réseau dans lequel un nœud a au moins k voisins. On peut extraire le k-core maximum en retirant itérativement des nœuds du réseau. En effet, le 1-core, où chaque noeud a au moins un lien, est le réseau entier. On forme le 2-core en enlevant tous les nœuds ayant une connexion dans le réseau 1-core. Puis, on extrait le 3-core en supprimant tous les noeuds ayant deux liens dans le 2-core, et ainsi de suite jusqu'à atteindre le nombre max k-core pour lequel il est impossible d'obtenir le (k+1)-core.

5 Niveau microscopique

5.1 Centralités

Trouver les nœuds les plus influents dans un réseau est sujet qui intéresse beaucoup de chercheurs dans le domaine des réseaux complexes. La détermination des nœuds stratégiques est primordiale dans beaucoup d'applications. Ces applications couvrent des applications telles que les stratégies de vaccination, les propagations d'épidémie, la robustesse des systèmes complexes, la recherche d'acteurs clés dans un réseau terroriste, la diffusion de l'information, etc.

Les centralités permettent de trouver les nœuds vitaux dans un réseau. Plusieurs mesures de centralités ont été proposées jusqu'à ce jour. Elles peuvent être classées en trois groupes [Lü, 2016]. le premier groupe concerne les centralités qui sont basées sur le voisinage et ont une portée locale. Il y a en qui sont basées sur le chemin et qui ont une portée globale. Le troisième groupe utilise un processus de réajustement itératif.

5.1.1 Centralité de degré

Le degré d'un nœud est le nombre de ses voisins de premier ordre. Étant donné un graphe $G(V, E)$, tel que V est l'ensemble des nœuds et E l'ensemble des liens, le degré k_i du noeud i est défini comme suit :

$$k_i = \sum_{j \in V, i \neq j} a_{ij}$$

a_{ij} est un élément de la matrice d'adjacence binaire de G tel que $a_{ij} = 1$ si i and j sont connectés, sinon, $a_{ij} = 0$.

5.1.2 Centralité de force pondérée

La centralité de force [Lü, 2016] d'un nœud est le cumul des poids de ses voisins de premier ordre. Étant donné un graphe $G(V, E, W)$, tel que V est le nombre de nœuds, E le nombre de liens, et W est l'ensemble des poids, la Force s_i du nœud i est définie comme :

$$s_i = \sum_{j \in V, i \neq j} w_{ij} \tag{1.24}$$

w_{ij} est un élément de la matrice d'adjacence de G tel que $w_{ij} = \alpha$ si i et j sont connectés, sinon, $w_{ij} = 0$, $\alpha \in \mathbb{R}_+^*$.

5.1.3 Centralité d'intermédiaire

La centralité d'intermédiaire d'un nœud est le pourcentage de plus court chemin qui passe par ce nœud. L'information transmise entre deux noeuds a une forte probabilité de passer par les noeuds qui ont une large valeur de centralité d'intermédiaire. La centralité d'intermédiaire d'un nœud i est définie comme suit :

$$b(i) = \sum_{i \neq j} \frac{\sigma_{jk}(i)}{\sigma_{jk}}$$

σ_{jk} est le nombre de plus courts chemins entre j et k. $\sigma_{jk}(i)$ est le nombre de plus courts chemins de j à k passant par i.

5.1.4 Centralité d'interdépendance pondérée

L'Entrelacement [Brandes, 2001] $b(i)$ d'un noeud i est la fraction du plus court chemin passant par lui. Lorsqu'elle est normalisée, l'Entrelacement du noeud i est défini comme suit :

$$b(i) = \sum_{i \neq j} \frac{\sigma_{jk}(i)^w}{\sigma_{jk}^w} \tag{1.25}$$

σ_{jk}^w est le nombre du plus court chemin entre j and k. $\sigma_{jk}(i)^w$ nombre du plus court chemin de j à k passant par i.

5.1.5 Centralité du PageRank

Initialement introduit pour classer les sites web, PageRank [Ding, 2011] est une mesure de raffinement itérative basée sur un processus de marche aléatoire. Bien que définie pour les réseaux dirigés, on peut utiliser deux liens dirigés pour les réseaux non dirigés. Pour un réseau pondéré, il est défini comme suit :

$$wpr_i(t) = \sum_{j=1}^{n} w_{ji} \frac{wpr_j(t-1)}{s_j^{out}} \tag{1.26}$$

s_j^{out} est la force de sortie du nœud j.

5.1.6 Centralité de proximité

Proximité pondérée : La centralité de proximité quantifie la distance entre un nœud et tous les autres nœuds d'un réseau. Plus un nœud est proche des autres nœuds, plus il est central. La centralité de proximité c_i du nœud i est définie comme suit :

$$c(i) = \sum_{j}^{n} \frac{1}{d_{ij}}$$

d_{ij}^{w} est la distance entre i et j.

5.1.7 Centralité de vecteur propre

La centralité du vecteur propre X d'un nœud i dépend de l'influence de ses voisins. Il s'agit d'une mesure globale, car l'influence des voisins dépend aussi de leurs voisins et ainsi de suite. Elle est définie par la formule suivante, dans laquelle le i-ième élément du vecteur X représente l'influence du nœud i.

$$AX = \lambda X$$

A est la matrice d'adjacence de G. λ est la valeur propre de A. Cette centralité peut être étendue aux réseaux pondérés en considérant la matrice d'adjacence A comme une matrice pondérée.

6 La robustesse

Avec la crise de la pandémie à la COVID19, nous avons pû constater de nombreuses perturbations sur le réseau mondial de transport aérien. Ainsi, l'étude de sa résilience, à laquelle sont consacrés les chapitres 5, 6 et 7, revêt donc un caractère important.

6.1 Les attaques

On peut considérer deux types d'attaques dans un réseau : les attaques aléatoires et les attaques ciblées. Une attaque aléatoire supprime des nœuds de façon aléatoire. Elle permet d'étudier la robustesse du réseau lorsqu'une défaillance inattendue se produit. En revanche, les attaques ciblées visent à supprimer les nœuds les plus vitaux pour la connectivité du réseau. [Wan, 2021 ; Wandelt, 2018a]. Les mesures de centralité décrivent généralement l'importance des nœuds [Lü, 2016 ; Ghalmane, 2019a ; Rajeh, 2020]. Dans une stratégie d'attaque forte, on supprime les nœuds du réseau par ordre décroissant de magnitude d'une mesure de centralité. Les attaques faibles utilisent l'ordre inverse.

6.2 Évaluations

La taille de la composante géante est la métrique la plus populaire pour évaluer la robustesse d'un réseau. Elle fait référence à la taille du plus grand ensemble de nœuds interconnectés lorsque le réseau se divise en plusieurs parties en raison de la suppression de nœuds. L'attaque la plus efficace est celle qui réduit le LCC à zéro avec le plus petit nombre d'aéroports ciblés. Le réseau est résilient si le LCC diminue lentement en fonction des nœuds supprimés.

L'efficacité [Crucitti, 2003] évalue l'échange d'informations entre les nœuds d'un réseau. Zhou et al. introduisent l'efficacité pondérée dans [Zhou, 2019] pour prendre en compte l'influence des poids des liens. Contrairement à l'efficacité non pondérée, l'efficacité pondérée peut être supérieure à 1. Elle est définie comme suit :

$$WE(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in V} \frac{1}{\sum_{l \in L_{ij}} \frac{1}{w_l}} \tag{1.27}$$

L_{ij} est l'ensemble des liens sur le plus court chemin entre i et j. w_l est le poids du lien l.

Pour évaluer la robustesse d'un réseau, cette mesure est recalculée après la suppression d'un nœud du réseau.

7 Contexte, contributions et organisation de la thèse

7.1 Contexte

Les structures mésoscopiques telles que la structure communautaire et la structure noyau-périphérie ont fait l'objet de beaucoup de travaux dans la science des réseaux. Ces deux structures sont présentes dans plusieurs réseaux réels. La structure communautaire et la structure multiple noyau-périphérie ont une caractéristique en commun qui est de mettre en évidence les parties denses d'un réseau. En effet, les communautés peuvent être définies comme les parties denses d'un réseau faiblement connectées. Dans une structure multiple noyau-périphérie, il y a plusieurs noyau-périphérie. Une structure noyau périphérie est composé d'un noyau et d'une périphérie. Le noyau regroupe les nœuds qui densément connectés, tandis que les nœuds de la périphérie sont peu connectés. Néanmoins, il y a une liaison assez forte entre les nœuds du noyau et de la périphérie. Une fois extraite, ces structures sont utilisées dans une multitude d'applications telles que la recherche des nœuds les plus importants, la visualisation, la propagation d'épidémie, la robustesse, etc. Néanmoins, les parties denses de ses structures ne sont pas exploitées dans la littérature des réseaux complexes.

Dans cette thèse, nous proposons, dans un premier temps, une nouvelle structure mésoscopique qui se base sur la structure communautaire ou sur la structure multiple noyau-périphérie. En effet, avec cette structure, il s'agit de détacher les parties denses et de garder les liaisons entre ces parties. Cette nouvelle structure offre de nouvelles perspectives pour explorer particulièrement le réseau mondial de transport aérien. Cette nouvelle structure a des applications dans l'étude des centralités, dans la resilience et la robustesse. Dans ce sens, et compte tenu de la sensibilité et de la criticité des réseaux, nous allons developper de nouvelles stratégies d'attaques en vue d'évaluer la robustesse du réseau de transport aérien.

7.2 Contributions

Les travaux de cette thèse ont abouti à plusieurs contributions :

— Nous introduisons une nouvelle structure mésoscopique appelée la structure en composante qui comprend les composantes locales et les composantes globales. Cette représentation permet une meilleure visualisation d'un réseau complexe. De plus, puisque chaque composante est un réseau, l'étude topologique de chaque composante apporte des informations plus fines sur la structure hétérogène des réseaux.

— La structure en composante présente de nombreuses applications. Au niveau des centralités, la structure en composante montre la pertinence qu'un nœud a une importance locale et une importance globale. Aussi, en étudiant la résilience d'un réseau, on peut évaluer en

même temps l'impact des attaques sur les composantes. De plus, avec la structure en composante, on peut établir plusieurs stratégies d'attaques.

— En prenant comme cas d'étude le réseau de transport aérien pondéré et non pondéré, cette thèse contribue de façon significative à sa compréhension. En effet, la composante locale du réseau mondial aérien capture les vols régionaux alors que la composante globale capture les vols inter-régionaux. Le réseau de transport aérien peut être vu comme un agrégat de réseaux régionaux et de réseaux inter-régionaux dont l'étude de leurs structures nous apportent plus d'information sur la structure du réseau mondial aérien. Ainsi, la structure hétérogène du réseau est mis en évidence.

— L'étude des centralités basée sur la structure en composante à montrer les manquements de la méthode classique. Nous avons montré qu'avec la méthode classique, beaucoup d'aéroports importants sont occultés. Avec la structure en composante, les aéroports qui ont une influence régionale et inter-régionale sont mis en évidence. En plus, on peut extraire plusieurs k-core avec la structure en composante.

— En faisant une étude comparative du réseau pondéré et non pondéré, nous avons trouvé que certains résultats changent radicalement. Avec le réseau aérien mondial non pondéré beaucoup aéroports les plus importants sont localisé en Europe. Alors qu'avec le réseau pondéré, beaucoup sont localisés en Amérique. Cependant, avec la structure en composante, les résultats sont plus cohérents et plus stables.

— L'étude de la robustesse du réseau à travers la structure en composante permet de comprendre l'interdépendance qu'il y a entre les différentes régions. En effet, l'attaque des aéroports les plus importants entraine l'isolement des régions. Ainsi, il y a certaines régions qui s'isolent presque toujours ensemble.

— La nouvelle stratégie d'attaque basée sur la structure en composante qu'on a mis au point est très efficace comparé aux attaques classiques. De plus, cette stratégie nécessite moins de temps de calcul.

7.3 Organisation de la thèse

Outre l'introduction générale et la conclusion générale, cette thèse est organisée de la manière suivante :

— Dans le Chapitre 3 qui est une contribution publiée dans [Diop, 2021], nous explorons la structure en composante du réseau mondial. Nous avons proposé en premier un algorithme d'extraction de la structure en composante. Ainsi, nous avons analysé les propriétés topologiques, la centralité de degré et le k-core du réseau mondial aérien en se basant sur la structure en composante.

— Dans le chapitre 4, nous avons fait une étude comparative de la structure en composante du réseau mondial pondéré par rapport au réseau non pondéré. En effet, nous avons comparé les propriétés qui sont étendues pour les réseaux pondérés pour comprendre l'influence du poids dans l'étude du réseau. Ces travaux sont soumis pour une publication dans un journal [Diop, 2022d].

— Dans les chapitres suivants, on s'intéresse à la robustesse du réseau mondial aérien. Dans ce chapitre 5 qui est une contribution publiée dans [Diop, 2022c], nous avons étudié l'impact des attaques classiques sur les composantes du réseau. Comme critère d'évaluation de la robustesse, nous avons utilisé uniquement la LCC.

— Ensuite, dans le chapitre 6, on utilise les attaques basées sur les centralités pondérées pour évaluer l'impact sur les composantes du réseau pondéré. En plus de la LCC, ce chapitre, nous avons ajouté l'efficacité pour évaluer la robustesse. Les resultats de ce chapitre ont fait l'objet deux publications dans [Diop, 2022a] et [Diop, 2022e].

— Dans le dernier chapitre 7, nous avons proposé une nouvelle stratégie d'attaque basée sur la composante globale pour étudier la robustesse du réseau non pondéré. Cette stratégie d'attaque est prometteuse puisqu'elle est efficace et moins couteuse en temps de calcul par rapport aux attaques classiques. Les résultats de ce chapitre sont publiés dans [Diop, 2022b].

Chapter 2

Context, contributions and organization of the manuscript

Summary

As the basic definitions of complex networks and the prerequisites given in the previous chapter are also taken up in detail in the following chapters, we will only limit ourselves here to introducing the context, main contributions, and organization of the thesis.

1 Thesis context

A complex network is a network that has an irregular architecture and topology, and that contains several parts or several elements combined in a way that is difficult to understand. It is, therefore, a network of interactions between entities whose global behavior cannot be deduced from the individual behaviors of said entities, hence the emergence of new properties. Like all networks, they are composed of nodes (or vertices or points) representing objects, interconnected by links (or edges or lines). These networks are abstract representations of relationships predominantly present in real life in various industrial, biological, financial, and technological systems. The study of complex networks has been the subject of great attention from the scientific community since the beginning of the 2001s and has proved useful in many fields such as physics, biology, telecommunications, computer science, sociology, and epidemiology. Complex networks are ubiquitous around us and have many applications in everyday life. We can cite the World Wide Web, the Internet, trophic networks (or food chains), metabolic networks, or transport networks. This great diversity of complex networks makes their classification according to their common properties difficult. We often distinguish three levels to examine a complex network : the macroscopic, microscopic, and mesoscopic levels. The macroscopic analysis characterizes the entire network topology through global measurements. Microscopic studies examine network properties at the level of nodes or links. Mesoscopic analysis concerns groups of nodes or links that share similar characteristics.

Mesoscopic structures such as the community and core-periphery structures have been the subject of much work in network science. Both structures are present in many real networks. The community structure and the multiple core-periphery structures have a common feature : highlighting the dense parts of a network. Indeed, communities can be defined as the dense parts of a weakly connected network. In multiple core-periphery structures, there are several core-peripheries. A core-periphery structure is composed of a core and a periphery. The core group nodes are densely connected, while the nodes in the periphery are loosely connected. Nevertheless, there is a fairly strong connection between the core and the periphery nodes. Once extracted, these structures are used in many applications, such as finding the most important nodes, visualization, epidemic propagation, robustness, etc. Nevertheless, the dense parts of these structures are not exploited in the complex network literature.

In this thesis, we propose, in the first step, a new mesoscopic structure based on the community structure or on the multiple core-periphery structures. Indeed, with this structure, it is a question of detaching the dense parts and keeping the connections between these parts. This new structure offers new perspectives to explore, especially the global air transport network. This new structure has applications in the study of centralities, resilience, and robustness. In this sense, and taking into account the sensitivity and criticality of the networks, we will develop new attack strategies to evaluate the robustness of the air transport network.

2 Main contributions

The work of this thesis has resulted in several contributions :

— We introduce a new mesoscopic structure called the component structure, which includes local and global components. This representation allows better visualization of a complex network. Moreover, since each component is a network, the topological study of each component provides more delicate information about the heterogeneous structure of networks.

— The component structure has many applications. At the hub level, the component structure shows a node's relevance in local and global importance. Also, studying a network's resilience can simultaneously assess the impact of attacks on the components. Furthermore, with the component structure, one can establish several attack strategies.

— By taking the weighted and unweighted air transport network as a case study, this thesis contributes significantly to its understanding. Indeed, the local component of the global air network captures regional flights, while the global component captures inter-regional flights. The air transport network can be viewed as an aggregate of regional and inter-regional networks whose study of their structures provides more information about the structure of the global air network. Thus, the heterogeneous structure of the network is highlighted.

— The study of centralities based on the component structure shows the shortcomings of the classical method. We have shown that many essential airports are hidden with the classical method. With the component structure, airports with regional and interregional influence are highlighted. In addition, several k-scores can be extracted with the component structure.

— By comparing the weighted and unweighted networks, we found that some results change dramatically. Many of Europe's most important airports are located with the unweighted global air network. While with the weighted network, many are located in America. However, the results are more consistent and stable with the component structure.

— The study of the robustness of the network through the component structure allows us to understand the interdependence between the different regions. Indeed, the attack on the most important airports leads to the isolation of the regions. Thus, there are certain regions that almost always isolate themselves together.

— The attack strategy based on the component structure we developed is very effective compared to conventional attacks. Moreover, this strategy requires less computation time.

3 Organization of the manuscript

Besides the general introduction and the conclusion, this thesis is organized as follows :

— In Chapter 3, which is a contribution published in [Diop, 2021], we explore the component structure of the global network. We first proposed an algorithm for extracting the component structure. Thus, we analyzed the topological properties, degree centrality and k-core of the aerial world network based on the component structure.

— In the chapter 4, we compared the component structure of the weighted global network versus the unweighted network. Indeed, we compared the properties that are extended for

weighted networks to understand the influence of weight in the study of the network. This work is submitted for publication in a journal [Diop, 2022d].

— In the following chapters, we focus on the robustness of the global air network. In this chapter 5 which is a contribution published in [Diop, 2022c], we studied the impact of classical attacks on the network components. As a robustness evaluation criterion, we used only the LCC.

— Then, in chapter 6, we use weighted centrality-based attacks to evaluate the impact on the weighted network components. In addition to LCC, we added efficiency to evaluate robustness in this chapter. The results of this chapter are published in [Diop, 2022a] and [Diop, 2022e].

— In the last chapter 7, we proposed a new attack strategy based on the global component to study the robustness of the unweighted network. This attack strategy is promising since it is efficient and less expensive in computation time than classical attacks. The result of this chapter is published in [Diop, 2022b]

Chapter 3

Revealing the component structure of the world air transportation network [1]

Summary

Air transportation plays an essential role in the global economy. Therefore, there is a great deal of work to understand better the complex network formed by the links between the origins and destinations of flights. Some investigations show that the world air transportation network exhibits a community and a core-periphery structure. Although precious, these representations do not distinguish the inter-regional (global) web of connections from the regional (local) one. Therefore, we propose a new mesoscopic model called the component structure that decomposes the network into local and global components. Local components are the dense areas of the network, and global components are the nodes and links bridging the local components. As a case study, we consider the unweighted and undirected world air transportation network. Experiments show that it contains seven large local components and multiple small ones spatially well-defined. Moreover, it has a main global component covering the world. We perform an extensive comparative analysis of the structure of the components. Results demonstrate the non-homogeneous nature of the world air transportation network. The local components structure highlights regional differences, and the global component organization captures the efficiency of inter-regional travel. Centrality analysis of the components allows distinguishing airports centered on regional destinations from those focused on inter-regional exchanges. Core analysis is more accurate in the components than in the whole network where Europe dominates, blurring the rest of the world. Besides the world air transportation network, this chapter demonstrates the potential of the component decomposition for modeling and analyzing the mesoscale structure of networks.

1. The results contained in this chapter are published in the paper :
[Diop, 2021] Issa Moussa Diop, Chantal Cherifi, Cherif Diallo, Hocine Cherifi. "Revealing the component structure of the world air transportation network". https ://doi.org/10.1007/s41109-021-00430-2, Applied Network Science, 2021, vol. 6, no 1, p. 1-50. Springer Nature, 2021.

1 Introduction

Air transport plays an essential role in the current context of globalization by reducing the distance between countries. Whether it is for the movement of millions of people or goods, thousands of flights are made per day, impacting the global economy and even public health [Colizza, 2006]. Subsequently, the spread of the COVID-19 pandemic is mainly due to the world air transportation network. This pandemic leads to the bankruptcy of several airlines and affects the tourism and trade industry. That is why researchers have been interested in the air transportation network for a long time to study its structure, dynamics, and robustness [Zanin, 2013].

One can consider three levels of analysis of a network (macroscopic, microscopic, mesoscopic). Macroscopic analysis characterizes the entire network topology through a set of global measures. Microscopic studies investigate the network properties at the node or link level. Mesoscopic analysis concerns groups of nodes or links sharing similar features. There are two popular mesoscopic structure models : the community structure and the core-periphery structure. Although there is no consensus on their definition, both are related to the non-homogeneous density observed in real-world networks. The common understanding is that communities are dense areas of the network sparsely connected [Fortunato, 2016a]. The core-periphery structure considers that a network comprises a dense, cohesive core and a sparse, unconnected periphery [Borgatti, 2000]. These mesoscopic structures are observed in most real-world networks. While they find numerous applications and explain a broad range of phenomena in networked systems, none of them is well-suited to disentangle the local interactions of nodes within their dense area from their global interactions with the other dense areas of the network. Inspired by the works reported in [Ghalmane, 2019a ; Guimera, 2005], we propose a new mesoscopic representation of a network, called the component structure. Indeed, the authors show that considering the interplay between intra-community links and inter-community links in a modular network allows defining effective centrality measures. While these works focus on computing centrality measures in modular networks, our proposition is more general. Our goal is to distinguish between the local and global influence of various groups of nodes or links. To this aim, the proposed model decomposes a network into local components and global components. The local components are isolated, dense parts of the network that can be uncovered using a community detection or a multicore detection technique. The global components are the subnetworks joining the local components. One extracts them easily, based on the links between local components. Although it also relies on dense areas of the network as the community or core-periphery structures, the component structure offers a complementary view of the network mesoscopic organization. Indeed, components are isolated networks that can be analyzed separately. Furthermore, as the overlapping community structure, it does not operate a partition of the network. Indeed, in overlapping communities, a node can belong to multiple communities. In the component structure, a node can belong to a local and a global component. It is the case of the nodes linked to the other groups.

Our work departs from recent studies focusing on robustness [Lordan, 2019], and multilayer modeling [Lordan, 2017 ; Dai, 2018] of the air transportation network. Our main concern is to

use the world air transportation network as a case study to evaluate the ability of the component structure representation to get a better understanding of the network mesoscopic organization. Therefore, we concentrate on the monolayer network representation of the world air transportation network. However, one may note that the component structure can be interpreted as a multilayer representation where the layers are the local components. Nodes and links connecting the various local components form the global components. The main contributions of this chapter are as follows :

1. We introduce an alternative mesoscopic network structure called the component structure where local components are dense subnetworks, and the global components account for their interactions.
2. We use the component structure to study the structure of the world air transportation network. The local components characterize its regional organization, and the subnetworks forming the global components represent their interactions.
3. The regional characterization of the world air transportation network is not based on geographical considerations, but it relies on the density of the interactions.
4. This representation allows us to distinguish the regional impact of an airport from its influence worldwide.
5. We perform an extensive topological analysis of the component highlighting the regional and inter-regional differences of the world air transportation network.

The rest of the chapter is organized as follows. Section 2 reports a review of related studies of the air transportation network. Section 3.2 introduces the definition of the component structure, and it gives an algorithm to uncover it. Section 4.2 describes the data used in the experiments and examines the network community structure. Section 5 reports the analysis of the local component structure and section 6 analyzes the global component structure. Comparisons with the whole world air transportation network are reported in section 7. Section 8 presents the results of a comparative analysis of the degree centrality of the components and the world air transportation network. Section 9 discusses the results of the core structure analysis. Finally, we conclude in section 9.

2 Litterature review

One can distinguish three levels of study of the air transportation network : worldwide, regional and national. Based on this classification we present some influential contributions. For more information the reader can refer to the following surveys [Rocha, 2017 ; Lordan, 2014c].

Several studies have been devoted to the properties of *the worldwide air transportation network*. In [Guimera, 2004] Guimera et al. conduct the first exhaustive analysis of the world air transportation network. Considering airports as nodes and direct connections as links, the authors show that the degree and betweenness centrality exhibit a power-law distribution. Their work also reveals that highly connected cities are not the ones with high betweenness centrality values. This feature contradicts the classical preferential attachment formation model with a

geographical distance constraint [Yook, 2002]. They propose a new model which generates nodes with large betweenness and small degree based on geopolitical considerations. Indeed, only a few airports in each country are permitted to connect to airports of other countries, regardless of the geographical distance.

In [Guimera, 2005], the community structure is investigated. Results show that communities correspond to geographical areas. Seven city roles are defined based on the proportion of intra-community links and extra-community links. Three roles distinguish the highly connected nodes. "Provincial hubs" connect cities in their community. The "Connector hubs" are in the majority linked to cities outside their community. Finally, "Kinless hubs" share their links uniformly with all the communities. In the four non-hub categories, the "Ultraperipherical" nodes have no inter-community links. "Peripherical" nodes share most of their links with nodes in their community. "Nonhub connector" nodes have many links with nodes outside their community, and "Nonhub kinless" nodes share their links uniformly with all the communities. The main contribution of these works is to highlight the importance of geopolitical considerations in the formation mechanism of the air transportation networks and the various role of cities.

Inspired by fractality analysis, in [Sun, 2017b] the authors consider six types of nodes from fine-grained to coarse-grained granularity : airport, city, spatial areas of 100 km diameter around hubs, spatial areas of 200 km diameter around hubs, sub-national territory, country. Links are direct connections between nodes. The network analysis shows that all networks are small-world and disassortative. Furthermore, the clustering coefficient increases and the average path length decreases as the aggregation level varies from fine to coarse. The community structures uncovered by Louvain are quite consistent. It contains about ten communities corresponding to different geographical boundaries.

[Cheung, 2020] explores the evolution of the world air transportation network during the period 2006 - 2016. In this weighted network, nodes are airports, and the total number of passengers per year weights the direct fly links. The authors propose a new metric called Global Airport Connectivity Index, measuring the importance of airports in global passenger movements. It combines degree, closeness, eigenvector centrality measures, flow betweenness, and an indicator of regional importance. Building on the work of [Guimera, 2005], they classify the airports into regional hubs or global hubs, depending on their embeddedness in their community and their Global Airport Connectivity Index. Results show that the average degree and the density increase over time. Furthermore, North America, Russia, and China focus on developing regional hubs, while West Europe and the Middle East concentrating on emerging global hubs.

Some studies focus on *the regional air transportation network*. In [Lordan, 2017], the authors investigate the European airport network where nodes are European cities, and links account for direct flights. It appears that the degree follows a two-regime power-law distribution and that the network possesses the small-world property. Using the k-core, they decompose the network into three layers : the core (max k-core), the periphery (k-core of degree one), and the bridges between the core and the periphery. The analysis shows that the core contains the global European cities. The leisure air travel origins and destinations compose the bridges. Finally, the local destinations constitute the periphery. A robustness study shows that the network is more

vulnerable to the isolation of a combination of core and bridge nodes.

In [Dai, 2018], the authors investigate the evolution of Southeast Asian air transportation over the period 1979 - 2019. Results indicate that the number of hubs increases in this scale-free network. Disassortative behavior increases with time due to a more pronounced hub-and-spoke configuration of small airports for better accessibility. Decomposition of the network into a core-bridge-periphery structure shows that the core comprises the regional capital cities, the most economically vibrant secondary cities, and tourist destinations. The periphery cities are in remote areas with declining connectivity. High volatility over time characterizes the bridge nodes. The number of connections and passengers increases mainly in the core layer and the bridge layer at the end of the 20th century.

Lordan et al. [Lordan, 2019] use the Official Aviation Guide (OAG) subdivision to partition the world air transportation network into seven global regions (Africa, Asia, Europe, Latin America, Middle East, North America, Southwest Pacific). In these networks nodes are cities and links represent direct connections. Analysis shows that these small-world networks exhibit a two-regime power-law degree distribution. Targeted attack experiments based on the network's decomposition into core, bridge, and peripheral nodes show that regional networks with a large core are more resilient than networks with a smaller core.

Much more works concern *the national air transportation network* of major countries (US, China, India, Brazil)[Wandelt, 2019]. We summarize the main findings of the following related studies where nodes are airports and links represent direct flights. The common characteristics of all these networks is that hey all share the small-world property and are disassortative. In [Guida, 2007] the authors investigate the Italian network during three non-overlapping periods (June 1, 2005, to May 31, 2006 - July 16 to August 14, 2005 - November 2005). Results are consistent across networks. Indeed, the degree distribution and the betweenness centrality follow a double Pareto law. Their findings also suggest that the networks exhibit a fractal structure. Furthermore, the clustering coefficients are comparable and lower than those observed in a corresponding random network. It also appears that some highly connected airports have a small betweenness centrality.

In [Bagler, 2008] the authors investigate the network of India. They consider an unweighted directed network and a network weighted by the number of flights by week. The unweighted network has a truncated power law degree distribution. It is disassortative with a clustering coefficient one order of magnitude higher than the corresponding random network. The weighted network presents a hierarchical structure. Its analysis shows that highly connected airports share almost all the traffic, forming high traffic corridors.

The Chinese network has been extensively studied [Du, 2017a ; Du, 2017b ; Yang, 2021]. In [Wang, 2011] the authors show that its structure diverges from other national networks. Indeed, the exponential is a better fit than the power-law for the degree distribution. The explanation lies in the influence of the three main metropolises (Beijing, Shanghai, and Guangzhou). The network is disassortative with highly connected cities surrounded by poorly connected cities with direct links. This phenomenon gets more pronounced as the degree increases. Indeed, small

airports in China tend to supply direct links to the top hubs bypassing the less developed regional ones.

Extensive research on the topology and the dynamics of the U.S. air transportation network have been performed [Jia, 2014 ; Xu, 2008]. In [Cheung, 2012] the authors analyze its evolution over the period 1991 - 2011, and the study reported in [Siozos-Rousoulis, 2021] concerns the period 2001 - 2016. Overall, one does not observe considerable changes in the topological properties of the networks. The number of airports and flight routes has increased according to user demand. The network exhibits a truncated power law degree distribution. It is highly disassortative, and this trend grows with time, suggesting a pronounced evolution towards a hub and spoke structure over time. Similarly, with the world air transportation network, some high betweenness centrality nodes such as Anchorage cannot be considered hubs. The clustering coefficient decreases over time, and the average shortest path increases. It is in line with hub and spoke organization where peripherical airports connect to hubs providing long-distance flights.

Several papers are devoted to the analysis of the Brazilian air transportation network [Rocha, 2009 ; Costa, 2018 ; Oliveira, 2020]. In Couto et al. [Couto, 2015], the authors consider three networks : the network of national flight, the network of international flight and the network with both type of flights. The network is scale-free. Six communities corresponding to geographical areas ("North", "Center/North", "Northeast", "Minas Gerals", "Southeast", "South/West") are discovered by Louvain. The network is not resilient to targeted attack. Viracopos and Guarulhos are the key airports in the national network and for international connections, and they have the largest values of degree centrality and betweenness centrality. In addition, the number of routes decreases while the number of passengers increases, causing a higher level of occupation of aircrafts.

The analysis of the Australian network [Hossain, 2017] shows that it is scale-free. Its clustering coefficient is higher than its random network version indicating a cohesive network where passengers can be easily rerouted. The average path length suggests that, on average, a passenger can reach every destination in 3 flights. Most of the traffic goes through an interconnected group of high-degree nodes surrounded by low-degree neighbors. Centrality analysis shows that the more connected nodes do not necessarily exhibit the largest betweenness and closeness centrality values.

This literature review's main findings are that the topological properties are pretty consistent across the various levels of studies. Indeed, the degree and betweenness centrality exhibit a heavy tail distribution. Networks are small world and disassortative, with most nodes poorly connected to a few highly connected nodes. Nevertheless, the preferential attachment mechanism is not sufficient to explain the formation of the networks. One needs to consider geographical constraints (borders) and political and economic issues to get a better understanding of the network's topology. The clustering coefficient is higher than in random networks, and the average path length allows to join any destination in few hops. Another critical finding concerns the mesoscopic properties of the world air transportation network. Some studies demonstrate that it exhibits a community and a core-periphery structure.

3 Component structure of a network

This section introduces the definition of the component structure and an algorithm to uncover it based on the community structure.

3.1 Definition

Community structure and core-periphery structure models assume that the density is not homogeneous in a network. Dense areas form either the communities or the core elements of the networks. These two mesoscopic representations share a different view of the remaining nodes or links. In the community structure approach, communities are supposed to be sparsely connected by inter-community links. In the core-periphery structure, peripherical nodes are poorly connected to each other and with core nodes. To define the component structure, we retain both approaches' common points of view, i.e., the network contains dense areas. Those dense areas are localized in the network. Indeed, the vast majority of nodes interact with nodes contained in their community or core. That is the reason why we call them local components. Indeed, they share information with the rest of the network through a set of proxy links and nodes that have a more global view of their environment. These subnetworks tie together the local components. Consequently, the definition of the component structure is quite simple. A network contains two sets of subnetworks : 1) The dense parts of the network form the local components 2) Nodes and links shared by any two local components form the global components. Note that once the dense areas are extracted, global components identification is straightforward. Furthermore, one can exploit the various definition of dense areas proposed either in the community detection literature or the multi-core-periphery studies to extract the local components. In the following, we propose an algorithm that uses the community structure approach.

3.2 Component structure detection algorithm

Building on the work of Ghalmane et Al [Ghalmane, 2019a], we propose a component structure extraction algorithm exploiting the community structure. Remember that a network is decomposed into local components and global components. The local components are isolated dense parts of the network. The global components are subnetworks joining the local components.

The algorithm to uncover the component structure of a network proceeds as follows :

1. Uncover the dense part of the network : Use a community detection algorithm to uncover the community structure.
2. Extract the local components : Remove the inter-community links from the community structure to form the local components.
3. Extract the global components : Remove the intra-community links from the community structure, and the subsequent isolated nodes.

Note that this representation is redundant. Indeed, a node can belong simultaneously to a local component and to a global component. Such nodes at the frontier of the communities are

important locally and globally. Figure 5.1 illustrates the decomposition process of a network into

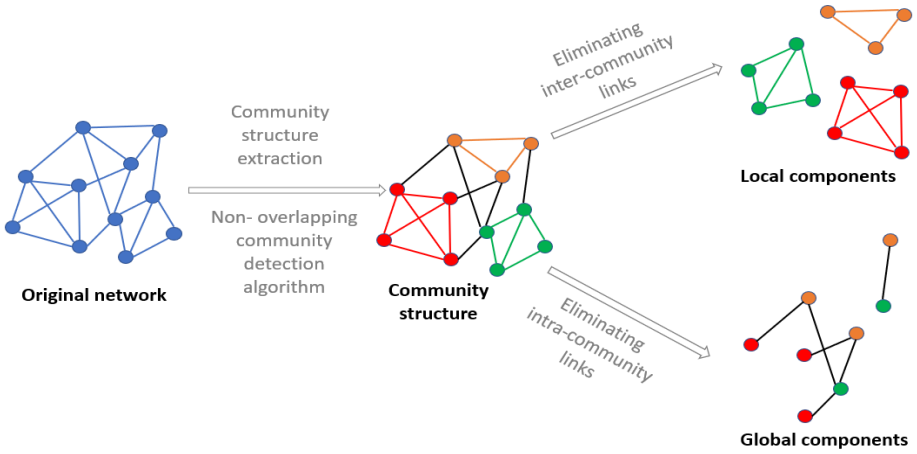


FIGURE 3.1 – The component structure detection algorithm. First, the community structure of the network is uncovered. Second, the local components are obtained by eliminating inter-community links from the community structure.Third, the global components are obtained by eliminating intra-community links and subsequent isolated nodes from the community structure.

its components on a toy example. First, one uses a community detection algorithm to partition the network into a set of non-overlapping communities. Inter-community links joining nodes in different communities are black. Nodes and intra-community links that bind nodes in the same community share the same color. We observe three communities respectively colored in red, yellow, and green. Removing the inter-community links allows us to isolate the three local components. Each community forms a local component, and it carries only local information. One obtains the global components removing the colored links from the community structure (intra-community links) and the subsequently isolated nodes. Global components act as bridges between the dense areas of the network. They are important actors in the information diffusion between the various dense parts of the network. As components are isolated sub-networks of the initial network, one can proceed to any type of topological analysis.

4 Uncovering the dense parts of the world air transportation network

In this section, we present the data set used in the experiments. We perform a comparative analysis of the community structure uncovered by two popular community detection algorithms used to extract the dense parts of the network.

4.1 Data

Information on world flights has been collected from FlightAware [Flightaware,]. The data covers six days (between May 17, 2018, and May 22, 2018), ensuring the inclusion of less frequent connections [Alves, 2020]. Nodes represent airports, and links represent direct flights between

two airports. For the sake of simplicity, the network is undirected and unweighted. However, one can consider weighted and directed networks using the appropriate analysis tools. The network contains 2734 nodes and 16665 links. Table 3.1 reports its basic topological properties.

TABLE 3.1 – Basic topological properties of the world air transportation network. N is the network size. $|E|$ is the number of edges. d is the diameter. L is the average shortest path length. μ is the density. ζ is the transitivity also called global clustering coefficient. λ is the assortativity also called degree correlation coefficient. η is the hub dominance.

N	$\|E\|$	d	L	μ	ζ	λ	η
2734	16665	12	3,86	0,004	0,26	-0,05	0.09

4.2 Community detection

The first step of the component structure detection algorithm consists in uncovering the dense part of the network using a community detection algorithm. Community detection is a very active field of research. Classically, the goal is to build a network partition into well-separated groups of nodes densely connected. In some situations, especially in social networks applications, nodes can belong to several groups. Therefore, overlapping groups are also investigated. Numerous algorithms inspired by different paradigms have been proposed so far. They are based on different methods such as modularity optimization, random walk or label propagation, among others. For complete coverage on this fundamental issue, one may refer to surveys reporting comparative studies, challenges, and open problems [Orman, 2012 ; Fortunato, 2016b ; Javed, 2018 ; Papadopoulos, 2012 ; Chunaev, 2020 ; Cherifi, 2018 ; Cherifi, 2019a].

In our experiments, we use two influential non-overlapping community detection algorithms, Louvain [Blondel, 2008] and Label propagation (LPA) [Raghavan, 2007]. The goal is to compare the uncovered community structures. Indeed, the community structure can be represented by various partitions with competitive quality score [Kirkley, 2021].

Louvain is a modularity-based algorithm. Its fundamental principle is to maximize the modularity at each call of the procedure. Two steps are repeated iteratively. Initially, each node represents a community. Then, one computes the gain in modularity obtained by moving each node in the community of its neighbors. The process stops when no modularity improvement is possible. The second step consists of creating a new network from the communities identified in the first step.

LPA is based on the label propagation method. It works as follows. At the initialization, assign a label at random to each node. Labels change at each iteration. Indeed, the nodes adopt the most common label of their neighbors until label change is no more possible. Groups of nodes sharing the same label form the communities uncovered by LPA.

4.3 Community structure analysis

We compare the community structures using two quality score functions : modularity and mixing parameter [Jebabli, 2018]. Modularity compares the proportion of intra-community links

with the ratio of expected links in a random network model. Typically, one considers that the community structure with the highest modularity corresponds to the best partition. The mixing parameter is the fraction of inter-community links. It reflects a strong community when it is near zero. The community structure with the smallest mixing parameter is considered the best one. Table 4.5 reports the results related to the quality metrics of the community structures uncovered by the two algorithms. The modularity values show that both community structures are well-defined. Moreover, it reveals that the community structure uncovered by Louvain is of better quality. The mixing parameter values indicate a low proportion of inter-community links. It corroborates the strong community structure measured by modularity.

Although both algorithms uncover well-separated and dense communities with few inter-community links, the community structures differ. Indeed, Louvain reveals 27 communities, while the community structure identified by LPA contains 130 communities. This result is not unusual. A wide range of different community structures can lead to community divisions with comparable quality scores. Even though these partitions differ, they are consistent. They relate through a set of "building blocks" that usually appear together in the same community [Riolo, 2020]. The Normalized Mutual Information (NMI) measures the information shared by two partitions. It is widely used to compare community structures [Orman, 2012]. When the NMI is equal to one, the two partitions are identical. When it is null, the two partitions are independent. Its high value reported in Table 4.5 confirms that both community structures have a lot in common.

Table 3.2 – Quality metrics of the community structures uncovered by Louvain and LPA community detection algorithms : Modularity, Mixing parameter, NMI

	Modularity	Mixing parameter	NMI
Louvain	0,63	0,13	0,77
LPA	0,59	0,14	

One can classify the communities uncovered by the community detection algorithms into two types. The first type corresponds to large size communities covering large geographical areas. Typically, these communities include several countries and contain hundreds of airports. The second type of community covers small geographical areas (within a country or fewer countries) and has less than one hundred airports. Louvain identifies seven large communities, and LPA discovers four large communities. Among these communities, three are of comparable size and cover the same geographical areas. Figure 3.2 represents the airports in these communities with the same colors. They are located respectively in North and Central America-Caribbean, East and Southeast Asia, and Russia-Central Asia-Transcaucasia. The fourth large community uncovered by LPA regroup into a single community, airports belonging to different communities uncovered by Louvain. Figure 3.3 illustrates this behavior. Indeed, parts of European (brown color) and Africa-Middle East-Southern Asia (lavender color) communities uncovered by Louvain merge in the same community (brown color) with LPA. Finally, the three other large communities uncovered by Louvain (Africa-Middle East India, South America, Oceania) split into multiple communities covering smaller geographical areas with LPA. Figure 3.4 illustrates this behavior.

For example, LPA divides into 28 communities the South America (orange color) community of Louvain.

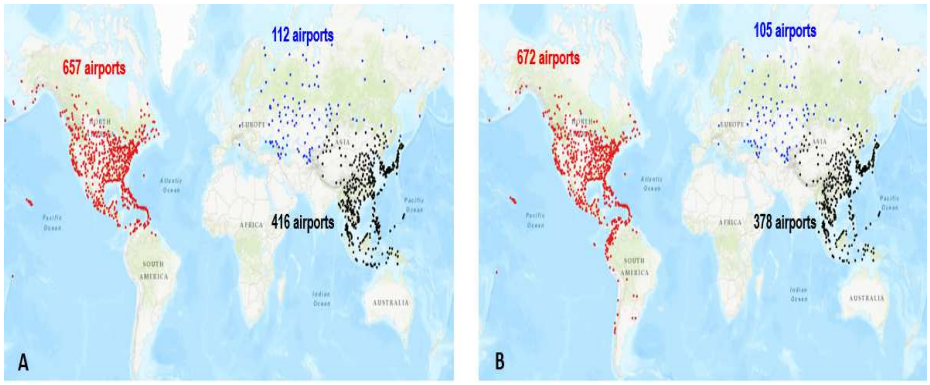


FIGURE 3.2 – The community detection algorithms uncover similar communities. (A) Louvain reveals 3 communities in North and Central America-Caribbean (airports in red), East-Southeast Asia (airport in black), and in Russia-Central Asia-Transcaucasia (airports in blue). (B) LPA finds similar communities in these areas.



FIGURE 3.3 – Communities uncovered by Louvain are grouped in a single community by LPA. (A) Louvain reveals two communities in Europe (airports in brown) and Africa-Middle East-Southern Asia (airports in lavender). (B) LPA finds a unique community covering the same geographical area.

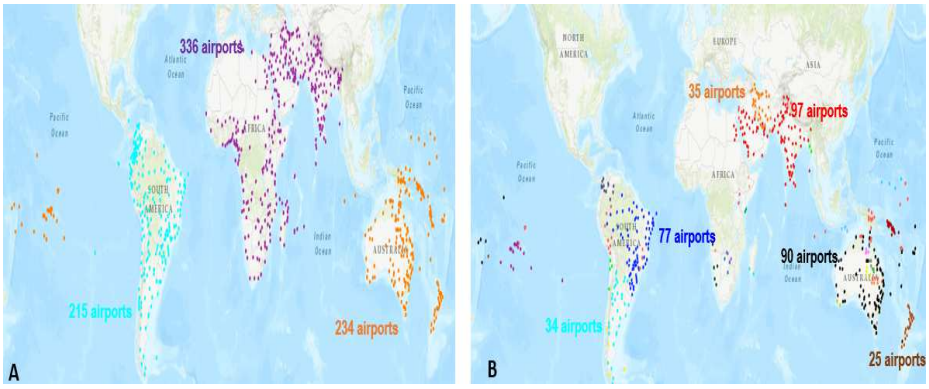


FIGURE 3.4 – Large communities uncovered by Louvain split into multiple communities by LPA. (A) Louvain reveals 3 communities. The indigo points represent the airports located in Africa-Middle East-Southern Asia. The orange points represent the airports located in Oceania. The cyan points represent the airports in South America. (B) LPA uncovers 57 communities in the same geographical areas. There are 16 communities in Africa-Middle East-Southern Asia, 28 in Oceania, and 13 in South America.

5 Local component structure of the world air transportation network

This section reports a comparative analysis of the macroscopic topological properties of the local components. To this end, we use the community structure uncovered by the Louvain algorithm to extract the components. Louvain uncovers 27 communities. Consequently, there are 27 local components. One can classify them according to their size into two categories. The seven large local components contain more than 100 nodes, and the size of the 20 small local components ranges from 2 to 60. Large components cover ample geographical areas, including several countries, while small components are localized in a single country.

Numerous schemes exist for partitioning the world into regions. The OAG divides it into seven regions : Europe, North America, Latin America, Africa, Asia, Middle East, Oceania. The subdivision corresponding to the local components is slightly different (see Figure 4.2). Indeed, the large components cover the following regions : North-Central America-Caribbean (657 airports), Europe (493 airports), East-Southeast Asia (416 airports), Africa-Middle East-Southern Asia (336 airports), Oceania (234 airports), South America (215 airports) and Russia-Central Asia-Transcaucasia (112 airports). Altogether, they include 90% of airports in the world. Note that the component division does not reflect strict geographical divisions. It instead corresponds more to the political, cultural, historical, and economic divide. For example, some African airports located in Morocco and Tunisia belong to the European component because of the solid economic and historical ties these countries share with Europe.

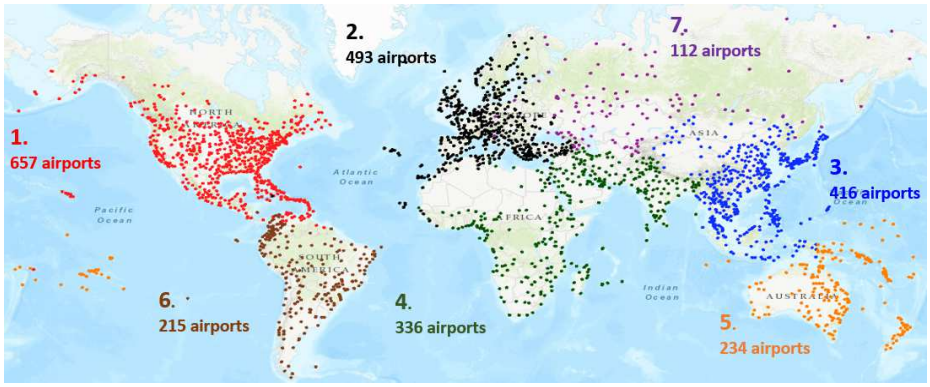


FIGURE 3.5 – The seven large local components. 1) Airports in the North and Central America-Caribbean are in red. 2) Airports in the European component are in black. 3) Airports in the East and Southeast Asia component are in blue.4) Airports in the Africa-Middle East and India component are in green. 5) Airports in the Oceania component are in orange. 6) Airports in the South America component are in brown. 7) Airports in the Russia-Central-Asia-Transcaucasia component are in indigo.

5.1 Analysis of the large local components

5.1.1 Basic macroscopic topological properties

The bar plots of Figure 3.6 report the values of basic topological properties (diameter, average shortest path, density, transitivity, assortativity, hub-dominance) used to compare the structure of the large local components. The *shortest path length* between node i and node j is the minimal number of links between those two nodes. The *diameter* is the largest shortest path between any two nodes of the network. It informs us about the longest route. Results show that between six or seven flights are needed to reach the most distant airports in the large local components. It is not the case in the Oceania and Russia-Central Asia-Transcaucasia components. Indeed, in the Oceania component, a maximum of nine flights are required. Whereas in the Russia-Central Asia-Transcaucasia component, one reaches the most distant airports in four flights. Consequently, if we consider this worst-case scenario, it is better to travel in Russia-Central Asia-Transcaucasia, and one should avoid Oceania.

The *average shortest path length* indicates how many flights are needed on average to reach a destination. North- Central America - Caribbean and East-Southeast Asia components are comparable with around 2.8 flights. It is easier to travel in Europe with its average shortest path of 2,58. Destinations are easy to reach in the Russia-Central Asia-Transcaucasia component. Indeed, one needs a little bit more than two flights on average to reach his goal. One needs more flights to reach his destination on average in Oceania, Africa, Africa-Middle East-Southern Asia, South America. Indeed, the average shortest path ranges from 3,15 to 3,5 in these regions.

Let N the number of nodes and $|E|$ the number of edges of a network. Its *density* μ measures the proportion of connected nodes as compared to the maximum number of possible connections. It is given by :

$$\mu = \frac{2 * |E|}{N(N-1)} \tag{3.1}$$

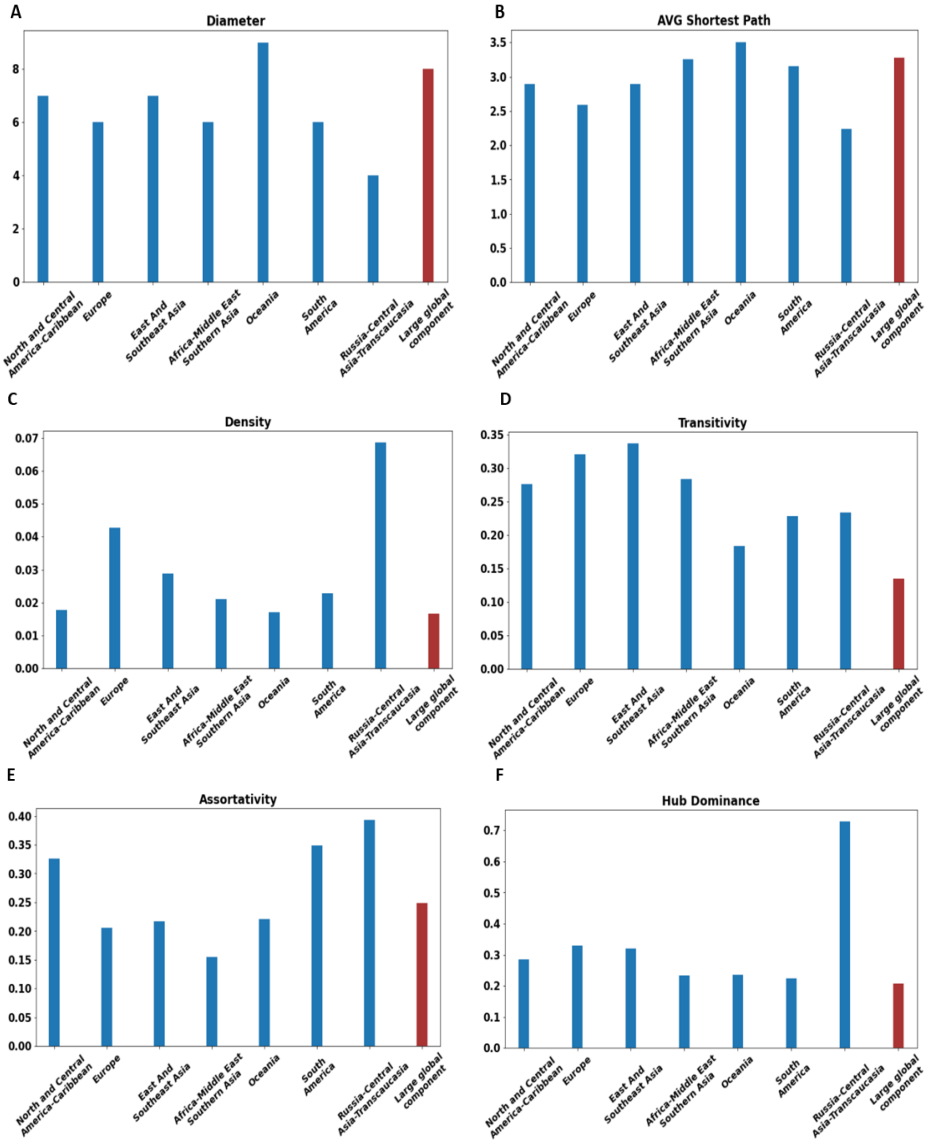


FIGURE 3.6 – Basic topological properties of the large components. Each bar plot concerns a different property (Diameter, Average shortest path, Density, Transitivity, Assortativity, and Hub dominance). The blue bar corresponds to the large local component. The red bar corresponds to the large global component. For assortativity, absolute values are plotted, but in reality, the values are negative.

Density is a marker of easiness to travel. Ideally, one would like to reach any destination with one flight. However, local components are all sparse for economic reasons. This parameter highlights different situations. Russia-Central Asia-Transcaucasia ranks first with a density value of 0,068. Europe (0,042) and East-Southeast Asia (0,029) follow. For the remaining components, the gap broadens. Indeed, South America (0,022) and Africa-Middle East-Southern Asia (0,021) have comparable densities. Oceania and North-Central America - Caribbean components (0,017) are at the end of the scale.

Transitivity measures the proportion of triangles into a network. It refers to the probability

that two nodes connected to a third one are also directly connected. It is given by :

$$\zeta = \frac{3 * \text{number of triangles}}{\text{number of triplets}} \tag{3.2}$$

Transitivity captures the local cohesiveness of a node. It reflects the easiness of reaching neighboring airports. The larger its value, the more easily passengers can be transferred to reach their destination if a direct flight becomes unavailable. From this point of view, Europe and East and Southeast Asia components are the most transitive, with around one-third of triangles. Transitivity is slightly below for the Africa-Middle East-Southern Asia (0,28) and North- Central America – Caribbean (0,27) components. One step beyond, Russia-Central Asia-Transcaucasia (0,23) and South America (0,22) present similar behavior. Finally, the Oceania component is the less cohesive, with 18% of triangles.

Assortativity or degree correlation indicates the mixing pattern of nodes. Assortativity value ranges between -1 and 1. If it is close to 1, nodes with a similar degree tend to be linked together. In this case, the network is assortative. Instead, if it is close to -1, low degree nodes tend to associate with high degree nodes. In this case, the network is disassortative. All the components are disassortative. This general pattern corroborates the tendency of hub-and-spoke organization, where a few interconnected hubs collect the traffic of peripherical low-degree destinations. Based on the observed values, we can form three groups. The first contains Russia-Central Asia-Transcaucasia, South America, North- Central America – Caribbean components with transitivity ranging from -0,39 to -0,32. The second group including Oceania, East, and Southeast Asia, and Europe, with values ranging from - 0,22 to -0,20 exhibits a less pronounced disassortative pattern. Finally, Africa-Middle East-Southern Asia with a value of -0,15 has a more homogeneous mixing. Indeed, airline deregulation is not yet fully implemented for political and geographical reasons. Hub dominance represents the proportion of nodes connected to the highest hub within a community [Lancichinetti, 2010]. In this work, we adopt the definition for the components. In this case, it represents the proportion of nodes connected to the largest hub within a component. It is computed as the ratio between the degree of the largest hub of the component c and the size of the component c. Its value ranges from 0 to 1. Let k_i, the degree of node i, n_c, the number of nodes of a component c and $K = \{k_1, k_2, ..., k_{n_c}\}$, the set of the nodes' degrees. The component hub dominance is defined as follows.

$$\eta(c) = \frac{max(K)}{n_c} \tag{3.3}$$

Hub-dominance allows identifying the level of centralization of the components. The higher the hub dominance, the more fragile the network to a targeted attack on its principal hub. We can distinguish three categories of components. The first category, the Russia-Central Asia-Transcaucasia component, is highly centralized. Indeed, the Domodedovo airport has a direct link with 73% of the airports. This ratio indicates over-centralization, with the majority of people from all over the region traveling through Moscow. This predominance is prejudicial to interregional relations. In the second category, the major hubs share direct flights with around one-third of airports in their components. Amsterdam, Beijing, and Atlanta international air-

TABLE 3.3 – Small-world property of the components. C and C_{rand} are respectively the clustering coefficients of the local components and the random network with same number of nodes and links. L and L_{rand} are respectively the average shortest path length of the components and their corresponding random network.

	C	C_{rand}	L	L_{rand}
North and Central America-Caribbean	0,491	0,0185	2,88	2,892
Europe	0,462	0,044	2,58	2,348
East and Southeast Asia	0,511	0,028	2,89	2,687
Africa-Middle East-Southern Asia	0,504	0,0187	3,25	3,178
Oceania	0,481	0,010	3,5	4,039
South America	0,359	0,015	3,15	3,528
Russia-Central Asia-Transcaucasia	0,495	0,069	2,23	2,531
Large global component	0,11	0,014	3,28	3,13
World air transportation network	0,46	0,005	3,86	3.45

ports are those major hubs respectively in Europe, East, and Southeast Asia, and North- Central America – Caribbean components. In the last category formed by Africa-Middle East-Southern Asia, Oceania, and South America, the main hubs (Dubai, Sydney, and São Paulo) are less influential. They allow reaching directly around 20% of the destinations in their region.

Small-world networks have short path lengths and high clustering. It is a common property of numerous real-world networks. Instead, small path length and low clustering are characteristics of random networks. Table 3.3 reports the clustering coefficients and the average shortest path lengths of the seven large local components and their corresponding random networks. It shows that the clustering coefficients C of the components are significantly higher than random network values C_{rand}. Furthermore, the average shortest path length values L of the large local components are in the same order of magnitude as their corresponding random network L_{rand}. Therefore, one can conclude that large local components are small-world.

5.1.2 Degree Distribution

The degree of a node corresponds to its number of links. The *degree distribution* $P(k)$ of a network is estimated using the fraction of nodes in the network with degree k. The *degree distribution* of many real-world networks is generally well-approximated by a power law ($P(k) = k^{-\alpha}$, where α is a positive exponent). Figure 4.7 represents the empirical degree distribution and four density estimates (power-law, truncated power-law, log-normal, stretched exponential) of the large local components. The power law possesses a single parameter, while its alternatives have two degrees of freedom. Table 4.6 reports the values of the Kolmogorov-Smirnov (KS test) goodness of fit test. The smaller the value, the better the fit. Note that the two-parameters density functions outperform the power-law in any case. These results corroborate previous studies [Broido, 2019]. Indeed, the power law is a good fit for a wide range of degrees before deviating for large degrees. One can explain this phenomenon with the cost of adding new destinations. Indeed, the preferential attachment mechanism, typical of scale-free networks, where new nodes connect

TABLE 3.4 – Kolmogorov-Smirnov goodness of fit values (KS-test) for the degree distribution of the components. Distributions under test are power law, truncated power law, log-normal, and stretched exponential. The smallest value (in bold) corresponds to the best fit.

	power law	truncated power law	log-normal	stretched exponential
North and Central America-Caribbean	0.073	0.012	0.011	**0.01**
Europe	0.136	**0.024**	0.038	0.032
East and Southeast Asia	0.105	0.157	0.025	**0.023**
Africa-Middle East-Southern Asia	0.111	0.021	**0.012**	0.015
Oceania	0.150	0.078	**0.050**	0.056
South America	0.061	**0.014**	0.015	0.027
Russia-Central Asia-Transcaucasia	0.176	0.051	**0.033**	0.035
Global component	0.103	0.025	**0.024**	0.123

preferentially to highly connected nodes, tends to weaken. Due to space and time constraints, highly connected airports tend to limit their connections. Therefore, the distributions with two parameters distributions better approximate this behavior. Note that none of the distribution under test is a better fit for all the local components. Whatever, they are very similar, and they are all characterized by heavy-tails.

If we refer to the power-law exponent, it seems that the large local components exhibit a hub-and-spoke configuration. Three categories emerge. The first category with the smallest exponent value concerns the European component. As it includes the most significant proportion of hubs, it is the one with the less pronounced hub-and-spoke structure. The North and Central America-Caribbean, East and Southeast Asia, Africa-Middle East-Southern Asia, and Russia-Central Asia-Transcaucasia components form the second category. These components have almost the same proportion of hubs. Finally, the last category includes the Oceania and South America components. These components have the smallest proportion of hubs. The vast majority of the air traffic goes through these hubs. Thus, these components are more hub-and-spoke. Remember that the hub-and-spoke configuration is economically efficient, but it is more vulnerable to targeted attacks.

5.1.3 Degree-degree distance distribution of the local components

The power law assertion for the degree distribution is controversial [Broido, 2019 ; Voitalov, 2019]. Thus, the authors of [Zhou, 2020] propose to characterize real-world networks using the degree-degree distance η, $\eta(i,j) = max\{k_i, k_j\}/min\{k_i, k_j\}$. η is defined for each link. They show that for many real-world networks, the power law is a better fit for the degree-degree distance distribution as compared to the degree distribution. This feature is more pronounced for dense networks.

The degree-degree distance distribution of each large component is estimated. The results reported in Figure 3.8 show that this distribution describes better the scale-free property of the components. Indeed, for three components (East and Southeast Asia, Africa-Middle East-Southern Asia and Russia-Central Asia-Transcaucasia), their degree-degree distance distribution

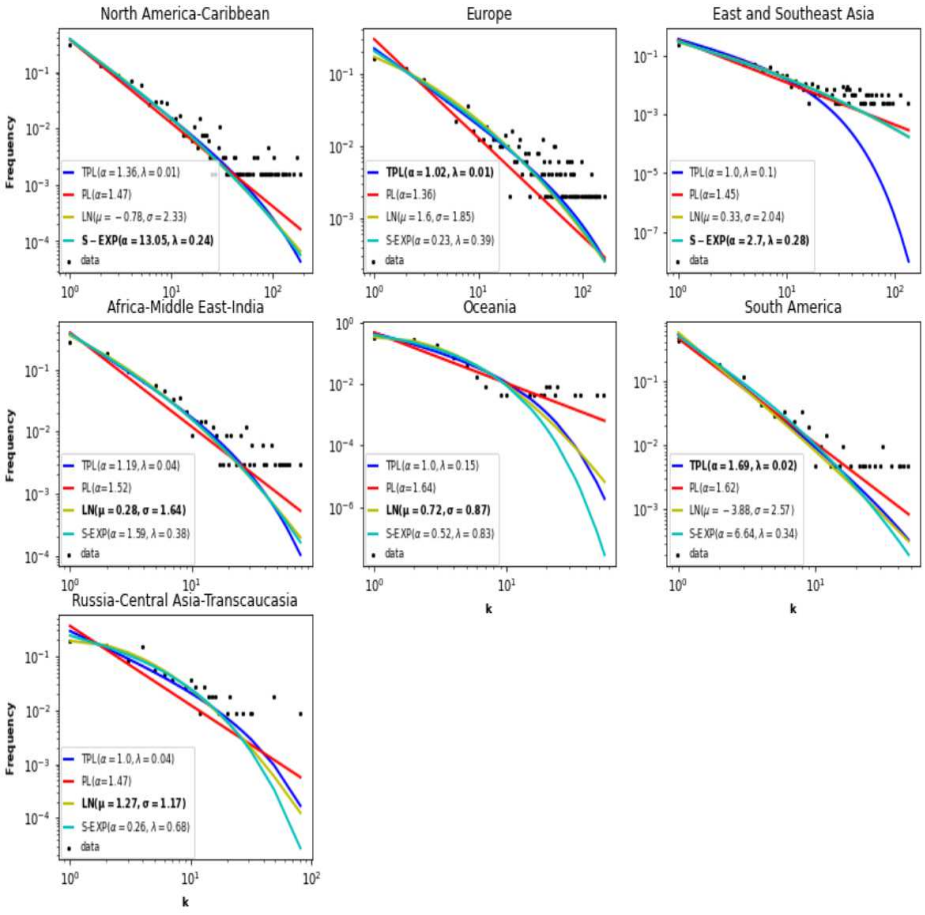


FIGURE 3.7 – Degree distributions of the large local components. Empirical distribution (**dot**), estimates (**line**).The distributions under test are : power law (PL), truncated power law (TPL), log-normal (LN) and stretched exponential (S-EXP). The values in bold are the parameters of the best fit according to the Kolmogorov-Smirnov test.

follows a truncated power law, which is a power law with a cutoff. Whereas, the degree distribution of two components follow a truncated power law. The degree-degree distance of the North and Central America-Caribbean, Europe, and South America components can be better modeled by a Log-Normal law. In the Oceania component, the degree-degree distance distribution follows a stretched exponential.

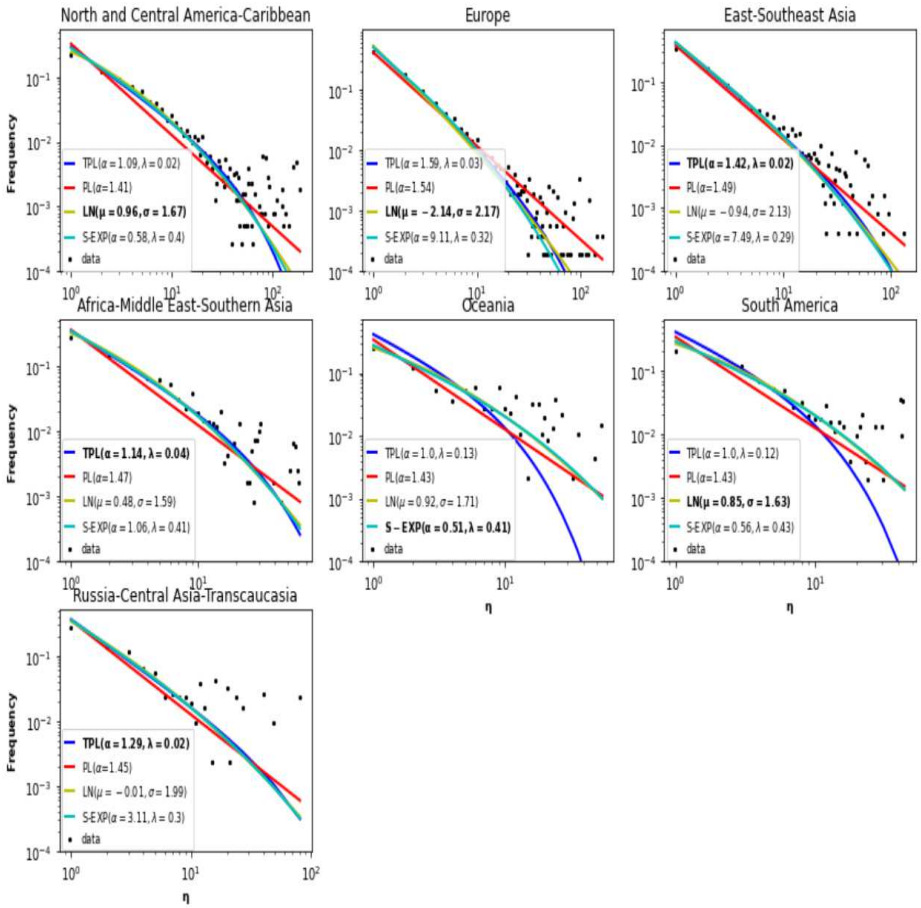


FIGURE 3.8 – Degree-degree distance distributions of the components. Empirical distribution (**dot**), estimates (**line**).The distributions under test are : power law, truncated power law, log-normal and stretched exponential. The values in bold are the parameters of the best fit according to the Kolmogorov-Smirnov test.

5.1.4 Distribution of airports by country

Figure 3.9 A illustrates the distribution of the number of airports across the large local components. One can see that they are a bit unevenly distributed. Indeed, the three most significant components (North-central America-Caribbean, Europe, East-Southeast Asia) contains almost two-thirds of the airports (64%). The Figures 3.9 [B to H] report the distribution of airports by country in the seven large local components. We can distinguish three typical cases. In the first case, one country predominates heavily. It is the case in the North America-Caribbean component. Indeed, 61% of the airports are located in the United States. It is followed by Canada, with 14% of the airports. The remaining countries have a share far below 10%. The same si-

TABLE 3.5 – Kolmogorov-Smirnov goodness of fit values (KS-test) for the degree-degree distance distribution. Distributions under test are power law, truncated power law, log-normal, and stretched exponential. The smallest value (in bold) corresponds to the best fit.

	power law	truncated power law	log-normal	stretched exponential
North and Central America-Caribbean	0.117	0.021	**0.009**	0.014
Europe	0.091	0.013	**0.01**	0.025
East and Southeast Asia	0.101	**0.012**	0.019	0.022
Africa-Middle East-Southern Asia	0.127	**0.013**	0.022	0.017
Oceania	0.146	0.282	0.060	**0.054**
South America	0.136	0.190	**0.025**	0.030
Russia-Central Asia-Transcaucasia	0.103	**0.022**	0.025	0.024
Global component	0.103	**0.022**	0.025	0.024

tuation characterizes the Russia-Central Asia-Transcaucasia component. Russia has 70% of the airports. Kazakhstan comes in the second position with 14% of airports.

In the second case, the gap between countries is less pronounced. (See Figures 3.9 [D,F,G]). It includes the East and Southeast Asia, Oceania, and South America components. In these components, the two first countries share more than 50% of the airports, and the others generally contain less than 10%. For instance, in the East and Southeast Asia components, China has the largest ratio of airports (37%), followed by Japan (14%). Australia (46%) and New Zealand (14%) dominate in Oceania. Brazil (37%), Colombia (20%), and Argentina (16%) dominate in South America.

The last case concerns the European and Africa-Middle East-Southern Asia components (See Figures 3.9 [C, E]). In these components, the proportion of airports by country is more evenly distributed. For instance, France covers 10% of airports, and the United Kingdom has 9% of airports in the European component. In Africa-Middle East-Southern Asia, 18% of airports are in India and 11% in Iran.

5.2 Analysis of the small local components

Figure 4.3 presents the airports included in the twenty small local components. They group airports from a single country or state. Among these components scattered in North and Central America-Caribbean, Europe, and Africa-Middle East-Southern Asia areas, nine contain less than five airports (3 in Canada, 3 in Europe, 2 in Africa). In the following, we consider only significant small components containing more than five airports in the analysis.

Alaska possesses five significant small local components. Three of comparable size serving around thirty airports and two with fourteen and eight airports. As many towns cannot be reached by road, air transportation is well-developed, leading to the small local component organization in the different regions of Alaska. One to three airports in each component receives flights from outside and serves cities located inside the component. For example, the Fairbanks Airport serves the majority of the airports in its component (See Figure 3.17 in appendix).

Canada contains two significant small components. The largest one connects 60 airports in the

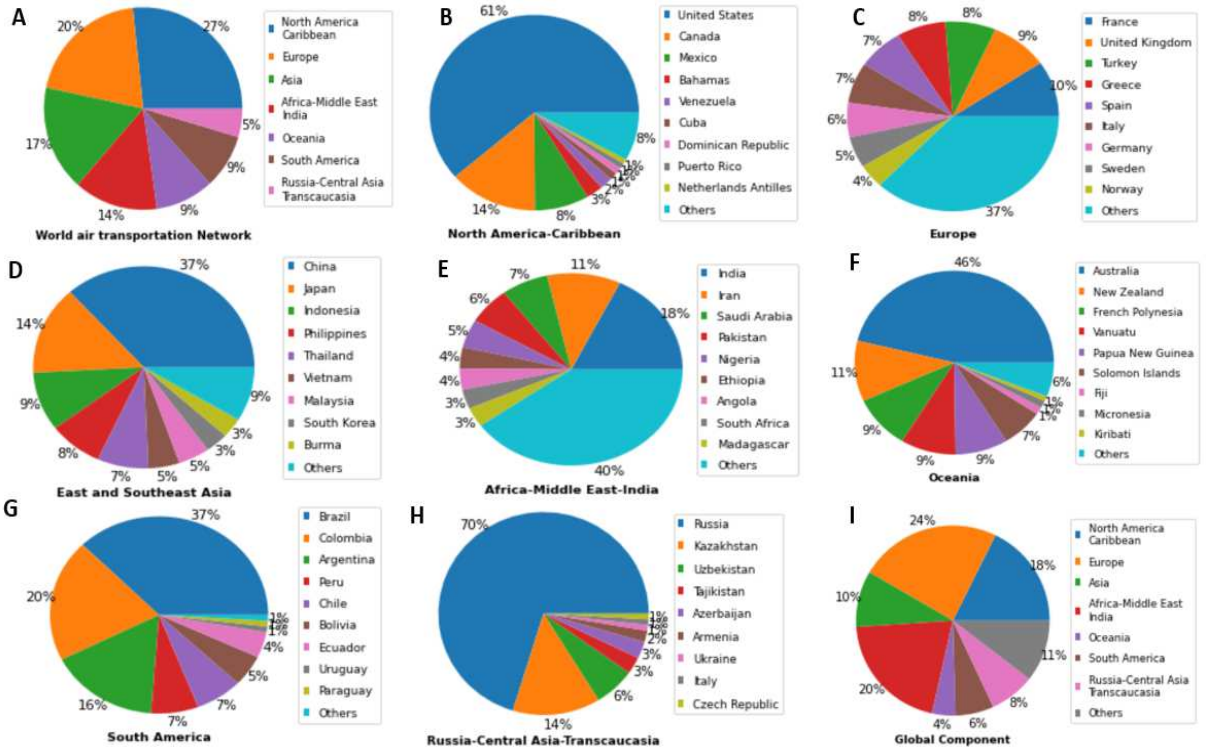


FIGURE 3.9 – Figure A represents the proportion of airports in each main local component of the world air transportation network. Figures B to H reports the distribution of airports in the various countries of the seven large local components. Figure I presents the distribution of airports per country in the large global component.

three territories : Northwest, Yukon, Nunavut.In these areas, cities are mainly reachable by plane due to climatic conditions and lack of road infrastructure. This component is well-connected with the rest of the country. Indeed, 12 airports share external liaisons with 17 airports outside the component. The second small component located in Ontario contains 25 airports. This star-shaped component is centered in Sioux Lookout, "The Hub of the North." It is reachable by three airports (Thunder Bay Airport, Red Lake Airport, and Kenora Airport)(See Figure 3.17 in appendix).

Europe has three significant small local components in Norway, Green-Land, and Scotland, with respectively 25, 10, and 7 airports. Located in northern regions, they are not easy to reach by other modes of transportation. The component in Norway is well-connected to the rest of the network, with ten airports sharing 23 liaisons with airports in other components. The Scottish component covers Orkney, an archipelago to the north of mainland Scotland. Only the airport of the largest city (Kirkwall) allows reaching this star-shaped component. Finally, one can attain the Green Land small local component that relies heavily on air transportation through two major airports (Kangerlussuaq and Godthaab/Nuuk)(See Figure 3.17 in appendix).

Finally, in Africa-Middle East-Southern Asia, the most significant small local components are in Algeria (9 airports) and Kenya (6 airports). In Algeria, these airports, located in the Sahara, serve tourist and oil-producing areas. The Tamanrasset and Ouargla airports connect the other airports of this component. The airports that are in the small local component in Kenya regroup touristic venue. The Nairobi Wilson Airport is the center of this star-shaped component (See Figure 3.17 in appendix).

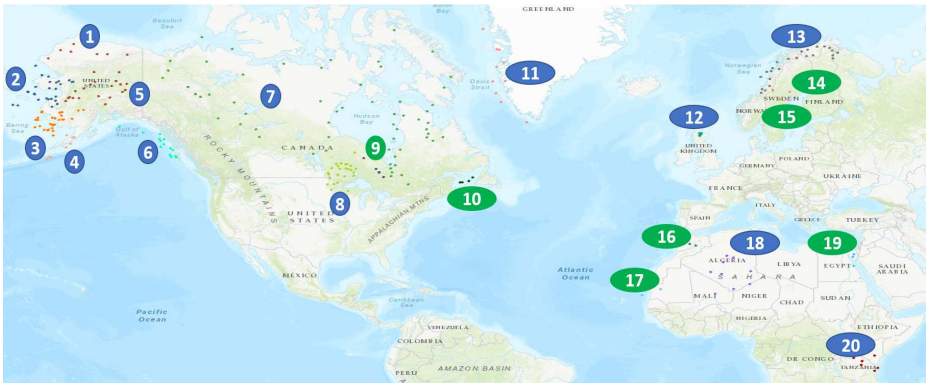


FIGURE 3.10 – The twenty small local components. The blue circles show the significant small components (size greater than five airports). The green circles show the small components with sizes up to five airports. The geographical areas not appearing in the figure do not contain small components.

5.2.1 Basic macroscopic topological properties

Table 3.6 reports the basic characteristics of the small components and the average shortest path length and transitivity of equivalent random networks. Note that we do not estimate the degree distribution because the components size is too small.

The *diameter* of the most significant small component in Canada is high. Indeed, it takes

twelve flights to travel between the most distant airports. Diameter is also high for the small components located in the USA, Canada, Norway, Green Land, and Algeria. It ranges from 6 to 4. Travel is easier in the Scotland component with a diameter of 2, and Kenya with at most three flights to reach the most distant destinations.

One can distinguish three categories of components based on the *average shortest path lenght* values. The first category contains the largest small local component, covering the Canadian territories. An average of five jumps is necessary to join any two airports. The second category regroups the components in Alaska, Ontario, and Norway. With an average shortest path length ranging from 2,8 to 2,25, one reaches its destination on average with less than three flights. Although destinations are national, it is pretty efficient. Indeed, the shortest path length values are comparable to the large local components. The third category contains Green land, Scotland, Algeria, and Kenya. With an average shortest path length ranging between 2 and 1,47, one needs no more than two flights to reach any destination on average in these star-based components.

Density shows that it is generally more challenging to get a direct flight for a destination in the small local components than in the large ones. Indeed, the components are sparse. It is particularly true for the components located in Canada, Alaska, and Norway. Density is much higher for the Green Land, Algeria, and Kenya components. The one located in Scotland is the densest.

Overall, *Transitivity* is very low. Small components in Algeria and Kenya have no triplets. In Alaska and Canada, their fraction of triangles is lower than 10%. In Europe (Scotland, Norway, Green Land), small components are more transitive. Indeed, the degree correlation values range from 30% to 50%.

The small local components are generally more *disassortative* than the large ones. Indeed, their degree correlation coefficient varies from -0,75 to -0,38. Canada's most significant small component is the only one deviating from this behavior with a much smaller value of -0,18.

A large *hub dominance* is a common feature of the vast majority of the small local components. The values ranging from 0.5 to 1 are characteristic of a star-based topology. Once again, Canada's most significant small component departs from this typical behavior. Indeed, with a hub dominance value equal to 0,18, its topology is more string-based.

Figure 3.11 illustrates the differences between these two topological organizations. In the string-based topology, the most connected nodes have almost the same degree. In contrast, a few hubs can reach more than half of the nodes in the star-based topology. Hubs in small components are gateways between other small airports and between large airports and small airports.

Five small local components out of twelve share the *small-world* property. Two in Europe (Norway, Greenland) and three in Alaska. Indeed, they have a similar or lower average shortest path than a random network. Additionally, their clustering coefficient is higher than in similar random networks. The largest component located in Canada is not small-world because of its high average shortest path length. The other components exhibit a low clustering coefficient compared to random networks.

To summarize, overall, small components appear in remote areas. They are usually poorly

TABLE 3.6 – Basic topological properties of the significant small local components (size greater than five airports). N is the network size. $|E|$ is the number of edges. d is the diameter. L is the average shortest path length. L_{rand} and C_{rand} are respectively the average shortest path and the average clustering coefficient of the corresponding random network. μ is the density. ζ is the transitivity. λ is the assortativity also called degree correlation coefficient. η is the hub dominance.

Location	N	$\|E\|$	d	L	L_{rand}	C	C_{rand}	μ	ζ	λ	η
Canada	60	80	12	5	3.77	0,28	0.05	0,045	0,2	-0,18	0,18
Canada	25	27	5	2,45	3.38	0,17	0.108	0.09	0.056	-0.57	0.7
Alaska	29	33	6	2,8	3.20	0,2	0.065	0.08	0.07	-0.56	0.68
Alaska	27	37	5	2,5	3.2	0,45	0.037	0.11	0.17	-0.38	0.62
Alaska	27	40	4	2,4	2.80	0,6	0.078	0.11	0.24	-0.44	0.54
Alaska	14	15	4	2,25	2.74	0,08	0.217	0.16	0.063	-0.63	0.62
Alaska	8	6	3	0,6	2.35	0.04	0.112	0.21	0	-0.75	0.5
Norway	25	49	4	2,37	2.37	0,68	0.142	0.16	0.39	-0.43	0.5
Greenland	10	14	4	2	2.08	0,49	0.229	0.31	0.38	-0.4	0.67
Scotland	7	11	2	1,47	1.57	0	0.523	0.52	0.5	-0.47	1
Algeria	9	10	4	2	1.78	0	0.33	0.28	0	-0.73	0.5
Kenya	6	5	3	1,86	2.06	0	0	0.33	0	-0.74	0.8

connected, making travels uneasy in these distant areas of the mainstream flows of travelers. They exhibit low transitivity and are highly disassortative. Their topology is usually star-based and more rarely string-based. They can be more or less connected to the main local component of their region.

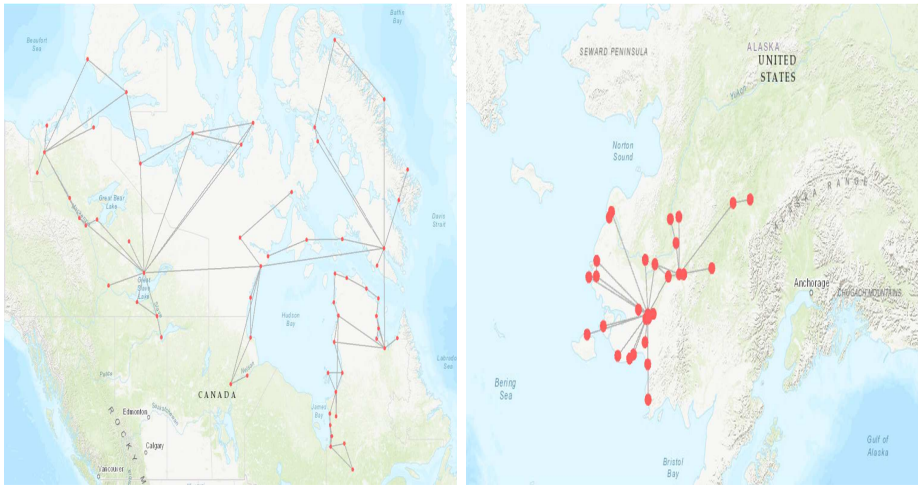


FIGURE 3.11 – Typical small local components structures. The left image presents an example of string-based topology. It is the largest small local component in Canada. The right image gives an example of star-based topology. It is the second largest small local component in Alaska.

6 Global component structure of the world air transportation network

The global components contain the airports serving destinations outside their local components through their inter-regional links. There are one large and eight small global components. Figure 4.4 represents these components. One can notice that the large global component contains 513 airports covering the world. Most of the small global components are in the North and Central America-Caribbean area. Except for one of them, which contains three airports, the other components contain only two airports. One can see in Table 3.19 reported in the appendix, that 6 of them link airports in Canada included in the large North-Central America-Caribbean with close airports from small local components. The two others link two airports at the periphery of their component. Therefore, we analyze the large global component that we call global component in the following for short.

6.1 Analysis of the large global component

Almost 20% of the airports in the world contribute to the global component. This significant proportion highlights the vitality of inter-regional exchanges. Its size (513 nodes) is in between the size of the European (493 nodes) and the North-Central America-Caribbean (657 nodes) large local components. In contrast, the number of links is much lower (2194). It compares to the number of links to East-Southeast Asia (2495) local components, which contains 20% fewer airports (416). Therefore, we focus the comparative analysis on these three local components.

6.1.1 Basic macroscopic topological properties

The *diameter* of the global component indicates that one needs a maximum of eight hops to join two airports. It is one flight more than in the North-Central America-Caribbean and East-Southeast Asia large local components, and two more flights compared with the local component of Europe.

The *average shortest path lenght* indicates that it is not as easy to travel in this component. Indeed, one needs more than three flights on average compared with less than three flights for Europe, North-Central America-Caribbean, and East-Southeast Asia local components. This parameter value is similar to the one observed in the Africa-Middle East-Southern Asia and South America local components.

Density values indicate that it is as challenging to travel in the global component as it is in North-Central America-Caribbean local component.

Transitivity is very low. Indeed, the fraction of triangles is 13%. It is almost two times less than in North-Central America-Caribbean component, and three times less than in the most transitive local component (East-Southeast Asia). Even Oceania, which is the less transitive local component, contains 18% of triangles. This result is expected. Indeed, it is more difficult to find an alternative to reach a destination in inter-regional routes often served by long-hauls.

The component is *disassortative*. Its degree correlation coefficient is comparable with that

of the European, the East and Southeast Asia, and the Oceania components.

It has the lowest *hub-dominance* compared to the large local components. Indeed, the most prominent hub (Frankfurt Airport) allows reaching only 20% of airports in the components. This value is slightly lower than in the Africa-Middle East-Southern Asia, Oceania, and South America local components

Table 3.3 indicates that it is *a small world*. Indeed, its average path length is in the same order of magnitude as the equivalent random network. Furthermore, its clustering coefficient, even low, is ten times higher than for an equivalent random network.

These results are pretty interesting. They indicate that despite being not under the control of any particular airline, the inter-regional network is quite efficient. Indeed, it compares favorably with the regional component organization of the most advanced local components. Quite naturally, with minimum planning, it allows making worldwide travels almost as easy as regional travel. Being less centralized, it is also more resilient. Unfortunately, its lack of transitivity makes it difficult to use alternative routes.

6.1.2 Degree and degree-degree distance distribution

Figure 3.13 shows the degree and degree-degree distance distributions of the large global component. Visually, it is not so easy to point the differences between both distributions. According to the KS goodness of fit test, the truncated power law better approximates the degree distribution. The log-normal function is a better fit for the degree-degree distance distribution. Whatever, both distributions are heavy-tailed.

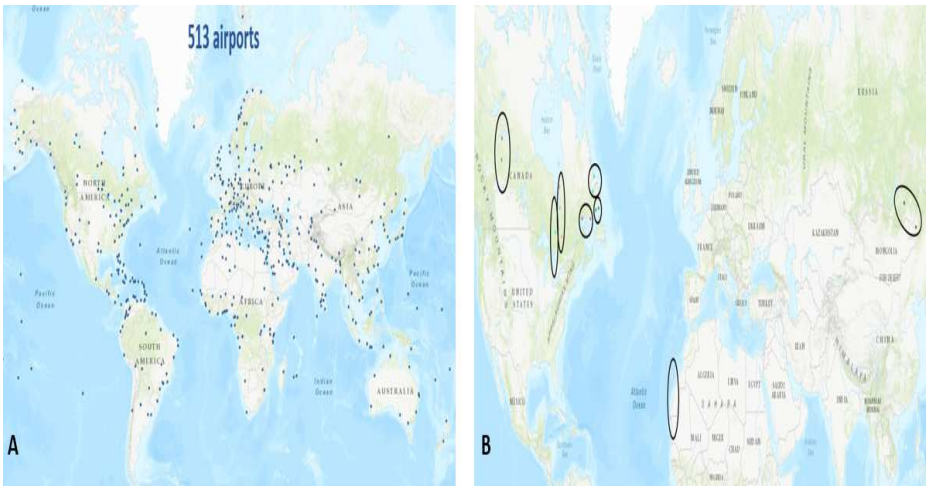


FIGURE 3.12 – (A) The airports in the large global component. (B) The 8 small global components which are circled. The size range are between 2 and 3 airports.

6.1.3 Distribution of the airports between regions

The pie-chart in Figure 3.9.I. illustrates the distribution of the global component airports between the seven regions corresponding to the local components. Comparing this chart with

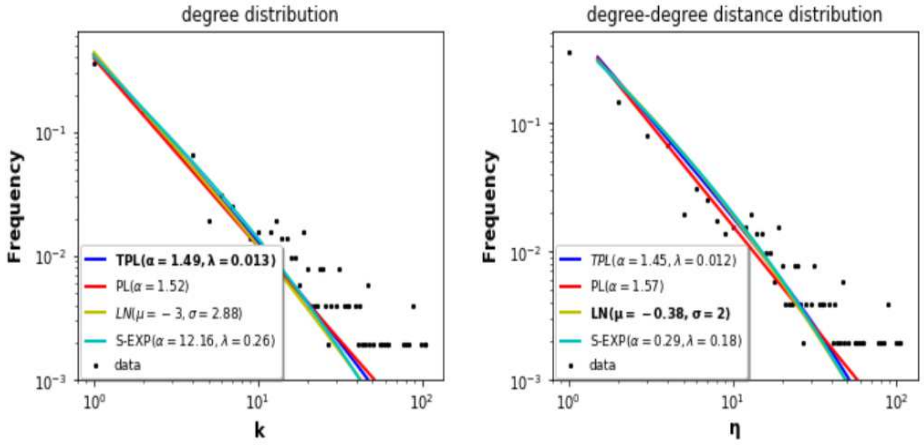


FIGURE 3.13 – The degree (**left**) and degree-degree distance (**right**) distributions of the large global component. Empirical distribution (**dot**), estimates (**line**).The distribution under test are : power law (PL), truncated power law (TPL), log-normal (LN) and stretched exponential (S-EXP). The values in bold are the parameters of the best fit according to the Kolmogorov-Smirnov test.

Figure 3.9. A. allows getting an idea about how the various regions are involved in the global air routes. One can see that these distributions, although correlated, can differ. Indeed, the regions with a big share of airports are not necessarily the most active in inter-regional traffic.

For example, the North and Central America-Caribbean local component contains the highest fraction of airports worldwide (27%). However, its share on the liaisons with the other regions is 18%. It indicates that a high proportion of airports serves local destinations. East and Southeast Asia exhibits a similar trend, with 17% of the airports in the world but only 10% of the inter-regional routes. It is also the case for South America and Oceania.

In contrast, Europe with its 20% of airports is more international with its 24% of airports exchanging flights with other regions. This behavior is more pronounced for Africa with its 14% of airports and 20% of inter-regional flights. It is probably due to political and historical reasons. Indeed, Africa still lacks global airlines. The air traffic flows mainly through Europe or the Middle East. Russia-Central Asia-Transcaucasia is intensely involved in inter-regional exchanges even if it has a low proportion of the world airports.

7 Comparison of the world air transportation network with the large components

7.1 Basic macroscopic topological properties

The basic topological properties of the world air transportation network are reported in Table 3.1.

The *diameter* of the world air transportation network is two to three times greater than the diameter of the large components. While the longer route ranges from 4 to 8 flights in the components, it takes 12 flights to join the furthest destinations worldwide.

Its *average shortest path lenght* is also higher compared to the components, but the difference is more moderate. Indeed, one needs on average 3,6 flights to reach any destination in the world compared to 3, 5 flights when traveling in the Oceania component.

Density shows that it is more uneasy to reach remote locations in the world air transportation networks than to travel in a single component. Indeed, the density is one degree of magnitude lower than in Europe

Transitivity is a good feature of the world air transportation network. With 26% of closed triangles, reaching a destination with a direct flight is as easy as traveling in the Africa-Middle East-Southern Asia component or the North-Central America-Caribbean component. Transitivity is higher than in the Oceania, South America, and Russia-Central Asia-Transcaucasia components.

It is slightly *disassortative*. Indeed, its degree correlation coefficient is approximately ten times higher in the local and global components.

It has the lowest *hub-dominance* compared to the components. Indeed, the largest hub (Frankfurt Airport) is linked to 9% of the airports worldwide. In the components, this value range from 20% to 73%.

It shares the *small world* property with the large components. Indeed, its average path length is low, and its transitivity is high compared to a similar random network, as indicated in Table 3.3.

7.2 Degree and degree-degree distance distribution

Degree and degree-degree distance distributions of the world air transportation network are reported in Figure 3.14. The truncated power law is the best fit in both cases. This result is in line with previous studies.

Overall, these results are very informative. They demonstrate the advantages of the component structure representation to get a clear picture of the diversity of the world air transportation network. Indeed, looking at the topological properties of the global air transportation network can create an inaccurate view of the actual situation. The regional and inter-regional situations are very diverse. Unfortunately, the network analysis offers an “average view” of a non-homogeneous world. Therefore, one needs to interpret these data very carefully.

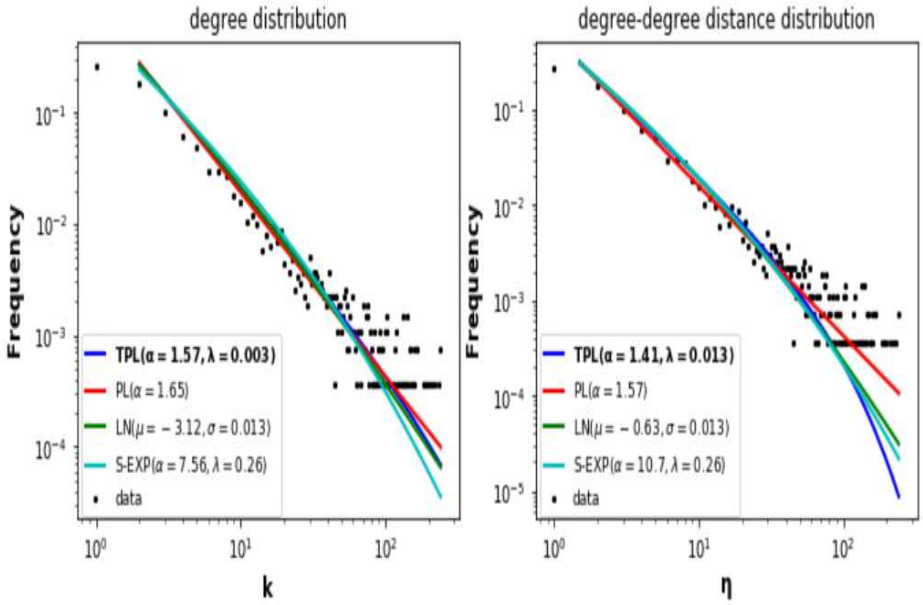


FIGURE 3.14 – The degree (**left**) and degree-degree distance (**right**) distributions of the world air transportation network. Empirical distribution (**dot**), estimates (**line**). The distribution under test are : power law (PL), truncated power law (TPL), log-normal (LN) and stretched exponential (S-EXP). The values in bold are the parameters of the best fit according to the Kolmogorov-Smirnov test.

8 Degree Centrality Analysis

Centrality measures aim to assess the importance of a node based on some topological features. Degree centrality is among the most commonly used measure. It measures the node influence with its number of links. This section investigates the degree centrality to assess the influential airports according to the different views carried by the component structure. First, we consider the large local components to uncover the top connected airports at the regional level. Then, we use the large global component to rank the inter-regional airports. Finally, we compare these results to the top connected airports in the world air transportation network. This process allows us to highlight the differences between these three levels of analysis.

8.1 Exploring the regional hubs in the large local components

Table 4.7 lists the top 5 connected airports in the seven large local components called regional hubs. It also reports the cumulative proportion of airports connected to the regional hubs.

The five top connected airports of the *North and Central America-Caribbean component* are in the United States. They are in four different states (Illinois, Texas, Colorado, Georgia). They are the most important of major airlines in the United States. Indeed, each of these airports receives more or less 50 millions passengers per year. Hartsfield J Atlanta Airport is Delta Air Lines' headquarter and one of its central hubs. It is one of the busiest airports worldwide, with almost one hundred millions passengers per year. The Dallas Fort Worth Airport ranks second. Located in Fort Worth, it is the primary hub of American Air Lines and Southwest Airlines. The

TABLE 3.7 – Degree Centrality in the seven large local components. The top five hubs for each large local component are reported. ID is the internal degree. It measures the number of connections of an airport with airports located in its large local component. LR is the airport local rank in its component by decreasing internal degree. CP is the cumulative fraction of connected airports represents the proportion of airports that the top x local hubs can reach in their components.

Large Local Component	Airport	City	Country	ID	LR	CP (%)
North and central America Caribbean	Hartsfield J	Atlanta	USA	187	1	23,89
	Dallas Fort Worth	Dallas	USA	172	2	35,92
	Chicago O'Hare	Chicago	USA	172	3	40,18
	Denver	Denver	USA	161	4	47,64
	George B Houston	Houston	USA	146	5	49,77
Europe	Schiphol	Amsterdam	Netherlands	162	1	32,86
	Munich	Munich	Germany	145	2	40,36
	London Stansted	London	UK	142	3	51,72
	Barcelona	Barcelona	Spain	137	4	55,78
	Frankfurt	Frankfurt	Germany	132	5	56,79
East and Southeast Asia	Beijing Capital	Beijing	China	133	1	31,97
	Guangzhou Baiyun	Guangzhou	China	112	2	37,25
	Shanghai Pudong	Shanghai	China	112	3	41,34
	Chengdu Shuangliu	Chengdu	China	95	4	44,47
	Taiwan Taoyuan	Taipei	Taiwan	87	5	46,63
Africa Middle East-Southern Asia	Dubai	Dubai	U A E	78	1	23,14
	King Abdudaziz	Jeddah	Saudi Arabia	63	2	30,56
	Indira Gandhi	Delhi	India	60	3	39,76
	Chhatrapati Shivaji	Mumbai	India	58	4	42,13
	Addis Ababa Bole	Addis Ababa	Ethiopia	56	5	51,33
Oceania	Sydney K Smith	Sydney	Australia	55	1	23,5
	Brisbane	Brisbane	Australia	48	2	31,19
	Auckland	Auckland	New Zealand	36	3	38,46
	Melbourne	Melbourne	Australia	33	4	40,59
	Port M Jacksons	Port Moresby	P New Guinea	24	5	48,71
South America	G G A F Montoro	Sao Paulo	Brazil	48	1	22,42
	El Dorado	Bogota	Colombia	44	2	40,65
	Viracopos	Campinas	Brazil	42	3	45,32
	Jorge Newbery	Buenos Aires	Argentina	41	4	60,28
	P J Kubistschek	Brasilia	Brazil	36	5	64,48
Russia Central Asia-Transcaucasia	Domodedovo	Moscow	Russia	81	1	72,32
	Pulkovo	St Petersburg	Russia	49	2	82,14
	Sheremetyevo	Moscow	Russia	49	2	84,82
	Tolmachevo	Novosibirsk	Russia	32	4	85,71
	Koltsovo	Yekaterinburg	Russia	31	5	85,71

third hub of this component, the Chicago O'Hare & Chicago Airport, is one of the main hubs of United Air Line (headquartered in Chicago) and American Air Line. Denver Airport is the fourth hub of this component. It has the largest area in the United States. Headquarter of the

low-cost carrier Frontier, it is a crucial hub for United Air Lines. The second important airport in Texas, the George B Houston Airport, is this component's fifth hub. It is one of the hubs of United Air Lines. Altogether, these first five hubs connect half of the airports in the component.

The top five hubs in the *European component* are in four different countries (Netherlands, Germany, UK, Spain). These airports are the stronghold of the biggest airlines in these countries. The number of passengers ranges from approximately 70 millions to 20 millions per year. The Amsterdam Schiphol Airport is the first hub of this component. It is the largest in the Netherlands and the most critical hub of KLM. It is also a very active platform for low-cost carriers such as EasyJet, Vueling, TUI Air Lines, Corendon Dutch. Munich Airport in Germany is second. It plays an essential role for Lufthansa Air Lines and its low-cost subsidiary (Air Dolomity, Eurowings) to serve European destinations. The London Stansted Airport in the United Kingdom is the third hub of the component. It is a major airport for low-cost carriers such as Ryanair, Pegasus, Easy Jet [Dobruszkes, 2006]. Due to its geographical position and the high level of tourism in Barcelona, its airport is very well connected. It receives about 20 millions passengers per year. Indeed, the Barcelona airport in Spain is the fourth most connected airport. It is the primary airport of the Vueling Air Lines. Frankfurt in Germany, the fifth hub, is the main historical airport of Lufthansa. While Munich is more focused on European destinations, Frankfurt is devoted to international destinations. These five hubs allow reaching about 56% of the airports of this component. The London Gatwick Airport (132 links) in the United Kingdom, Charles de Gaulle Airport (129 links) in France, Dublin Airport (128 links) in Ireland, Düsseldorf Airport (121 links) in Germany, and Ataturk Airport (121 links) are the five other hubs. The top first ten hubs can reach 66% of airports of the component. They are the home of the leading European carriers.

The largest four hubs in the *East and Southeast Asia component* are in different provinces and municipalities of China. The fifth hub is in Taiwan. These airports receive tens of millions of passengers per year. The most connected, Beijing Capital Airport, is one of the busiest airports in the world, with almost 100 millions passengers yearly. The Guangzhou Baiyun airport, the second hub of this component, is situated in the capital of the provincial Guangdong in South China. It is a hub of the China Southern Air Lines, which is headquartered in Guangzhou. The Shanghai Pudong Airport, situated in East Asia, has the same degree as the Guangzhou Baiyun airport. It is a major hub of the Shanghai and China Eastern Air Lines. The fourth hub, called Chengdu Shuangliu Airport, is in the capital of Sichuan province. Serving Central-western China, it is the central hub of the Sichuan and Chengdu Air Lines. The Taiwan Taoyuan Airport is the fifth regional hub in East and Southeast Asia. Serving Taipei, it is the largest airport on the island. Together, these hubs are connected to 46,6% of East and Southeast Asia airports.

The first five hubs of the *Africa-Middle East-Southern Asia component* span Africa, the Middle East, and Southern Asia. Indeed, two are in the Middle East, two in India and one in Africa. The first, Dubai airport, is one of the most important airports in the world. It can handle more or less 100 millions passengers. It is the most critical hub of Emirates Air Lines. King Abdulaziz Airport in Saudi Arabia ranks second. It is the largest airport in terms of surface area. Located near Mecca, it receives many visitors from different Muslim countries. The biggest

airport in India, Indira Gandhi Airport, is third. Situated in the capital Delhi, a cosmopolitan and tourist city, it receives a lot of people. The Chhatrapati Shivaji Airport, located in the megacity of Bombay, the economic capital of India, is the following hub. It is the central hub of India Air Lines. The Addis Ababa Bole Airport in the capital of Ethiopia is the fifth hub of the component. Due to its political and geographical position, it connects Africa with the rest of the world. It is the hub of Ethiopia Air Lines, the first airline serving Africa's airports. These five hubs connect about 51% of airports of the Africa-Middle East-Southern Asia component.

Among the top five hubs in the *Oceania component*, three are in Australia, one in New Zealand, and one in Papua New Guinea. These airports receive less than 50 millions passengers per year. The Sydney Kingsford Smith Airport is the largest hub of the Oceania component. Indeed, located in the capital city of New South Wales in Australia, Sydney is a metropolis and touristic city attracting people coming from different countries. It is the main hub of Qantas, the largest airline in Australia. The Brisbane Airport, situated in the capital city of Queensland, is the second most connected airport in this component. Like Sydney, Brisbane is also a tourist city that receives many people from different horizons. It is the primary airport of Virgin Australia and a hub for Qantas. The third hub, Auckland Airport, is the top airport in New Zealand and the base of Air New Zealand. The Melbourne-Tullamarine Airport located in Melbourne ranks fourth. It is essential for Jetstar Air Lines, based in Melbourne and Qantas. Melbourne is also a metropolitan city that receives numerous tourists. The fifth hub of this component, the Port Moresby Jacksons Airport, is the most influential airport of Papua New Guinea. Located near Port Moresby, this airport is the largest hub of Air Niugini, headquartered in Port Moresby. These five hubs serve 48,7% of the airports in the component.

Three of the top five hubs in the *South America component* are in Brazil, one in Colombia, and one in Argentina. Guarulhos G A F Montoro Airport, the first hub, is in São Paulo, the most populated city in Brazil. The busiest in Brazil [Couto, 2015], it is an essential airport for GOL (based in Rio de Janeiro) and ITA Air Lines (based in São Paulo). The second hub in Columbia, the El Dorado Airport, is in Bogotá, the capital of Colombia. It is the base for Avianca Air Lines. The Viracopos Airport in Campinas ranks third. It has a regional vocation in Brazil [Couto, 2015]. It is a hub for Azul Brazilian Airlines, a low-cost airline. The Jorge Newbery Airport is the fourth hub of this component. Located in Buenos Aires, the capital of Argentina, it is a hub for the largest Argentina airline (Aerolíneas Argentinas). The fifth hub, the Presidente J Kubistschek, is in Brazilia, the capital of Brazil. It is one of the hubs of GOL and LATAM Brasil Air Lines. These five largest hubs reach around 64% of airports in this component.

In the *Russia-Central Asia-Transcaucasia component*, the top five airports are in the major cities of Russia. They receive less than 50 millions passengers per year. The Domodedovo Airport, located in Moscow, is the first hub. It centralizes about 72% of the routes. Sheremetyevo Airport in Moscow and Pulkovo Airport, the main airport of St Petersburg, are tied for second. Together the two top airports in Moscow concentrate 80% of the destination of the component. Therefore, Moscow is the heart of the component and the Russian air network in particular. This result corroborates the findings reported in [Tarkhov, 2017]. Indeed, this chapter shows that since the fall of the Soviet Union, the Russian air network is more and more centralized in Moscow. Almost

all flights (regional and international) transit through this city. The Domodedovo Airport is a hub of the Domodedovo Air Lines, and the Sheremetyevo Airport is a hub of Aeroflot–Russian Airlines, the biggest in Russia. The Pulkovo Airport is the main airport of St Petersburg, the second-largest city in Russia. It is also an important airport for Rossiya Airlines. Another hub of this component is the Tolmachevo Airport, located near Russia's major city, Novosibirsk. This airport has a regional vocation. It is an important airport for the Novaport Air Lines. Situated in Yekaterinburg city, the Koltsovo is the fifth most connected airport of this component. It is a vital hub for the Ural Air Lines, based in the same city as the airport. These hubs reach around 85% of airports in this component.

8.2 Exploring the inter-regional hubs in the large global component

We consider the large global component to extract the high degree airports. For comparative purposes, we classify the so-called inter-regional hubs in seven regions based on their local component. Indeed, an airport in a global component is also a member of a local component. Table 4.8 reports their coordinates and their external degree (number of links with airports outside their local component). Their global rank in the component is computed according to the decreasing order of their external degree. The global rank in their region measures their relative global position in their region. Four out of five of the top inter-regional hubs in the North and central American-Caribbean area are in the USA. These hubs are located on the coasts, closer to the other regions. The largest hub of this component is the John F Kennedy Airport, located in New York. This gateway to the United States is the busiest airport for international flights, with around fifty millions passengers yearly. It is an essential platform for American Airlines, Delta, and Jet Blue. The most important airport in Canada, the Lester B Pearson Airport, is the second-largest inter-regional hub. It connects Canada and the US to the other regions. Situated in Toronto, it is the primary hub of Air Canada. Newark Liberty Airport ranks third. This crucial hub of United Airlines serves the New York metropolitan area. Miami and Los Angeles complete the top five. These airports allow reaching the United States from South America and Asia, respectively. The top five European airports, Frankfurt, Charles de Gaulle, Ataturk, Heathrow, Amsterdam, are the primary hubs of the leading airlines in Europe, Lufthansa, Air France, Turkish Airlines, British Airways, and KLM, respectively. Note that Frankfurt and Amsterdam Schiphol airports are also among the five most important regional hubs. Charles de Gaulle Airport, the first French airport, is located in Paris, the world-leading destination in the world. Thanks to its central position, the Ataturk Airport is a vital hub where several air flights transit. Indeed, it is at a gateway for Asia, the Middle East, and Africa. The fourth inter-regional hub is London Heathrow, located in the top tourist city of the United Kingdom, London. It is the leading international airport in London.

While China dominates the regional hubs, inter-regional hubs are scattered in different countries of East and Southeast Asia. Beijing Capital Airport is the inter-regional hub. The Suvarnabhumi airport located in Bangkok, the capital city of Thailand, ranks second. It is an important hub serving local (Thai Airways and Bangkok Airways) and foreign airlines connecting to Asia, Oceania, Europe, and Africa. Narita, located in Tokyo, the capital of Japan, is a hub

TABLE 3.8 – Degree centrality in the large global component. The top five inter-regional hubs in each region are reported. ED is the external degree. It measures the number of connections of an airport with airports located in the large global component. GRR is the global rank of the airport in its region. GRC is the global rank in the global component is the airport rank in the global component by decreasing external degree.

Region	Airport	City	Country	ED	GRR	GRC
North and central America-Caribbean	John F Kennedy	New York	USA	62	1	9
	Lester B Pearson	Toronto	Canada	45	2	19
	Newark Liberty	Newark	USA	39	3	24
	Miami	Miami	USA	36	4	27
	Los Angeles	Los Angeles	USA	34	5	30
Europe	Frankfurt	Frankfurt	Germany	106	1	1
	Charles de Gaulle	Paris	France	104	2	2
	Atatürk	Istanbul	Turkey	99	3	3
	Heathrow	London	UK	89	4	4
	Schiphol	Amsterdam	Netherlands	80	5	7
East and Southeast Asia	Beijing Capital	Beijing	China	61	1	10
	Suvarnabhumi	Bangkok	Thailand	57	2	12
	Narita	Tokyo	Japan	56	3	13
	Incheon	Seoul	South Korea	48	4	15
	Hong Kong	Hong Kong	Hong Kong	47	5	16
Africa-Middle East-Southern Asia	Dubai	Dubai	U A E	89	1	4
	Hamad	Doha	Qatar	59	2	11
	Abu Dhabi	Abu Dhabi	U A E	42	3	22
	Cairo	Cairo	Egypt	34	4	30
	Indira Gandhi	Delhi	India	31	5	32
Oceania	Sydney K Smith	Sydney	Australia	26	1	41
	Melbourne	Melbourne	Australia	20	2	57
	Auckland	Auckland	New Zealand	13	3	96
	Perth	Perth	Australia	13	3	96
	Brisbane	Brisbane	Australia	12	4	106
South America	Guarulhos-G A F M	São Paulo	Brazil	36	1	26
	El Dorado	Bogotá	Colombia	30	2	36
	Jorge Chávez	Lima	Peru	22	3	53
	Ministro Pistarini	Buenos Aires	Argentina	19	4	61
	Rio G–T Jobim	Rio de Janeiro	Brazil	18	5	69
Russia-Central Asia-Transcaucasia	Sheremetyevo	Moscow	Russia	84	1	6
	Domodedovo	Moscow	Russia	77	2	8
	Pulkovo	St Petersburg	Russia	47	3	16
	Vnukovo	Moscow	Russia	31	4	32
	Heydar Aliyev	Baku	Azerbaijan	23	5	51

for Japan Air Lines and All Nippon Airways. The largest airport in South Korea, the Incheon Airport, is the fourth inter-regional hub. Located on the coast in Seoul, which receives a million tourist passengers per year, this airport is an important hub for Korean Air Lines. The Hong Kong Airport is the primary hub of Cathay Pacific and Dragon Air Lines, headquartered in Hong Kong.

The regional hubs, Dubai and Indira Gandhi Airport, are respectively the first and the fifth inter-regional hubs in the Africa-Middle-East Southern Asia area. The Hamad airport is the second inter-regional hub in this zone. This airport, located in Doha, the capital of Qatar, is a gateway to Europe, Asia, Africa, and Oceania. It is the most important hub for Qatar Airways. Abu Dhabi, the second-largest airport in the United Arab Emirates, is also strategically located. It is the primary hub of Etihad Airways. The fourth hub located in Cairo is the vital hub of Egypt Air.

In Oceania, the four regional hubs (Sydney K Smith Airport, Melbourne Airport, Auckland Airport, and Brisbane Airport) are also the inter-regional hubs. In addition, Perth Airport is an inter-regional hub. This airport, located on the coast of the capital city of Western Australia, is an essential airport for Alliance Airlines. One can see that Australian airports are the most influential in this area.

The two largest inter-regional hubs in South America (Guarulhos in Brazil and El Dorado in Colombia) are also the two largest regional hubs. The third hub, Jorge Chávez Airport, is located in Lima, the capital of Peru. It is the hub of different airlines such as ATSA, LATAM Perú Airlines. The largest airport in Argentina, the Ministro Pistarini Airport, located in the capital city of Argentina, has a vital role in international flights. It is the hub of Aerolíneas Argentinas.

In the Russia-Central Asia-Transcaucasia area, the top four inter-regional hubs are in Russia. Note that the top three are also present in the list of the top five regional hubs. The Vnukovo Airport is the fourth inter-regional airport. Located in Moscow, it is a hub for the Azur Air Lines. It corroborates the extreme centralization of the Russian air transport network in Moscow. The Heydar Aliyev Airport, located near the coast in Baku, the capital city of Azerbaïdjan, is the fifth inter-regional hub of the area. This airport is essential for Azerbaijan Air Lines.

8.3 Comparison of the regional and inter-regional hubs with the hubs of the world air transportation network

We now compare the degree centrality of the nodes in the world air transportation network with the various degree centrality measures of the components. Table 4.9 reports information about the top 36 airports worldwide according to their total degree (the sum of their internal and external degree). It also contains their total degree, worldwide rank, local rank, and eventually global rank. For comparative purposes, airports are classified in the seven regions identified by the local components. At first glance, one can see that the distribution among regions is not homogeneous. Indeed, Europe and the North and Central America-Caribbean areas dominate, with 29 out of 36 airports in the world's top ranks.

Thirteen airports of the North and Central America-Caribbean component belong to the top 36 of the world air transportation network. The top five are the top five regional airports. The eight other airports include the first five inter-regional hubs. The other three airports (Charlotte Douglas, McCarran, Detroit M Wayne) are regional hubs ranked 6, 7, and 9 in the local component. In this part of the world, regional hubs are more influential than inter-regional hubs. Indeed, the first inter-regional hub ranks nine in the world network. It shows

that the air traffic in the North and Central America-Caribbean zone focuses mainly on regional destinations. The rank of the regional airports in the world air transportation network can be misleading because they are not the most active in inter-regional traffic.

With 16 airports, European airports are the most numerous in the top 36 in the world network. They are also among the most influential. Indeed, the top four world airports are in Europe. The top of the ranking includes the top inter-regional and regional airports (Frankfurt, Charles de Gaulle, Ataturk, London Heathrow, Amsterdam, Munich, London Stansted, Barcelona). Five regional airports complete this list. Three are in the UK (London Gatwick, Dublin, Manchester), one is in Germany (Dusseldorf), and one is in Austria (Vienna). Note that Dusseldorf uses to be the second airport in Germany before Munich took the lead. Two airports located in Rome and Madrid are more focused on inter-regional traffic. These results suggest that European airports are more oriented towards inter-regional connections.

There are four airports from the East and Southeast Asia area in the top 36 hubs of the world air network. Two are in the top five inter-regional hubs (Beijing Capital, Incheon), and two are in the leading regional hubs (Shanghai Pudong, Guangzhou Baiyun). Remember that Beijing Capital airport, which ranks seven in the world air network, is also the region's top regional and inter-regional airport. This repartition between regional and inter-regional hubs indicates a balanced regional and inter-regional traffic dominated by China.

Dubai is the only airport in the top 36 world air network hubs for the Africa-Middle East-Southern Asia area. Ranked 10 in the world network, it is an essential regional hub (1st) and inter-regional hub (4th).

There is no airport of Oceania in Table 4.9. The first airport from this region, the Sydney K Smith Airport, ranks 84 worldwide. It is also the case for South America. Indeed, Guarulhos-G A F M, the highest degree airport in Brazil, ranks 74 in the world air network.

Two airports from the Russia-Central Asia-Transcaucasia zone are in the first 36 hubs of the global air network. These airports in Moscow are the most important hubs in the region at both regional and inter-regional levels.

To summarize, Europe and the United States dominate the world transportation network in terms of centrality. The main drawback of using world air transportation to evaluate the airports' influence is that it cannot distinguish regional from inter-regional importance. Consequently, large parts of the world are hidden due to striking regional discrepancies. The component structure allows taking into account these regional disparities. Indeed, airports at the bottom in the world network ranking, which are very influential in their region, emerge. It is the case for airports such as the London Stansted Airport, the Chengdu Shuangliu Airport, King Abdulaziz Airport, Sydney K Smith Airport, Guarulhos-G A F M Airport, and the Pulkovo Airport that are the most connected in their region. Similarly, it brings to the attention that airports, such as Suvarnabhumi, Hamad, Melbourne, El Dorado, and Heydar Aliyev, are crucial for inter-regional traffic.

TABLE 3.9 – The top 36 hubs of the world air transportation network. WR is the worldwide rank in the world air transportation network. It is computed using the total degree (internal and external). LR is the local rank computed in the large local components. GR is the global rank in the global component. No indication is given when an airport does not belong to the global component. Top inter-regional airports are in bold. Top regional and inter-regional airports are in bold and italic. Others are top regional airports.

Region	Airport	City	Country	Degree	WR	LR	GR
North and central America-Caribbean	Hartsfield J Atlanta	Atlanta	USA	212	5	1	43
	Chicago O'Hare	Chicago	USA	200	6	3	39
	Dallas Fort Worth	Dallas	USA	188	8	2	78
	Denver	Denver	USA	165	12	4	197
	George B Houston	Houston	USA	165	12	5	66
	John F Kennedy	**New York**	**USA**	**160**	**14**	**15**	**9**
	Newark Liberty	Newark	USA	152	19	10	24
	Lester B. Pearson	**Toronto**	**Canada**	**147**	**22**	**13**	**19**
	Los Angeles	**Los Angeles**	**USA**	**144**	**24**	**11**	**30**
	Charlotte Douglas	Charlotte	USA	141	26	6	137
	Miami	**Miami**	**USA**	**138**	**30**	**14**	**27**
	McCarran	Las Vegas	USA	133	33	7	162
	Detroit M Wayne	Detroit	USA	133	33	9	120
Europe	***Schiphol***	***Amsterdam***	***Netherlands***	***242***	***1***	***1***	***7***
	Frankfurt	***Frankfurt***	***Germany***	***241***	***2***	***5***	***1***
	Charles. de Gaulle	***Paris***	***France***	***233***	***3***	***6***	***2***
	Ataturk	***Istanbul***	***Turkey***	***220***	***4***	***10***	***3***
	Munich	Munich	Germany	186	9	2	23
	Heathrow	**London**	**UK**	**167**	**10**	**33**	**4**
	L da Vinci-F	Rome	Italy	158	15	16	14
	London Gatwick	London	UK	157	17	6	45
	Barcelona	Barcelona	Spain	156	18	4	63
	Brussels	Brussels	Belgium	150	20	11	36
	Adolfo Suárez	Madrid	Spain	150	20	17	17
	Stansted	London	UK	143	25	3	46
	Dublin	Dublin	Ireland	141	26	8	101
	Düsseldorf	Düsseldorf	Germany	141	26	9	59
	Manchester	Manchester	UK	141	26	13	46
	Vienna	Vienna	Austria	136	31	14	50
East and Southeast Asia	***Beijing Capital***	***Beijing***	***China***	***194***	***7***	***1***	***10***
	Shanghai Pudong	Shanghai	China	147	22	3	28
	Guangzhou Baiyun	Guangzhou	China	135	32	2	52
	Incheon	**Seoul**	**South Korea**	**133**	**33**	**4**	**15**
Africa-Middle East-Southern Asia	***Dubai***	***Dubai***	***U A E***	***167***	***10***	***1***	***4***
Russia-Central Asia-Transcaucasia	***Domodedovo***	***Moscow***	***Russia***	***158***	***15***	***1***	***8***
	Sheremetyevo	***Moscow***	***Russia***	***133***	***33***	***3***	***6***

9 Core Analysis

In this section, we investigate the core structure of the components, and we perform a comparative analysis with the core of the world transportation network. To this end, we compute the max k-core. The k-core of a network is a sub-network in which a node has at least k neighbors. One can extract the maximum k-core by removing nodes iteratively from the network. Indeed, the 1-core, where each node has at least one link, is the whole network. One forms the 2-core by removing all the nodes with one connection in the 1-core network. Then, one extract the 3-core by removing all the nodes with two links in the 2-core, and so on until one reaches the core number max k-core for which it is impossible to obtain the (k+1)-core.

9.1 Exploring the max k-core in the large local components

Figure 4.12 reports the airports included in the maximum k-core of the seven large local components with the value of the maximum k-core uncovered. In the following, we use indifferently the words max k-core and core for short.

The core of the North and Central America-Caribbean component contains 42 airports sharing at least 29 connections. Except for Lester B Pearson Airport (first local hub in Toronto, Canada) and the Cancun Airport (second hub in Mexico), these airports are in the USA. Table 4.10 (in the appendix) reports the list of airports included in the max k-core with their number of connections in the core, their rank in the local component, and their national rank according to their local degree centrality. One can see that top regional hubs outside the USA are not in the core. In addition, Nashville Airport, managed by Allegiant Air Lines, a low-cost airline, and Portland Airport, operated by Alaska Air Lines, is also absent from the core. In contrast, US domestic airports of minor importance in the local component but crucial at the national level integrate the core. They are mainly located on the East coast of the United States. The presence of the Lester B Pearson Airport and the Cancun Airport in the core reflects their high integration in the US transportation network. Indeed, Toronto is a main entry point to the US, and Cancun is a valued holiday destination in the US.

The core of the European component includes 65 airports. This result is close to the result reported in [Lordan, 2017]. Each airport in this core has at least 31 connections. In contrast with the North and Central America-Caribbean component, these airports are scattered in 27 countries. The European component is not very large compared to the others. One can consider it as a single geographic and economic space facilitating the connection of the member countries. The proportion of airports by country is not homogeneous. Germany, France, United Kingdom, Spain, and Italy dominate the core. Indeed, those are the major countries in Europe. The list of airports in the core correlates well with the top regional hubs of the component (See Table 4.11 in the appendix). Indeed, 59 airports ranked in the top 65 regional hubs belong to the core. The six missing airports are in Belgium (Brussels South Charleroi Airport), Italy (Il Caravaggio Airport), United Kingdom (Birmingham Airport, Glasgow Airport), Portugal (Faro Airport), and Netherlands (Eindhoven Airport). Although well-connected, these airports operate in the short and medium-haul segment, and they do not serve all of Europe. Six airports in the core

are not included in 65 first regional hubs : Ben Gurion Airport, the first airport in Israel, Naples Airport, and Catania-Fontanarossa Airport in Italy, the Bordeaux-Merignac Airport and the Toulouse-Blagnac Airport in France, and the Menara Airport in Marrakech, Morocco. Except for Israel with more links with Europe than with the Middle East for political issues, those are mainly attractive regional tourist destinations (See Table 4.11 in the appendix).

The core of the East and Southeast Asia component includes 38 airports. Each airport has at least 24 links with the others. One observes a very similar behavior to the one encountered in the North and Central America-Caribbean component. Indeed, China, with 33 airports, dominates the core (See Table 4.12 in the appendix). These airports are in provincial and municipal capitals. Quite naturally, its "satellites" in the region belong to the core. Indeed, four airports are located in Hong Kong (Hong Kong Airport), Singapore (Singapore Changi Airport), Taiwan (Kaohsiung Airport and Taiwan Taoyuan Airport). Incheon Airport (South Korea) completes the core. Indeed, there is a vast community of Chinese people in Korea and strong economic and political relations between the two countries. Comparison with the top 38 regional airports (see Table 4.12 in the appendix) shows that South-East Asia's leading regional airports (Malaysia, Thailand, Philippines, Vietnam) do not share enough links to belong to the core. It is also the case for Japan's main airports due to the political divide.

In the Africa-Middle East-Southern Asia component, the core contains 26 airports distributed across twelve countries (See Table 4.13 in the appendix). At least, each airport in the core has 13 connections. These airports are essentially in the Middle East and Southern Asia. Indeed, the Arabic peninsula includes ten airports ; one is in Jordan and one in Lebanon. Eight are in India, three in Pakistan, and one in Sri Lanka. There is no African airport in the core, except for Cairo, predominantly linked to the Arab world. There is also no airport from Iran. It must be related to the geopolitical context. The core airports reveal the significant air traffic between the Middle East and Southern Asia. Comparison with the top regional 26 airports reported in table 4.13 shows that 20 airports in the core belong to the 26 top regional hub set. Leading regional airports of Iran (Mehrabad Airport and Mashhad airport), Ethiopia (Addis Ababa Bole Airport), Kenya (Jomo Kenyatta Airport), and Nigeria (Murtala Muhammed Airport) have less than 12 connections with the core airports. Consequently, Africa is not part of the component core.

There are eight airports in the core of the Oceania component (See Table 3.14 in the appendix). Each of them has seven connections. Thus, the core is a complete graph ($K7$). Once again, a country dominates the core. Indeed, all airports are in Australia. Three top regional hubs located in New Zealand (Auckland Airport), Papua New Guinea (Port Moresby Jacksons Airport), and French Polynesia (Faa'a Airport) are not in the core airports.

The South America component behaves similarly. Its core includes 16 airports situated in Brazil (See Table 4.14 in the appendix). These airports are mainly located on the East coast. The less connected core member has eight links. Comparisons with the top regional hubs reported in Table 4.14 show that six regional hubs do not have enough links with Brazilian airports to belong to the core. These are the airports from Argentina (Jorge Newbery Airpark, and Ministro Pistarini Airport), Columbia (El Dorado Airport), Peru (Jorge Chávez Airport), Chile

(Comodoro A M Benítez Airport) and Brazil (Tancredo Neves Airport).

The core of the Russia-Central Asia-Transcaucasia component is also dominated by a single country Russia (See Table 3.16 in the appendix). Indeed, 14 airports are in Russia, 2 in Kazakhstan, and one in the capital cities of Uzbekistan, Tajikistan, Armenia, and Azerbaijan. Each airport shares at least nine connections with the core. Those ex-USSR countries have strong political relations that translate into solid interconnections between their metropolises. Note that four of the 19 top regional hubs from Russia are not in the core airports (Yakutsk Airport, Roshchino Airport, Irkutsk Airport, and Khabarovsk-Novy Airport).

To summarize, results show that leading countries widely influence the core of the local components. Indeed, the USA, China, Australia, Russia, Brazil dominate their region. The core reveals meaningful information about the national air traffic with its foreign satellite destinations. The core of the European component is more homogeneously distributed between various countries, even though major countries dominate the core. Finally, it appears that in the core of the Africa-Middle-East Southern -Asia component, Cairo is the unique core airport in Africa. The air integration concerns the Middle-east and Southern Asia essentially.

9.2 Exploring the max k-core in the large global components

Figure 4.13.B shows the core airports of the large global component. One can see that they cover the different regions of the world, except Oceania and Africa. Five global core airports are in North and Central America-Caribbean, nine in Europe, five in East and Southeast Asia, four in Africa-Middle East-Southern Asia, one in South America, and one in Russia-Central Asia-Transcaucasia. The maximum core of the global component contains 24 airports in 18 countries. These airports are mainly located in the capital of the countries and are managed by the largest airlines (See Table 4.15 in the appendix). Six of the top 24 inter-regional hubs are not at the core of the global component. These airports are in the United States (Newark Liberty), Spain (Adolfo S M–Barajas Airport), Thailand (Suvarnabhumi Airport), Singapore (Singapore Changi Airport), Malaysia (Kuala Lumpur Airport), and Russia (Pulkovo Airport). Although they have many links in the global component, most of these connections are not with major core airports. In contrast, some airports have fewer connections, but their links are with the core airports. It is the case for two airports in the North America-Caribbean region (Chicago O'Hare Airport, Los Angeles Airport, and Washington Dulles Airport), Switzerland (Zurich Airport), one in India (Indira Gandhi Airport), and one in Brazil (Guarulhos-Governador A F Montoro Airport). The less connected airports in the core have at least 15 long destinations throughout the world. It compares favorably with the core of a number of local components (Africa-Middle East-Southern Asia, Oceania, South America, Russia-Central Asia-Transcaucasia). Overall, Europe, North and Central America Caribbean, and East and Southeast Asia are particularly well connected.

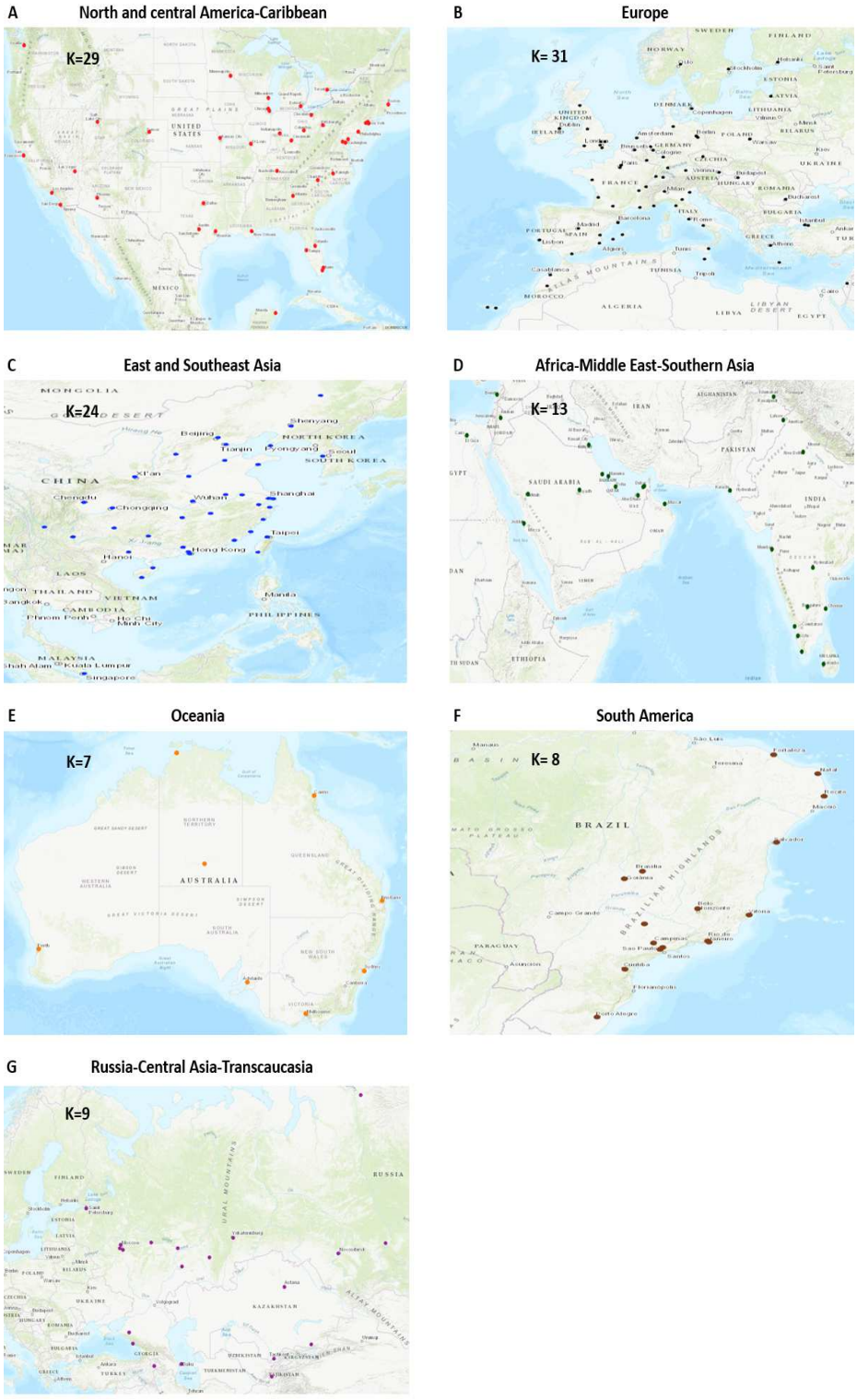


FIGURE 3.15 – Core of the large local components. The dots represent the airports, and K is the maximum k-core. The core of the North and Central America-Caribbean component contains 42 airports. The core of Europe contains 65 airports. The core of East and Southeast Asia contains 38 airports. The core of Africa-Middle East-Southern Asia contains 26 airports. The core of Oceania contains 8 airports. The core of South America contains 16 airports. The core of Russia-Central Asia-Transcaucasia includes 19 airports.

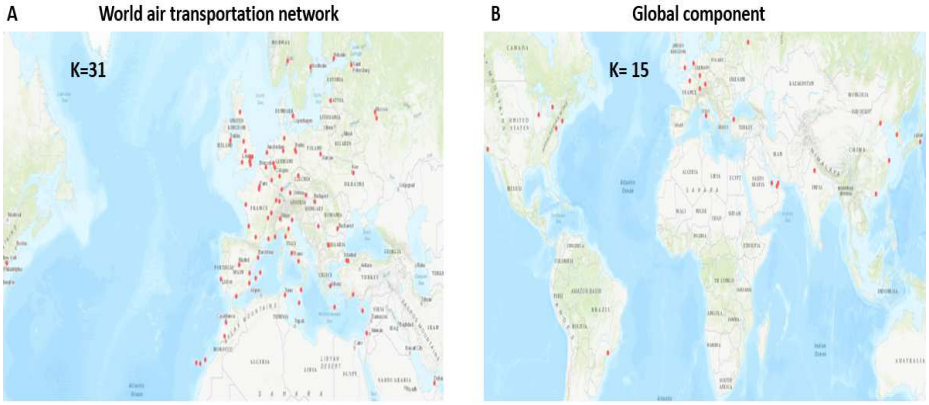


FIGURE 3.16 – Figure A presents the 79 airports in the maximum k-core of the world air transportation network. Figure B presents the 24 airports in the maximum k-core of the large global component. The red dots are the airports and K is the maximum k-core.

9.3 Comparison of max k-core of the components with the max k-core of the world air transportation network

This section compares the core of the world air transportation network with the core of the components. Figure 4.13.A shows the core airports of the world air transportation network. It contains 79 airports listed in Table 3.18 in the appendix. Similar to the core of the European component, each airport has at least 31 links. Only five airports are not in the European local component. One is in the USA (John F Kennedy), one is in the Middle East (Dubai), and three are in Russia (Sheremetyevo, Pulkovo, and Domodedovo). Note that these top inter-regional hubs are in the gateway cities of their region. It also includes all the airports of the European local components. In addition, European capitals (Kyiv, Sofia, Belgrade), tourist destinations (Antalya, Larnaca, Tunis, Fuerteventura, Catania), and Hannover airport in Germany join the core. Note that Hannover is the base of low-cost companies serving metropolitan and leisure destinations.

Once again, these results reveal the importance of the component structure decomposition. Indeed, the core of the world transportation network brings to fore Europe, and it blurs the rest of the world. In contrast, the component structure allows dealing with the non-homogeneous structure of the world transportation networks, highlighting the similarities and the differences between the components.

10 Conclusion

In this work, we present a mesoscale structure called the component structure. Each dense area of the network forms a local component. The subnetworks of nodes and links joining the local components are the global components. Although quite simple, this alternative mesoscopic representation presents multiple advantages. The exploration of the world air transportation network illustrates its usefulness. First of all, it allows partitioning the world air transportation

network into meaningful regional entities. Indeed, similarly with communities, the local components do not partition the world according to strict geographical considerations but on the internal exchanges in a region.

Experiments reveal seven large local components corresponding to geographical area (North and Central America-Caribbean, Europe, East-Southeast Asia, Africa-Middle East-Southern Asia, Oceania, South America, and Russia-Central Asia-Transcaucasia) with blurred boundaries. Indeed, for example, some airports in North Africa belong to the European component, because they share more links with Europe than with Africa. Twenty small local components generally cover remote parts of a single country. There is one large global component distributed over the world. Eight small global components join a large local component with neighboring small local components or close airports at the periphery of their local components. The analysis of the component structure reveals the regional diversity of the world air transportation network. Nevertheless, the large components share some common characteristics. They are disassortative, small world, and have a heavy-tail degree and degree-degree distance distributions. A star-based topology generally characterizes the small components.

The comparative analysis of the degree centrality in the various components allows disentangling the diverse role of airports. Local component centrality measure highlights the regional importance of the airports, while the global component allows quantifying its influence in inter-regional exchanges. Note that some major airports are pretty active at both levels.

In the same vein, the core analysis reveals that considering the air transportation network as a whole can lead to misleading interpretation. Indeed, as Europe is more densely connected than the rest of the world, most airports in the core are in Europe. Failing to take into account the regional variations blur the picture. In contrast, core analysis of the components proves very instructive about various regional and inter-regional situations.

In the next chapter, we explore the components of the weighted world air transportation network in order to compare the topological properties with the unwweighted network.

11 Appendix

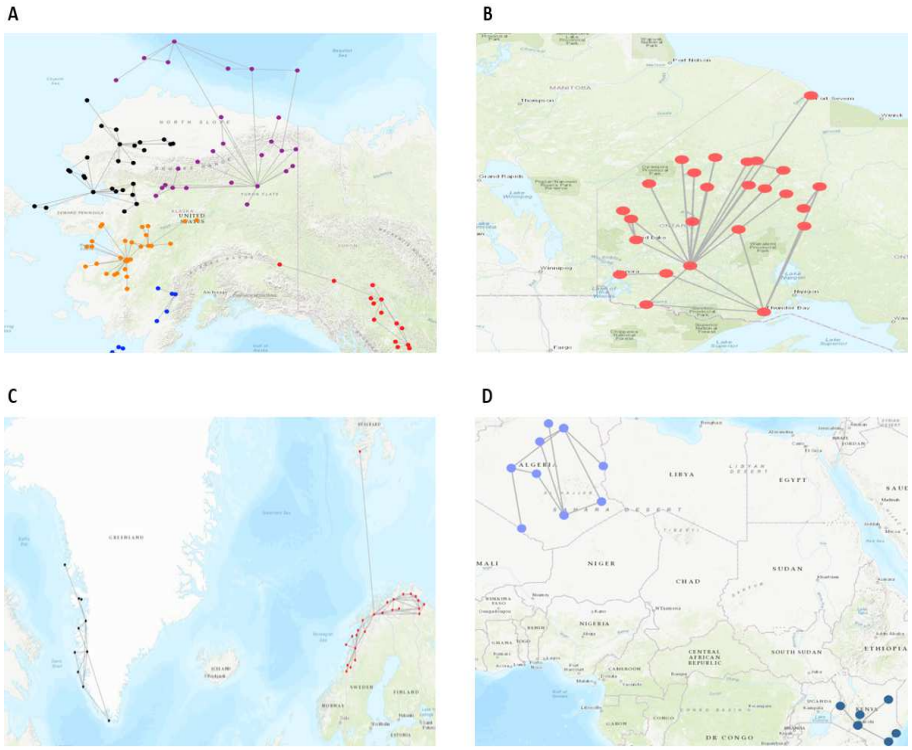


FIGURE 3.17 – Small local components (size greater than 5 nodes). (A) Small components located in Alaska (USA). (B) small components in Ontario (Canada). (C) Small components in Green-Land and Norway (Europe). (D) Small component in Algeria and Kenya (Africa).

TABLE 3.10 – Core airports of the North and Central America-Caribbean component (42). Airports not in the top 42 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 42 regional hubs and not in the k-core are in red.

Airport	City	Country	Core degree	Local rank	National rank
Hartsfield J Atlanta	Atlanta	United States	41	1	1
Dallas Fort Worth	Dallas-Fort Worth	United States	40	2	2
Chicago O'Hare	Chicago	United States	40	3	3
Denver	Denver	United States	41	4	4
George Bush Houston	Houston	United States	39	5	5
Charlotte Douglas	Charlotte	United States	41	6	6
McCarran	Las Vegas	United States	39	7	7
Minneapolis-St P/W-Chamberlain	Minneapolis	United States	41	8	8
Detroit M W County	Detroit	United States	41	9	9
Newark Liberty	Newark	United States	39	10	10
Los Angeles	Los Angeles	United States	40	11	11
Philadelphia	Philadelphia	United States	41	12	12
Lester B. Pearson	Toronto	Canada	38	13	13
Miami	Miami	United States	35	14	14
John F Kennedy	New York	United States	35	15	15
Orlando	Orlando	United States	40	16	16
Washington Dulles	Washington	United States	37	17	17
Salt Lake City	Salt Lake City	United States	31	18	18
Ronald Reagan Washington	Washington	United States	37	19	19
Phoenix Sky Harbor	Phoenix	United States	39	20	20
Fort Lauderdale Hollywood	Fort Lauderdale	United States	38	21	21
General E L Logan	Boston	United States	41	22	22
San Francisco	San Francisco	United States	37	23	23
Seattle Tacoma	Seattle	United States	31	25	24
Chicago Midway	Chicago	United States	31	26	25
La Guardia	New York	United States	29	27	26
Baltimore/W T Marshall	Baltimore	United States	38	28	27
Cancun	Cancun	Mexico	32	29	2
Tampa	Tampa	United States	36	30	28
St Louis Lambert	St. Louis	United States	39	32	29
San Diego	San Diego	United States	30	35	31
Cleveland Hopkins	Cleveland	United States	35	37	32
Nashville	Nashville	United States	36	38	34
Cincinnati N Kentucky	Cincinnati	United States	33	40	35
Kansas City	Kansas City	United States	35	42	36

TABLE 3.10 – Core airports of the North and Central America-Caribbean component (42). Airports not in the top 42 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 42 regional hubs and not in the k-core are in red.

Airport	City	Country	Core degree	Local rank	National rank
Austin Bergstrom	**Austin**	**United States**	**30**	**45**	**38**
Louis A N Orleans	**New Orleans**	**United States**	**35**	**47**	**39**
Raleigh-durham	**Raleigh-durham**	**United States**	**34**	**48**	**40**
Pittsburgh	**Pittsburgh**	**United States**	**31**	**50**	**41**
Indianapolis	**Indianapolis**	**United States**	**30**	**53**	**44**
John Glenn Columbus	**Columbus**	**United States**	**29**	**55**	**45**
General Mitchell	**Milwaukee**	**United States**	**29**	**64**	**50**
Licenciado Benito Juarez	Mexico City	Mexico	-	24	1
Montreal/Pierre Elliott Trudeau	Montreal	Canada	-	31	2
Vancouver	Vancouver	Canada	-	33	3
Portland	Portland	United States	-	34	30
Calgary	Calgary	Canada	-	36	4
William P Hobby	Houston	United States	-	38	33
Tocumen	Panama City	Panama	-	41	1

TABLE 3.11 – Core airports of the European component (65). Airports not in the top 65 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 65 regional hubs and not in the k-core are in red.

Airport	City	Country	Core degree	Local rank	National rank
Schiphol	Amsterdam	Netherlands	62	1	1
Munich Airport	Munich	Germany	61	2	1
London Stansted	London	UK	34	3	1
Barcelona	Barcelona	Spain	62	4	1
Frankfurt	Frankfurt	Germany	54	5	2
London Gatwick	London	UK	52	6	2
Charles de Gaulle	Paris	France	56	7	1
Dublin	Dublin	Ireland	58	8	1
Duesseldorf	Duesseldorf	Germany	48	9	3
Ataturk	Istanbul	Turkey	48	10	1
Brussels	Brussels	Belgium	57	11	1
Palma De Mallorca	Palma de Mallorca	Spain	42	12	2
Manchester	Manchester	UK	49	13	3
Vienna	Vienna	Austria	49	14	1
Malaga	Malaga	Spain	49	15	3
Leonardo da V–F	Rome	Italy	57	16	1
Madrid–Barajas	Madrid	Spain	58	17	4
Kastrup	Copenhagen	Denmark	52	18	1
S Arlanda	Stockholm	Sweden	46	19	1
Paris-Orly	Paris	France	39	20	2
Zurich	Zurich	Switzerland	54	21	1
Berlin-Tegel	Berlin	Germany	47	22	4
Alicante	Alicante	Spain	32	23	5
E. Venizelos	Athens	Greece	44	24	1
Oslo Lufthavn	Oslo	Norway	42	25	1
Geneva Cointrin	Geneva	Switzerland	53	26	2
Cologne Bonn	Cologne	Germany	37	27	5
London Luton	London	UK	34	28	4
Edinburgh	Edinburgh	UK	43	29	5
Nice-Côte d'Azur	Nice	France	46	30	3
Hamburg	Hamburg	Germany	47	31	6
Stuttgart	Stuttgart	Germany	33	32	7
London Heathrow	London	UK	47	33	6
H. Delgado	Lisbon	Portugal	51	34	1
Tenerife South	Tenerife	Spain	38	36	6
Vaclav Havel	Prague	Czech Republic	52	37	1
Marseille	Marseille	France	36	38	4
Gran Canaria	Gran Canaria	Spain	38	39	7
Malpensa	Milano	Italy	51	40	2
Saint-Exupéry	Lyon	France	37	42	5

TABLE 3.11 – Core airports of the European component (65). Airports not in the top 65 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 65 regional hubs and not in the k-core are in red.

Airport	City	Country	Core degree	Local rank	National rank
Sabiha Gökçen	Istanbul	Turkey	31	43	2
Bristol	Bristol	UK	34	44	7
Warsaw Chopin	Warsaw	Poland	33	46	1
Liszt Ferenc	Budapest	Hungary	46	47	1
Basel-Mulhouse	Mulhouse	France	40	48	6
Malta	Malta	Malta	39	49	1
Venice Marco Polo	Venice	Italy	42	51	5
Berlin-Schonefeld	Berlin	Germany	35	52	8
G Marconi	Bologna	Italy	36	53	6
Pisa	Pisa	Italy	31	54	7
Ibiza	Ibiza	Spain	31	55	8
Helsinki Vantaa	Helsinki	Finland	35	56	1
Henri Coandaf	Bucharest	Romania	34	57	1
Riga	Riga	Latvia	34	58	1
Valencia	Valencia	Spain	31	59	9
N Kazantzakis	Heraklion	Greece	31	60	2
Nantes Atlantique	Nantes	France	34	62	7
Mohammed V	Casablanca	Morocco	31	63	1
Luxembourg-Findel	Luxemburg	Luxemburg	31	65	1
Ben Gurion	**Tel-aviv**	**Israel**	**38**	**70**	**1**
Naples	**Naples**	**Italy**	**35**	**72**	**6**
Bordeaux-Merignac	**Bordeaux**	**France**	**31**	**76**	**8**
Catania-Fontanarossa	**Catania**	**Italy**	**35**	**77**	**7**
Menara	**Marrakech**	**Morocco**	**32**	**80**	**2**
Toulouse-Blagnac	**Toulouse**	**France**	**31**	**82**	**10**
Brussels South Charleroi	Charleroi	Brussels	-	35	2
Il Caravaggio	Bergamo	Italy	-	41	3
Birmingham	Birmingham	UK	-	45	8
Faro	Faro	Portugal	-	50	2
Glasgow	Glasgow	UK	-	61	9
Eindhoven	Eindhoven	Netherlands	-	64	2

TABLE 3.12 – Core airports of the East and Southeast Asia component (38). Airports not in the top 38 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 38 regional hubs and not in the k-core are in red.

Airport	City	Country	Core degree	Local rank	National rank
Beijing Capital Airport	Beijing	China	36	1	1
Guangzhou Baiyun	Guangzhou	China	35	2	2
Shanghai Pudong	Shanghai	China	33	3	3
Chengdu Shuangliu	Chengdu	China	36	4	4
Taiwan Taoyuan	Taipei	Taiwan	35	5	1
Kunming Changshui	Kunming	China	35	6	5
Incheon	Seoul	South Korea	29	7	1
Hong Kong	Hong Kong	Hong Kong	36	8	1
Shenzhen Bao'an	Shenzhen	China	34	9	6
Chongqing Jiangbei	Chongqing	China	36	10	7
Xi'an Xianyang	Xi'an	China	33	11	8
Singapore Changi	Singapore	Singapore	24	12	1
Hangzhou Xiaoshan	Hangzhou	China	29	13	12
Xiamen Gaoqi	Xiamen	China	33	14	13
Changsha Huanghua	Changcha	China	32	17	14
Shanghai Hongqiao	Shanghai	China	26	19	15
Wuhan Tianhe	Wuhan	China	32	21	16
Taoxian	Shenyang	China	26	22	17
Zhoushuizi	Dalian	China	25	23	18
Liuting	Qingdao	China	31	24	19
Nanjing Lukou	Nanjing	China	29	25	20
Zhengzhou Xinzheng	Zhengzhou	China	30	26	21
Taiping	Harbin	China	24	27	22
Tianjin Binhai	Tianjin	China	29	29	23
Fuzhou Changle	Fuzhou	China	29	32	24
Longdongbao	Guiyang	China	29	34	25
Taiyuan Wusu	Taiyuan	China	28	35	26
Sanya Phoenix	Sanya	China	28	37	27
Nanning Wuxu	Nanning	China	28	38	28
Ningbo Lishe	**Ninbo**	**China**	**26**	**41**	**29**
Nanchang Changbei	**Nanchang**	**China**	**29**	**44**	**30**
Kaohsiung	**Kaohsiung**	**Taiwan**	**24**	**45**	**31**
Haikou Meilan	**Haikou**	**China**	**28**	**46**	**32**
Yaoqiang	**Jinan**	**China**	**27**	**50**	**33**
Hefei Luogang	**Hefei**	**China**	**26**	**52**	**34**
Wenzhou Longwan	**Wenzhou**	**China**	**24**	**53**	**35**
Guilin Liangjiang	**Guilin**	**China**	**26**	**55**	**36**
Lijiang	**Lijiang**	**China**	**24**	**61**	**37**

TABLE 3.12 – Core airports of the East and Southeast Asia component (38). Airports not in the top 38 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 38 regional hubs and not in the k-core are in red.

Airport	City	Country	Core degree	Local rank	National rank
Suvarnabhumi	Kuala Lumpur	Malaysia	-	15	1
Suvarnabhumi	Bangkok	Thailand	-	16	2
Tokyo Haneda	Tokyo	Japan	-	18	1
Ninoy Aquino	Manila	Philippines	-	20	1
Don Mueang	Bangkok	Thailand	-	28	2
Narita	Tokyo	Japan	-	28	2
Kansai	Osaka	Japan	-	31	3
Tan Son Nhat	Ho Chi Minh City	Vietnam	-	33	1
Noi Bai	Hanoi	Vietnam	-	36	2

TABLE 3.13 – Core airports of the Africa-Middle East-Southern Asia component (26). Airports not in the top 26 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 26 regional hubs and not in the k-core are in red.

Airport	City	Country	Core degree	Local rank	National rank
Dubai	Dubai	United Arab Emirates	22	1	1
King Abdulaziz	Jeddah	Saudi Arabia	23	2	2
Indira Gandhi	Delhi	India	19	3	1
Chhatrapati Shivaji	Mumbai	India	20	4	2
Sharjah	Sharjah	United Arab Emirates	23	6	2
King Khaled	Riyadh	Saudi Arabia	24	7	2
Hamad	Doha	Qatar	19	9	1
Muscat	Muscat	Oman	25	10	1
Kuwait	Kuwait	Kuwait	22	11	1
King Fahd	Dammam	Saudi Arabia	24	12	3
Abu Dhabi	Abu Dhabi	United Arab Emirates	23	13	3
Cairo	Cairo	Egypt	14	14	1
Bahrain	Bahrain	Bahrain	22	15	1
Kempegowda	Bangalore	India	16	17	3
Queen Alia	Amman	Jordan	15	20	1
Chennai	Madras	India	18	21	4
Jinnah	Karachi	Pakistan	14	22	1
Prince Mohammad Bin Abdulaziz	Madinah	Saudi Arabia	16	24	3
Beirut Rafic Hariri	Beirut	Lebanon	13	25	1
Rajiv Gandhi	Hyderabad	India	13	26	6
Bandaranaike Colombo	**Colombo**	**Sri Lanka**	**17**	**27**	**1**
Benazir Bhutto	**Islamabad**	**Pakistan**	**13**	**28**	**2**
Alama Iqbal	**Lahore**	**Pakistan**	**14**	**33**	**3**
Cochin	**Kochi**	**India**	**17**	**34**	**7**
Calicut	**Calicut**	**India**	**14**	**37**	**8**
Trivandrum	**Trivandrum**	**India**	**14**	**38**	**9**
Addis Ababa Bole	Addis Ababa	Ethiopia	-	5	1
Jomo Kenyatta	Nairobi	Kenya	-	8	2
Mehrabad	Teheran	Iran	-	16	1
Mashhad	Mashhad	Iran	-	18	2
Murtala Muhammed	Lagos	Nigeria	-	19	1
Netaji Subhash Chandra Bose	Kolkata	India	-	23	5

TABLE 3.14 – Core airports of the Oceania component (8). Airports not in the top 8 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 8 regional hubs and not in the k-core are in red.

Airport	City	Country	Core degree	Local rank	National rank
Sydney Kingsford Smith	Sydney	Australia	7	1	1
Brisbane	Brisbane	Australia	7	2	2
Melbourne	Melbourne	Australia	7	4	3
Cairns	Cairns	Australia	7	6	4
Perth	Perth	Australia	7	7	5
Adelaide	**Adelaide**	**Australia**	**7**	**10**	**6**
Darwin	**Darwin**	**Australia**	**7**	**15**	**7**
Alice Springs	**Alice Springs**	**Australia**	**7**	**18**	**10**
Auckland	Auckland	New Zealand	-	3	1
Port Moresby Jacksons	Port Moresby	Papua New Guinea	-	5	1
Faa'a	Papeete	French Polynesia	-	8	1

TABLE 3.15 – Core airports of the South America component (16). Airports not in the top 16 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 16 regional hubs and not in the k-core are in red.

Airport	City	Country	Core degree	Local rank	National rank
Guarulhos-Governador A F Montoro	Sao Paulo	Brazil	13	1	1
Viracopos	Campinas	Brazil	13	3	2
Presidente Juscelino Kubistschek	Brasilia	Brazil	14	5	3
Rio Galeao–Tom Jobim	Rio De Janeiro	Brazil	14	6	4
Tancredo Neves	Belo Horizonte	Brazil	15	7	5
Congonhas	Sao Paulo	Brazil	10	10	6
Deputado Luiz E Magalhaes	Salvador	Brazil	12	11	7
Salgado Filho	Porto Alegre	Brazil	8	13	8
Afonso Pena	Curitiba	Brazil	10	14	9
Santos Dumont	Rio De Janeiro	Brazil	11	15	10
Pinto Martins	**Fortaleza**	**Brazil**	**8**	**17**	**12**
Guararapes-Gilberto Freyre	**Recife**	**Brazil**	**8**	**18**	**13**
Santa Genoveva	**Goiania**	**Brazil**	**8**	**25**	**16**
Eurico de Aguiar Salles	**Vitoria**	**Brazil**	**8**	**26**	**17**
Governador Aluízio Alves	**Natal**	**Brazil**	**8**	**29**	**18**
Leite Lopes	**Ribeirao Preto**	**Brazil**	**8**	**32**	**20**
El Dorado	Bogota	Colombia	-	2	1
Jorge Newbery Airpack	Buenos Aires	Argentina	-	4	1
Jorge Chavez	Santiago	Chile	-	9	1
Ministro Pistarini	Buenos Aires	Argentina	-	12	2
Tancredo Neves	Belo Horizon	Brazil	-	7	5

TABLE 3.16 – Core airports of the Russia-Central Asia- Transcaucasia component (19). Airports not in the top 16 regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 19 regional hubs and not in the k-core are in red.

Airport	City	Country	Core degree	Local rank	National rank
Domodedovo	Moscow	Russia	17	1	1
Pulkovo	St. Petersburg	Russia	18	2	2
Sheremetyevo	Moscow	Russia	15	3	3
Tolmachevo	Novosibirsk	Russia	14	4	4
Koltsovo	Yekaterinburg	Russia	18	5	5
Tashkent	ashkent	Uzbekistan	16	6	1
Vnukovo	Moscow	Russia	12	7	6
Astana	Tselinograd	Kazakhstan	9	8	1
Almaty	Alma-ata	Kazakhstan	10	9	2
Yemelyanovo	Krasnoyarsk	Russia	13	10	7
Zvartnots	Yerevan	Armenia	11	13	1
Krasnodar Pashkovsky	Krasnodar	Russia	11	11	8
Dushanbe	Dushanbe	Tajikistan	12	15	1
Ufa	Ufa	Russia	9	16	9
Heydar Aliyev	Baku	Azerbaijan	10	18	1
Sochi	Sochi	Russia	12	19	9
Kazan	**Kazan**	**Russia**	**9**	**22**	**11**
Nizhny Novgorod Strigino	**Nizhniy Novgorod**	**Russia**	**9**	**23**	**12**
Kurumoch	**Samara**	**Russia**	**10**	**24**	**13**
Norilsk-Alykel	**Norilsk**	**Russia**	**9**	**26**	**14**
Yakutsk	Yakutsk	Russia	-	12	9
Roshchino	Tyumen	Russia	-	14	10
Irkutsk	Irkutsk	Russia	-	17	13
Khabarovsk-Novy	Khabarovsk	Russia	-	20	18

TABLE 3.17 – Core airports of the global component (24). Airports not in the top 24 inter-regional hubs are in bold. The core degree of a node is its number of links in the maximum k-core. Global rank is the rank in the large local component. Airports in the top 24 inter-regional hubs and not in the k-core are in red.

Airport	City	Country	Core degree	Global rank
Frankfurt am Main	Frankfurt	Germany	16	1
Charles de Gaulle	Paris	France	16	2
Atatürk	Istanbul	Turkey	16	3
London Heathrow	London	United Kingdom	16	4
Dubai	Dubai	United Arab Emirates	19	5
Sheremetyevo	Moscow	Russia	19	6
Amsterdam Schiphol	Amsterdam	Netherlands	16	7
John F Kennedy	New York	United States	19	9
Beijing Capital	Beijing	China	18	10
Hamad	Doha	Qatar	17	11
Narita	Tokyo	Japan	17	13
Leonardo da Vinci–Fiumicino	Rome	Italy	16	14
Incheon	Seoul	South Korea	18	15
Hong Kong	Hong Kong	Hong Kong	17	16
Lester B. Pearson	Toronto	Canada	17	19
Abu Dhabi	Abu Dhabi	United Arab Emirates	17	22
Munich	Munich	Germany	15	23
Guarulhos-Governador A F Montoro	**Sao Paulo**	**Brazil**	**16**	**26**
Shanghai Pudong	**Shanghai**	**China**	**17**	**28**
Zurich	**Zurich**	**Switzerland**	**15**	**29**
Los Angeles	**Los Angeles**	**United States**	**16**	**30**
Indira Gandhi	**Delhi**	**India**	**16**	**32**
Washington Dulles	**Washington**	**United States**	**16**	**38**
Chicago O'Hare	**Chicago**	**United States**	**17**	**39**
Domodedovo	Moscow	Russia	-	8
Suvarnabhumi	Bangkok	Thailand	-	12
Adolfo S Madrid–Barajas	Madrid	Spain	-	17
Pulkovo	St. Petersbourg	Russia	-	18
Singapore Changi	Singapoer	Singapore	-	20
Kuala Lumpur	Kuala Lumpur	Malaysia	-	21
Newark Liberty	Newark	United States	-	24

TABLE 3.18 – These airports are not in the core of the European local component, but they belong to the core of the world air transportation network. Airports in the European local component are in black. Airports from other local components are in red. The core degree is the number of links in the maximum k-core network. Worldwide rank is the rank in the world air transportation network. Local rank is the rank in the large local components. Global rank is the rank in the global component.

Airport	City	Country	Core degree	Worldwide rank	Local rank	Global rank
Dubai	Dubai	United Arab Emirates	40	10	1	5
John F Kennedy	New York	United States	32	14	15	9
Domodedovo	Moscow	Russia	35	16	1	8
Pulkovo	St. Petersburg	Russia	36	63	2	18
Sheremetyevo	Moscow	Russia	49	36	3	6
Boryspil	Kiev	Ukraine	35	145	86	91
Tunis Carthage	Tunis	Tunisia	32	165	103	86
Sofia	Sofia	Bulgaria	31	218	96	305
Fuerteventura	Fuerteventura	Spain	31	158	65	-
Antalya	Antalya	Turkey	31	137	65	156
Larnaca	Larnaca	Cyprus	33	156	88	111
Belgrade Nikola Tesla	Belgrade	Serbia	31	178	81	238
Hannover	Hannover	Germany	32	171	75	476
Catania-Fontanarossa	Catania	Italy	37	177	75	440

TABLE 3.19 – Small global components. The Local component number corresponds to the number in Figure 4.3.

Airport	City	Country	Local component number
Postville	Postville	Canada	10 (Canada)
Goose Bay	Goose Bay	Canada	North and Central America-Caribbean
Val-d'Or	Val D'or	Canada	North and Central America-Caribbean
Chisasibi	Chisasibi	Canada	7 (Canada)
Waskaganish	Waskaganish	Canada	7 (Canada)
Natashquan	Natashquan	Canada	9 (Canada)
Sept-Iles	Sept-Iles	Canada	North and Central America-Caribbean
Lourdes de Blanc Sablon	Lourdes-De-Blanc-Sablon	Canada	North and Central America-Caribbean
St Augustin	St-Augustin	Canada	9 (Canada)
Fort Chipewyan	Fort Chipewyan	Canada	7 (Canada)
Fort McMurray	Fort Mcmurray	Canada	North and Central America-Caribbean
Timmins/Victor M. Power	Timmins	Canada	North and Central America-Caribbean
Moosonee	Moosonee	Canada	9 (Canada)
Nouadhibou	Nouadhibou	Mauritania	17 (Mauritania)
Gran Canaria	Gran Canaria	Spain	Europe
Dongshan	Hailar	China	East and Southeast Asia
Chita-Kadala	Chita	Russia	Russia-Central Asia-Transcaucasia

Chapter 4

Exploring the Regional Structure of the Worldwide Air Traffic and Route Networks [1]

Summary

The topological structure of the world air transportation network has been the subject of much research. However, to better understand the reality of air networks, one can consider the traffic, the number of passengers, or the distance between flights. This paper studies the weighted world air transportation network through the component structure, recently introduced in the network literature, by using the number of flights. The component structure is based on the community or multiple core-periphery structures and splits the network into local and global components. The local components capture the regional flights of these two mesoscopic structures (dense parts). The global components capture the inter-regional flights (links between the dense parts). We perform a comparative analysis of the world air transportation network and its components with their weighted counterparts. Moreover, we explore the strength and the s-core of these networks. Results display fewer local components well delimited and more global components covering the world than the unweighted world air transportation network. Centrality analysis reveals the difference between the top airports with high traffic and the top airports with high degrees. These difference is more pronounced in the global air network and the largest global component. Core analysis shows similitude between the s-core and the k-core for the local and global components, even though the latter includes more airports. For the world air network, the North and Central America-Caribbean airports dominate the s-core, whereas the European airports dominate the k-core.

1. The results contained in this chapter are submitted for a publication in the paper :
[Diop, 2022d] Issa Moussa Diop, Cherif Diallo, Chantal Cherifi, Hocine Cherifi. "Exploring the Regional Structure of the Worldwide Air Traffic and Route Networks". (submitted to IEEE Access Journal).

1 Introduction

The air transport system connects all countries in the world. This infrastructure has a direct impact on society and the global economy. Indeed, millions of people and goods transit through the air every day. Recently, we have seen how the COVID-19 pandemic spread rapidly because of air transport. Thus, understanding the air transportation system can help policymakers make decisions that can improve or affect this system. Network science provides a simple way to represent and understand this infrastructure. Thus, several studies are devoted to the air transportation network, including structure, dynamics, and robustness. In a previous study, we analyzed the world air transportation network through a new mesoscopic structure called the component structure[Diop, 2021]. A network contains two types of components. The dense parts of the network form the local components. The interactions between the local components are called the global components. Therefore, one must extract the dense areas to build the component structure. To do so, the community structure or the core-periphery structure are good candidates. Indeed, the communities constitute cohesive groups of nodes sparsely connected [Fortunato, 2016b ; Cherifi, 2019b ; Newman, 2004]. The core-periphery [Borgatti, 2000 ; Lee, 2014 ; Zhang, 2015] structure contains two groups of nodes (core and periphery) with three types of connections. The core nodes are tightly connected. The periphery nodes are almost not connected. The links between the core and periphery nodes are relatively dense. Networks can exhibit a multi-core-periphery structure [Kojaku, 2017a]. One can extract the dense parts using any algorithm to uncover the communities or the various cores to form the local components. In the world air transportation network, the local components capture the regional destinations, while the global components represent the inter-regional flights. In the previous study, we analyzed the unweighted world transportation network. Therefore, the results of our investigations are related to the network infrastructure. Indeed, it uses no information on the dynamic of the infrastructure. This simplification can lead to centrality anomalies [Kojaku, 2017a ; Rajeh, 2021] and hide critical information about the flow of flights and passengers in the infrastructure. Indeed, the traffic in the various routes can be pretty different, and failing to integrate these differences can lead to misleading conclusions.

This paper aims to extend and compare the infrastructure of the world air transportation network with its traffic. To this end, we incorporate the number of flights to weight the links of this network. Note that either number of flights or passengers during a given period quantifies the traffic between two destinations. As the component structure has shown promises in disentangling regional with inter-regional routes, we adopt this representation in this study. For the comparative purpose, we rely on the community structure framework to extract the components. Indeed, we can exploit community detection algorithms tailored to weighted networks to identify the components. It allows us to explore the dynamics of the different regional and inter-regional flights. Moreover, we do a comparative analysis with the results found in [Diop, 2021]. The main contributions of this paper are as follows :

1. This paper investigates the weighted world air transportation network at the macroscopic, mesoscopic, and microscopic levels, highlighting the advantages of the component structure.
2. We compare the weighted and unweighted world air transportation and its component structure at each level to understand better the differences between the infrastructure and traffic of the world air transportation network.
3. We perform the s-core analysis of the weighted global air network and its components.

The rest of the paper is organized as follows. Section 2 reports a review of related studies of the air transportation network. Section 3 describes the data and section 4 examines the network community structure. Section 5 reports the analysis of the component structure. Section 6 explores the topological properties of the local and global components. Section 7 present the topological properties of the world air

transportation network. Section 8 presents the results of a comparative analysis of the strength centrality of the components and the world air transportation network. Section 9 discusses the results of the core structure analysis. Finally, we conclude in section 9.

2 Literature review

The literature reports numerous studies of unweighted air transportation networks [Guimera, 2005 ; Zanin, 2013 ; Cheung, 2020 ; Bagler, 2008]. They cover national, regional, and worldwide networks and include investigations at the macroscopic, mesoscopic, and microscopic scales. All these networks share some common characteristics. They are generally small-world and scale-free. Moreover, some exhibit a community structure, and one can observe that the most connected cities do not have the largest betweenness [Guimera, 2005].

In contrast, the weighted air transportation network is not much studied. It is particularly true for the world air transportation network. In the following, we briefly report related studies of the national, regional, and worldwide weighted air transportation networks.

In [Bagler, 2008], the authors study the weighted airport network of India. The weight indicates the number of weekly flights between two airports. The distribution of strength reveals the heterogeneity of this network. Moreover, the author shows that the strength of a node correlates well with its degree. Comparing nodes' unweighted and weighted clustering coefficients indicates that the hubs tend to form interconnected groups. However, the airport network in India is disassortative. The weighted assortativity shows that the hubs with many flights are more connected.

In [Xu, 2008] the authors investigate the air transportation network of the United States from 2002 to 2005, by quarter. They explore 16 networks in this period using topological and weighted metrics. Cities are the nodes, and routes between two cities are the edges. They consider three types of weight : the non-stop distance between cities, the average passengers, and the average one-way fare. The degree and the strength distribution of these three attributes exhibit a scale-free behavior. Additionally, they are correlated. Then, high-degree nodes tend to have high-strength links. The analysis of the degree-degree correlation shows a rich-club phenomenon. Indeed, there is large traffic among the hubs. Moreover, the interconnected hubs make long distances at an accessible price due to the multiplication of point-to-point flights. Finally, the authors propose a weighted network model.

In [Hossain, 2017], the authors study the Australian airport network's structure and dynamic flow. The weights correspond to the number of flights between two airports. The authors find that the strengths of the nodes evolve in the same direction as the degree. Sydney airport, the most prominent hub, handles the highest fraction of traffic. The weighted clustering coefficient is lower than the unweighted. Consequently, the edges with low weights form the topological clustering. The weighted and unweighted degree-degree correlation reveals a disassortative behavior.

TABLE 4.1 – Basic topological properties of the world air transportation network and the global components. N is the network size. $|E|$ is the number of edges. d is the diameter. L is the average shortest path length. μ is the density. ζ and ζ^w are, respectively, the unweighted and weighted average clustering coefficients. λ and λ^w are respectively the unweighted and weighted assortativity, also called the degree-degree correlation coefficient. η is the hub dominance.

	N	$\|E\|$	d	L	μ	ζ	ζ^w	λ	λ^w	η
network	2734	16665	12	3,86	0,004	0,046	0,007	-0,046	0,048	0.09

In [Cheung, 2020], the authors analyze the characteristics of the Asian international passenger aviation

market in 2014 and 2018. The number of passengers weights the links between two airports. To conduct their analysis, they consider 28 Asian top airlines, with 7 Low-Cost Carriers and 21 Full-Service Carriers. The airports of the Low-Cost Carriers increased, and their degree distribution shows the evolution of the Low-Cost Carriers into a hub-and-spoke system. Developing countries like China, Vietnam, etc., influence the air transport network of this region. To characterize the airports' influence, they explore eight centrality measures : the degree, mean association (normalized strength), betweenness, weighted betweenness, page rank, weighted page rank, reverse page rank, and weighted reverse page rank. For most centrality measures, Changi Airport, Incheon Airport, Narita Airport, and Hong Kong Airport are in the top position in 2014 and 2018 [Cheung, 2020].

In [Barrat, 2004], the author proposes to extend the definitions of the clustering coefficient and the assortativity to a weighted network. They use the world air transportation network as a typical example. The link weight is the number of seats available on the flights between two airports. They show that the degree and strength distributions are heavy-tailed. Moreover, there is a linear relationship between a node's degree and strength. The network exhibits the rich-club phenomenon. Indeed, the interconnected neighbors of the hubs handle the most significant proportion of the traffic. The weighted degree-degree correlation is assortative.

To summarize, most of the previous works focus on weighted national air transportation networks. One is limited to a regional air transport network[Cheung, 2020]. The weighted world air network analysis concentrates only on the macroscopic properties [Barrat, 2004]. This paper investigates the weighted world air transportation network in light of its component structure. We perform an extensive comparative analysis of the route (unweighted) and traffic (weighted) networks at the macroscopic, mesoscopic, and microscopic levels.

3 Data and Tools

This section describes the data used to conduct our experiment. Then, two weighted community detection algorithms are applied to the data. The purpose is to compare their community structure. Finally, we also compare the communities from the unweighted and weighted world air transportation network.

3.1 Data

This network of 2734 nodes and 16665 links originates from FlightAware [Flightaware,]. It collects the flights between May 17, 2018, and May 22, 2018 [Alves, 2020]. Nodes represent airports, and links are direct flights between airports. The link weight is the number of flights between two airports. Table reftab :watn reports basic properties (unweighted and weighted). One needs at least 12 flights to reach the most distant airports. On average, four flights allow reach any destination. The network is not very dense, and globally, there are few triplets by airports with many flights ($C^w \geq C$). The unweighted assortativity reveals that the hubs tend to connect to the airport with few connections. In addition, the global weighted assortativity shows numerous flights between these hubs and these types of airports ($K^w \geq K$).

3.2 Tools

Here, we briefly recall the algorithm to extract the component structure. As it relies on the communities to form the local components, we describe two community detection algorithms that uncover these dense parts of the network. Finally, we present the evaluation measures.

3.2.1 Extracting the component structure

The process of extracting the component structure described in [Diop, 2021] consists of three steps :

1. Extracting the dense areas of the network by using a community or multiple core-periphery detection algorithms.
2. Extracting the local components by removing all the links between the dense areas.
3. Extracting the global components by removing all the links within the dense areas.

Note that a node can belong to the local and global components. In this paper, weighted community detection algorithms are used [Orman, 2011b ; Orman, 2011a]. Indeed, they take into account the dynamic flow in the network.

3.2.2 Extracting the communities

We use two popular community detection algorithms : Louvain[Blondel, 2008] and Combo[Sobolevsky, 2014]. They are weighted non-overlapping community detection algorithms based on modularity. Our purpose is to evaluate the impact of the community structure variations induced by the community detection algorithms on the component structure.

Louvain

Louvain constructs the communities in two phases, repeated iteratively. At first, each node is a community. Then, one evaluates the modularity after grouping neighboring nodes. The second step considers the group of nodes (community), maximizing the modularity as new nodes. The algorithm stops when there are no more modifications, and the modularity can no longer be maximized. Note that the definition of modularity depends on the nature of the network (unweighted or weighted).

Combo

Combo combines three strategies to uncover the community structure. Indeed, at the initial, all nodes form a community. Then, for several iterations, one can merge communities, split communities, or recombine nodes between communities according to the optimization of the objective functions until any gain is possible. One can use two objective functions (modularity and code length). In this paper, we use the algorithm based on modularity.

3.2.3 Quality metrics

We use three quality metrics to compare the Louvain and Combo weighted community structures.

Modularity

Modularity is the most popular. It compares the actual community structure with a null model without community structure. Its values range between -1 and 1. Several community detection algorithms aim to optimize modularity. The best partition is the one that is nearer to 1. The modularity of weighted networks[Blondel, 2008] is defined as follows :

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{4.1}$$

A_{ij} is the weight of link between i and j ; k_i is the strength of the node i ; c_i is the community that i belongs to ; the Kronecker delta $\delta(a, b)$ is 1 if $a = b$ and 0 otherwise ; and $m = \frac{1}{2} \sum_{ij} A_{ij}$.

Mixing Parameter

For a weighted network, the mixing parameter μ^w of a node i is the proportion of weight pointing outside its community. The mixing parameter of the network is the average of the nodes mixing parameters. The communities are well separated when the average mixing parameter is near 0. The smallest the

mixing parameter, the stronger the community structure. The mixing parameter of a network is defined as follows :

$$\mu^w = \frac{1}{n} \sum \mu_i^w \tag{4.2}$$

where $\mu_i^w = \frac{w^{ext}}{w_i}$, w^{ext} is the sum of weight from the node is i to the communiities that don't contain i, and w_i is the strength of the node i.

Normalized mutual information

The Normalized Mutual Information measures the similarity of two partitions. The partitions are similar when the NMI is near one and almost independent when close to 0. The NMI of two partitions P_1 and P_2 is defined as follows :

$$NMI(P_1, P_2) = 2 * \frac{I(P_1, P_2)}{H(P_1) + H(P_2)} \tag{4.3}$$

$I(P_1, P_2)$ is the Mutual Information between P_1 and P_2. $H(P)$ is the entropy of P. Here, the sum of entropy normalizes the Mutual Information.

Jaccard Index

The Jaccard Index to compare the similarity of two sets. It is defined as follows :

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{4.4}$$

When the two sets are identical, the Jaccard index equals 1. It is equal to zero if the two sets have no element in common.

TABLE 4.2 – Quality metrics of the community structures uncovered by Louvain and Combo community detection algorithms : Modularity, Mixing parameter, NMI

	Modularity	Mixing parameter	NMI
Louvain	0,47	0,043	0,87
Combo	0,47	0,046	

TABLE 4.3 – The Jaccard Index of the five similar communities uncovered by Louvain and Combo.

Community	Jaccard Index
North and Central America-Caribbean	0,94
Europe-Russia-Central Asia-Central Asia	0,98
East and Southeast Asia-Oceania	0,85
Africa-Middle East-Southern Asia	0,88
South America	0,85

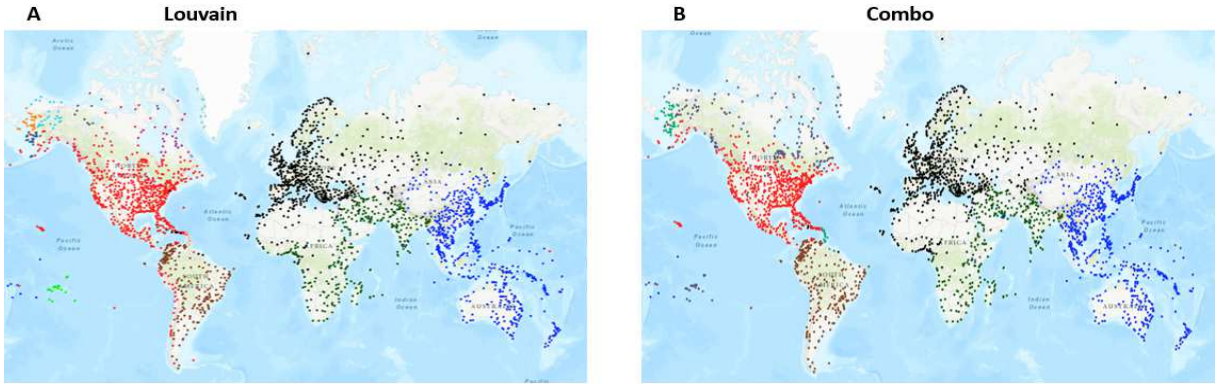


FIGURE 4.1 – A) represents the communities identified by the Louvain community detection algorithm. It includes seventeen communities. B) presents the communities uncovered by the Combo community detection algorithm. It contains seven communities. For both, each color represents a community. Similar communities are similar colors.

4 Community structure analysis

This section first compares the weighted community structures uncovered by Louvain and Combo. Then we compare the communities of the weighted and unweighted world air transportation network uncovered by Louvain. We use the modularity and the mixing parameter as quality measures of the community structures. We also perform a comparative qualitative evaluation highlighting the similitudes and differences between weighted and unweighted networks.

4.1 Comparing the weighted community structure uncovered by Louvain and Combo

The number of communities extracted from the networks by the algorithms is quite different. Louvain uncovers 17 communities. The largest contains 725 airports, whereas the smallest includes two airports. Combo identifies seven communities. The largest includes 703 airports, and the smallest consists of 70 airports. The table4.2 displays the quality metrics of the two community structures. Their modularity is identical. Its value of 0.47 indicates that the communities are dense with a medium proportion of inter-community links. The community structure of Louvain contains few connections between the communities with strong weights. In contrast, Combo reveals more inter-community links with smaller weights. In both cases, the mixing parameter values demonstrate that the communities are well-separated.

Although the Louvain algorithm uncovers more than two times more communities than the Combo algorithm, their community structures have numerous similitudes. Indeed, Combo group some communities of Louvain. The high value of the NMI (0.87) confirms their similarity. Five communities are very similar at first glance. They cover the same geographical areas. In Figure4.1, these areas correspond to the communities with the same color. They are in North and Central America-Caribbean, Europe-Russia-Central Asia-Central Asia, East and Southeast Asia-Oceania, Africa-Middle East-Southern Asia, and South America. In addition, the Table4.3 shows their Jaccard Index. Its value for all similar communities is greater than 0.85. For North and Central America-Caribbean and Europe-Russia-Central Asia-Central Asia, it is greater than 0.9, indicating a high similarity between the communities uncovered by the two algorithms. The twelve other communities from Louvain, mainly located in Canada and Alaska, are grouped into Combo. To sum up, although the community structures uncovered by Louvain and Combo differ, the largest communities have a lot in common. Therefore, in the rest of the paper, we adopt the communities

of Louvain to build the structure of the components used in subsequent analysis.

4.2 Comparing weighted and unweighted community structures uncovered by Louvain

Louvain uncovered fewer communities in the weighted network (17 communities) than in the unweighted one (27 communities). The weighted network contains five large communities and 12 small communities. In contrast, there are seven large and 20 small communities in the unweighted network. Three correspond to similar areas with the unweighted communities (North America, Africa-Middle East-Southern Asia, and South America). Moreover, their Jaccard Index with their unweighted counterpart reported in Table 4.4 is very high. The two other large communities in the weighted network regroup into two communities four communities in the unweighted network. The Europe-Russia-Central Asia community breaks into Europe and Russia-Central and Asia-Transcaucasia communities. In the same way, the community covering the East-Southeast Asia-Oceania region separates into the East-Southeast Asia community and Oceania. Weighted Louvain tends to group nodes linked by heavyweights in the same community. The Jaccard Index reported in Table 4.4 show that the regrouped weighted communities are very similar to the unweighted communities. These results reveal several flights between Europe and Russia and East-Southeast Asia and Oceania, even if the number of connections between these regions is limited. Apart from grouping communities of the unweighted network, the communities of the weighted network present some particular singularities. Indeed, some airports belong to communities far from their geographical area.

The airports in North and West Africa are in the Europe-Russia-Central Asia-Central Asia community in the weighted network, while they are in the Africa-Middle East-Southern Asia in the unweighted network. Indeed, these airports have more flights to Europe despite sharing more connections to the other airports in Africa and the Middle East. John F Kennedy airport in the United States is in the Europe-Russia-Central Asia-Central Asia community. Likewise, Frankfurt airport in Germany and London Heathrow in the United Kingdom are respectively in the Africa-Middle East-India region and the North and Central America-Caribbean area. These examples confirm that the communities of the weighted network correspond to areas of influence, while in the weighted network, they correspond to geographical regions.

The large communities of the weighted network include several small communities of the unweighted network. Moreover, like in the large communities, a few small communities of the unweighted network are grouped in the weighted network.

The modularity and the mixing parameters show a higher community structure strength for the unweighted network. Nevertheless, in both cases, the community structure strength is in a medium range indicating a clear community structure. The mixing parameter values corroborate these findings.

TABLE 4.4 – The Jaccard Index of the communities uncovered by the weighted and unweighted Louvain algorithm. Three are similar (North and Central America-Caribbean, Africa-Middle East-Southern Asia, and South America), and two are regrouped by the weighted algorithm. The East and Southeast Asia-Oceania regroup the East and Southeast Asia and the Oceania components. Europe-Russia-Central Asia-Central Asia regroups the Europe and Russia-Central Asia components.

Community	Jaccard Index
North and Central America-Caribbean	0,81
Europe-Russia-Central Asia-Central Asia	0.87
East and Southeast Asia-Oceania	0.97
Africa-Middle East-Sourthern Asia	0,87
South America	0,78

TABLE 4.5 – Quality metrics of the community structures uncovered by the unweighted and weighted Louvain community detection algorithms : Modularity, Mixing parameter, NMI

	Modularity	Mixing parameter	NMI
Weighted	0,47	0,14	0,78
Unweighted	0,63	0,13	

5 Component structure

We categorize the uncovered components as large or small. The large components include more than 100 airports and cover large geographical areas. In this section, we describe their features and compare them to the component structure of the unweighted network [Diop, 2021].

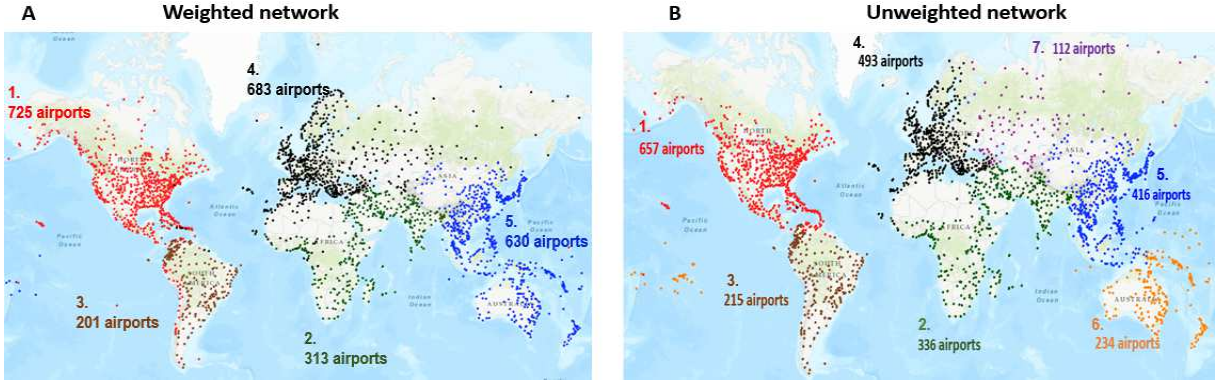


FIGURE 4.2 – A represents the airports in the large components of the weighted network. B represents the airports in the large components of the unweighted network. Each color is associated with a component. The North and central America-Caribbean component is red (1). The Africa-Middle East-India component is green (2). The South American component is brown (3). The Europe-Russia-Central Asia-Central Asia component is black (4). It splits in Europe in black (4) and Russia in purple (7) in the unweighted network. The East and Southeast Asia-Oceania component (5) breaks in East and Southeast Asia in blue (5) and Oceania in orange (6) in the unweighted network.

5.1 Local components

The local components are the dense parts of the networks. They correspond to the 17 communities uncovered by Louvain in the weighted network. There are five large and twelve small local components. The large local components do not reflect strict geographical divisions. They correspond more to political, cultural, historical, and economic divides. For example, some African airports in Morocco, Tunisia, and West Africa belong to the European component. It is because of the solid economic and historical ties these countries share with Europe. The small local components are in a single country (the United States, Canada, French Polynesia, Greenland, Israel, Australia, United Arab Emirates) or a few countries (Caribbean).

5.1.1 Large local components

The large local components cover 1) North and Central America-Caribbean (725 airports), 2) Europe-Russia-Central Asia-Central Asia (683 airports), 3) East-Southeast Asia-Oceania (630 airports), 4) Africa-Middle East-Southern Asia (313 airports), and 5) South America (201 airports). Altogether, they regroup more than 93% of the world's airports. One can distinguish two typical behaviors when comparing the

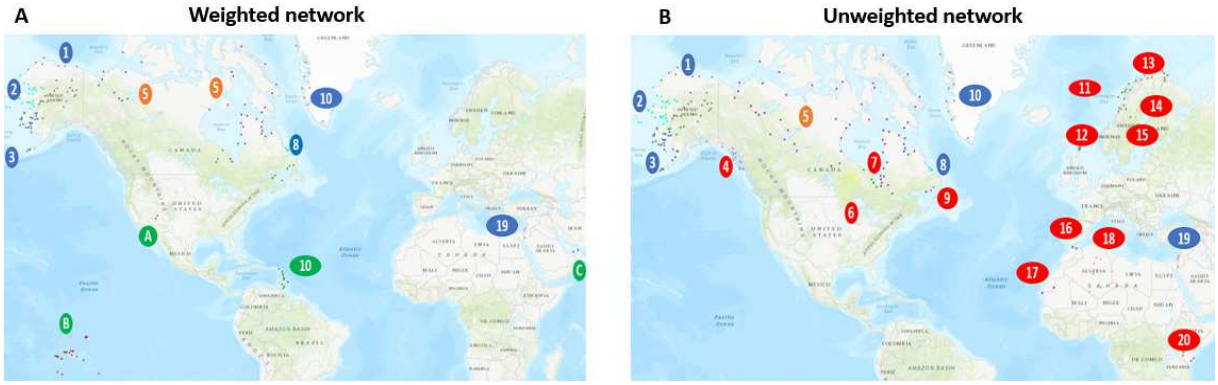


FIGURE 4.3 – Small local components in the weighted (A) and unweighted (B) world air transportation network. The red circles are the components disappearing in the weighted network. The blue circles are similar small components in both networks. The green circles are the components that do not appear in the unweighted network. The orange circles are the components splitting in the weighted network. Geographical areas outside the figure do not contain small components.

large local components of the weighted and unweighted network illustrated in Figure4.2. In the first case, the components are very similar. In the second case, separated components in the unweighted network merge into a single component.

There are three similar components (North and Central America-Caribbean, Africa-Middle East-Southern Asia, and South America). We quantify their similarity using the Jaccard index. The higher the similarity, the closer the Jaccard index is to 1

The North and Central America-Caribbean weighted component contains 10% more airports than the unweighted. Their Jaccard index is high (0.81). Note, however, that 39 airports of the unweighted component disappear in the weighted component. They are mainly in French Antilles and Venezuela. More surprisingly, John F Kennedy airport is not in the weighted component.

Additionally, some new airports emerge in the weighted component (107). Most are in Canada, Alaska, Peru, and Chile. London Heathrow, the most important airport in the United Kingdom, also appears in the weighted component. Indeed, it has several destinations in the United States, and numerous flights

from North America land at this airport. The most unexpected airports in this component are the Marshall Airport in the Marshall Islands and the Osmany Airport in Bangladesh. The first collaborates with United Airlines situated in the United States. The second has flights to and from London Heathrow.

The Jaccard index of the unweighted (336 airports) and weighted (313 airports) Africa-Middle East-Southern Asia component is also high (0.87). The 35 airports of the unweighted component that disappear in the weighted are in the Middle East and West Africa. In contrast, eleven new airports emerged (6 are in Kenya, 2 in West Africa, and 3 in Europe) in the weighted component. The Frankfurt am Main airport, the largest airport in Germany is unexpected in this component. Indeed, it has high traffic with Saudi Arabia and India. The Rzeszów-Jasionka airport in Poland and Araxos airport in Greece are also in this component. Both airports have dense traffic with Frankfurt am Main.

The Jaccard Index of the unweighted (215 airports) and weighted (201 airports) South America components is 0.78. Thirty-two airports in this unweighted disappeared in the weighted. These airports are in Chile and Peru. As aforementioned, the airports of these two countries are now in the weighted North and Central America-Caribbean component. The airports appearing in the weighted South America component are in Venezuela, Colombia, and Cuba.

There are two merged components. The Europe-Russia-Central Asia component regroups the European and Russia-Central Asia-Transcaucasia components from the unweighted world air transportation network. Similarly, the "East and Southeast Asia-Oceania" component includes the unweighted network's East and Southeast Asia and Oceania components. We join the unweighted European (493 airports) and Russia-Central Asia-Transcaucasia (112 airports) components and the unweighted East and Southeast Asia (416 airports) and Oceania (234 airports) components to compare with their counterpart in the weighted network.

The Jaccard Index of the European-Russia component is 0.87. Nine airports present in the unweighted disappear, while 87 new airports appear in the weighted component. They are mainly in Norway, West and North Africa, Iran, United Arabs Emirates, and the French Antilles. Beyond their geographical localization, the countries of this weighted component have political, historical, and economic relations. Indeed, from the airports in Moscow and St Petersburg, the Russian airports join the other world regions throughout Europe. North Africa, West Africa, and the French Antilles are associated with Europe politically and historically. It translates into high traffic between these regions.

The unweighted and weighted East and Southeast Asia-Oceania components are very similar. Indeed, their Jaccard index equals 0.97. These two regions, through the largest airports, have significant traffic. In the weighted component, twenty-three airports in French Polynesia disappeared, and three airports from Russia and India emerged. Vladivostok and Yuzhno-Sakhalinsk airports are near China and North Korea. Their exchanges are with China, Japan, and South Korea. The Trichy airport in India has heavy traffic with Malaysia, Singapore, and Sri Lanka.

5.1.2 Small local components

Figure 4.3.A presents the twelve small local components in the weighted world air transportation network. They are either in a single country (the United States, Canada, French Polynesia, Greenland, Israel, Australia, United Arab Emirates) or cover a few countries or subregions (Caribbean). The biggest small component in Alaska contains 30 airports. The smallest includes two airports. In the following, we concentrate on small components with a size greater than five airports.

Alaska has three small components with comparable sizes (around 27 airports). These components have a star shape. Indeed, most of the traffic goes through a leading airport. Fairbanks Airport, in Northeast Alaska, leads the first component. The Nome Airport dominates the Northwest. The third



FIGURE 4.4 – A) The airports in the large global component. B) The 11 small global components are circled. The size range are between 2 and 3 airports.

component in the Southwest is centered around the Bethel Airport. These components are also in the unweighted world air transportation network. It reveals the isolation of Alaska in terms of traffic and connections.

Canada possesses two small components. The largest contains 25 airports, mainly on Nunavut and Quebec coasts. It is dominated by Iqaluit and Quujjuaq airports. The second includes ten airports in the Northwest Territories. The Inuvik airport serves several flights in this component. Note that the unweighted network merges these two components.

The significant small component in Greenland contains ten airports. The most frequent flights operate through Godthaab / Nuuk, Kangerlussuaq, Ilulissat, and Sisimiut airports. This component is in the unweighted network. Indeed, in Greenland, air transport is a must.

Twenty-two airports in French Polynesia constitute a small local component. This country is a set of islands. Thus, air transportation is very developed. The major airport of this country, Faa'a airport, reaches 18 airports with 547 flights. The Bora Bora airport is the second most important, with 302 flights. This component is not in the unweighted network. So, the flow of flights between the airports of French Polynesia is essential, even though the connections are not dense.

The last significant small component includes French Antilles, Quebec, and Ontario airports. Indeed, there is a strong community from French Antilles based in these two cities of Canada. Therefore, there is high traffic between these airports. The major airports in Trinidad and Tobago (Piarco Airport and Tobago-Crown Point Airport) capture most of the traffic. This component also exists in the unweighted network.

Globally, the small local components cover 6.6% of airports. They are in North and Central America-Caribbean (3 in Alaska and 2 in Canada and 1 in the Caribbean), Europe (1 in Greenland and 1 in Israel), East and Southeast Asia-Oceania (1 in French Polynesia and 1 in Australia), Africa-Middle East-Southern Asia (1 in the United Arab Emirates). Among them, five components include less than five airports.

5.2 Global components

Figure 4.4 shows the global components extracted from the weighted world air transportation network. There are one large and 11 small global components. The large global component contains 557 airports (20.44% of the world's airports). It covers the world. The small global components, which include 36 airports, are principally in the North and Central America-Caribbean region.

The Jaccard Index between the weighted and unweighted large global components is not high (0.63). The large global component of the weighted world air network contains 14 more airports than the unweighted. However, their content is very different. Indeed, 100 airports disappear from the weighted largest global component, and 144 new airports integrate it. Airports disappear in the largest weighted global component because Oceania and Russia merged with their neighboring regions reducing the inter-component links. New airports appear because of the singularities in the North and Central America-Caribbean, Europe-Russia-Central Asia, and Africa-Middle East-India components. Indeed, highly connected airports not localized in their natural geographical components, such as London Heathrow and John F. Kennedy, increase the links between the local components attracting new airports in the global component. Moving an airport from one local component to another modifies the large connected component drastically.

The small global components include 36 airports in North and Central America-Caribbean and East and Southeast Asia-Oceania. Their size ranges from 4 airports to 2 airports. Canada contains most of them (11). Only two components are shared with the unweighted world air transportation network. In the following, we neglect these components.

6 Topological properties of the large components

This section investigates the clustering coefficient and degree-degree correlation. We recall the definitions of these topological properties in weighted networks.

6.1 Clustering coefficient

The clustering coefficient C_i of a node i reflects the cohesiveness of its neighbors. The closer it is to 1, the more interconnected its neighbors are. Whether $C(k) \approx k^{-1}$ the network has a hierarchical organization [Ravasz, 2003]. The weighted clustering coefficient C_i^w of a node i is defined as follows [Barrat, 2004] :

$$C_i^w = \frac{1}{s_i(k_i - 1)} \sum_{j,h} \frac{(w_{ij} + w_{ih})}{2} a_{ij} a_{ih} a_{ij} jh \tag{4.5}$$

One can define $C(k)$ and $C^w(k)$ as the average clustering coefficient of nodes with degree k for, respectively, unweighted and weighted networks. The relation $C^w > C$ indicates that the interconnected triplets tend to be formed by links with high weights. The opposite $C^w < C$ shows that lower-weight edges produce interconnected triplets.

6.1.1 Large Local components

Figure 4.5 represents the average (weighted and unweighted) clustering coefficient versus degree for the large local components.

The average unweighted clustering coefficient decreases monotonically with the degree. Low-degree nodes have a large clustering coefficient. In contrast, hubs are less cohesive. This characteristic is shared with numerous networks such as the India air transport network [Bagler, 2008], the actor-network, the World Wide Web network, etc. Fitting $C(k)$ by the law $k^{-\gamma}$, we observe that $\gamma \approx 0.3$ for the Africa-Middle East-Southern Asia and the South America components. For the other components, $\gamma \approx 0.2$.

In contrast, the average weighted clustering coefficient is independent of the degree. This result differs from previous analysis with real-world networks [Barrat, 2004 ; Bagler, 2008]. Moreover, weighted average clustering coefficients are always lower than their unweighted equivalent. It shows that interconnected triplets tend to be formed by edges with low weight in the large local components.

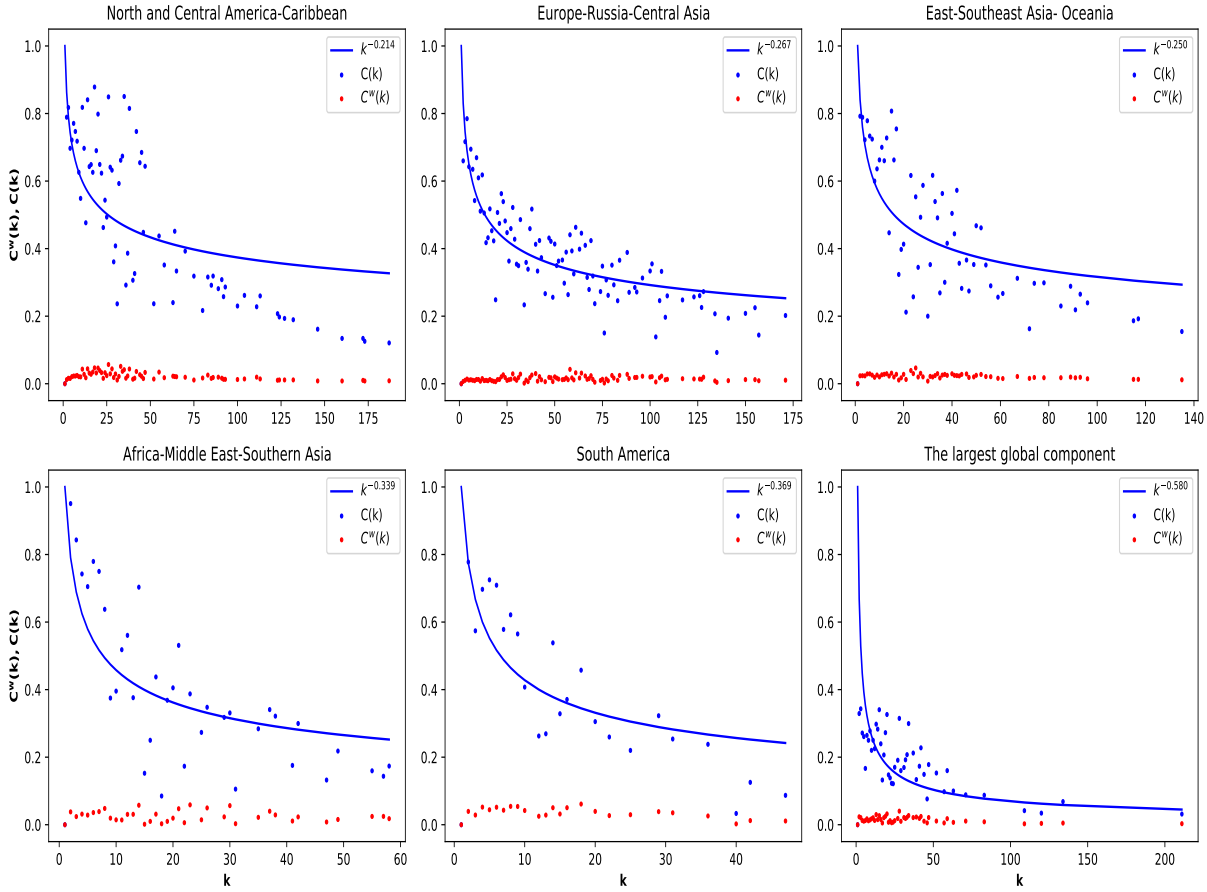


FIGURE 4.5 – Average clustering coefficient versus degree of the large components. The blue dots represent the average clustering coefficient for the unweighted network. The red dots represent the average clustering coefficient for the weighted network. The blue curve is the estimated $k^{-\beta}$ function.

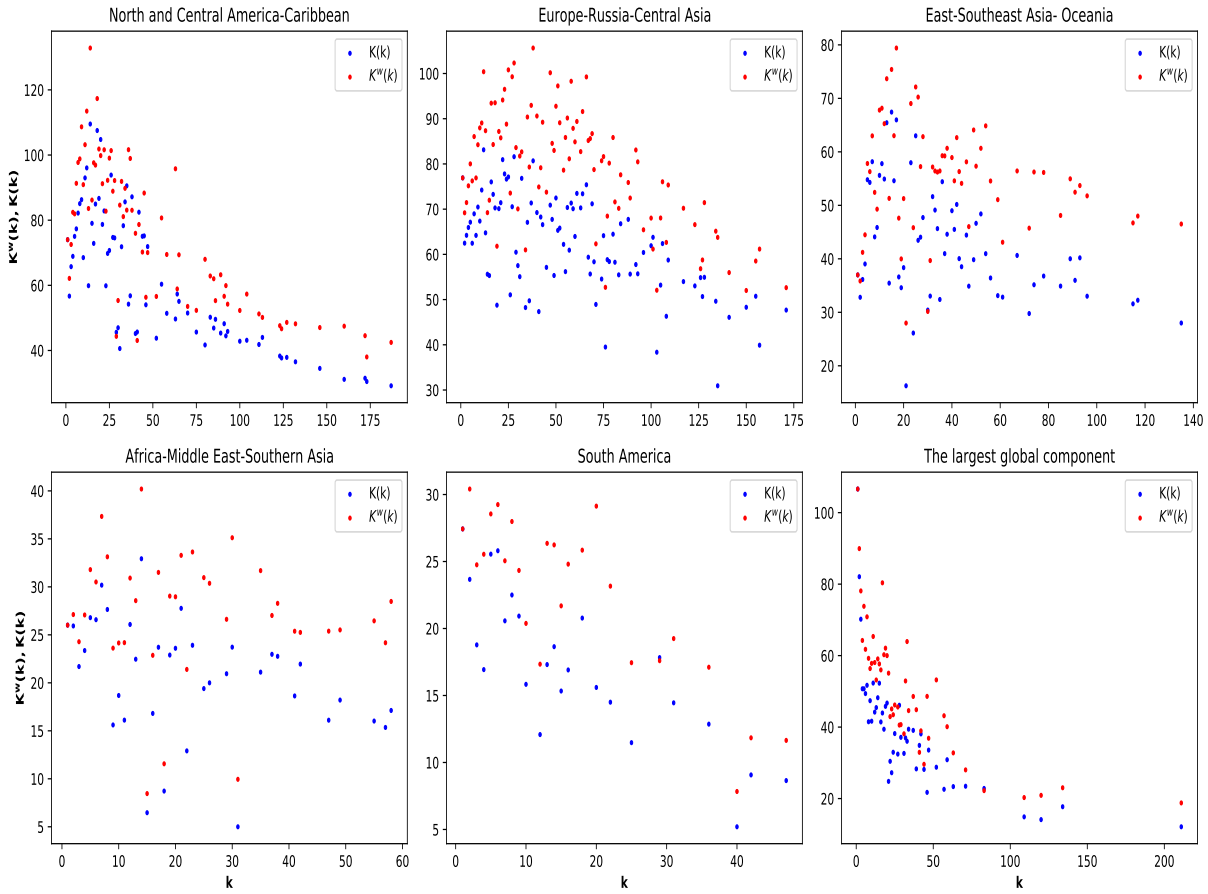


FIGURE 4.6 – Average degree-degree correlation versus degree of the large components. The blue dots represent the average degree-degree correlation for the unweighted network. The red dots represent the average degree-degree correlation for the weighted network.

6.1.2 Large Global component

Figure 4.5 reports the evolution of the average clustering coefficient versus the degree of the large global components. Overall, its behavior is similar to the local components. For the unweighted network, the coefficient of the fitted power law function is $\gamma \approx 0.58$. Consequently, there are fewer interconnected triplets in the global component compared to the local ones. Indeed, the global component includes long-distance traffic airports. Among these airports, numerous are hubs in their country in a hub and spoke configuration with poor rerouting capacity. Like in the large local components, the unweighted average clustering coefficient of a degree node k is greater than the weighted average clustering coefficient. Thus, the connected triplets are mainly between nodes with low-weight edges.

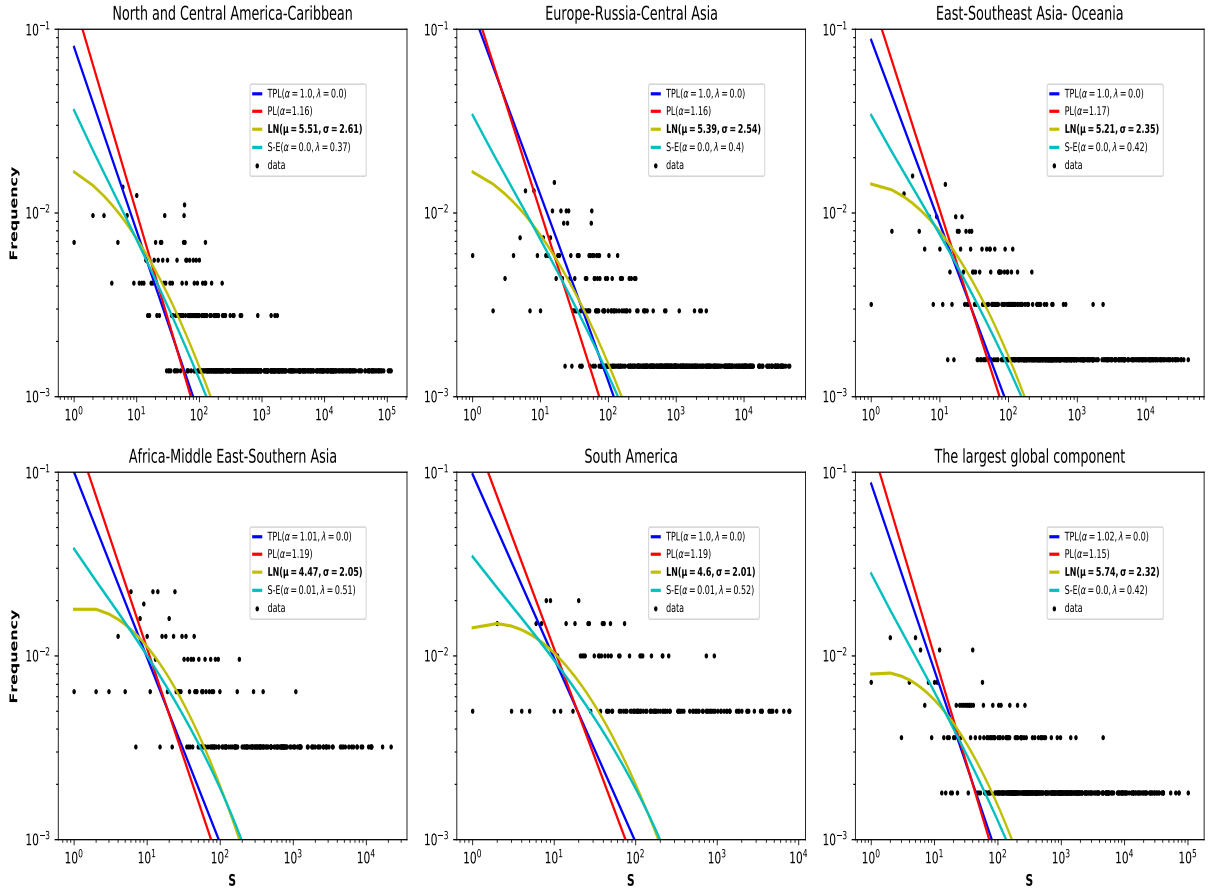


FIGURE 4.7 – Strength distributions of the large local components and the largest global component. Empirical distribution (**dot**), Estimates (**line**).The distributions under test are : Power Law (PL), Truncated Power Law (TPL), Log-Normal (LN), and Stretched Exponential (S-EXP). The values in bold are the parameters of the best fit according to the Kolmogorov-Smirnov test.

6.2 Degree-degree correlation

$$k^w_{nn,i} = \frac{1}{s_i} \sum_{j=1}^{N} a_{ij} w_{ij} k_j \tag{4.6}$$

One can compare weighted ($k^w_{nn,i}$) and unweighted ($k_{nn,i}$) degree-degree correlation. If $k^w_{nn,i} > k_{nn,i}$, edges with high weights tend to connect high degree nodes. If $k^w_{nn,i} < k_{nn,i}$ the edge weights connect the

degree nodes with opposite magnitudes (low or high).

6.2.1 Large Local components

Figure 4.6 presents the average degree-degree correlation evolution versus the degree for the large local components. One can observe that the weighted degree-degree correlation is greater than the unweighted degree-degree correlation. It indicates a high fraction of traffic transit between hubs in the local components.

However, the curve trend of the unweighted and weighted degree-degree correlation as a function ok k is different for the large local components. Indeed, above a threshold k, the weighted and unweighted degree-degree correlation of the North-Central America-Caribbean ($k \approx 70$), and South America ($k \approx 25$), decrease. The high-degree nodes tend to connect to low-degree nodes. Thus, the disassortative aspect of these unweighted and weighted components appears clearly. The Europe-Russia-Central Asia component's degree-degree correlation decreases slowly compared to the South America and North-Central America-Caribbean components. Moreover, the degree-degree correlation of the high-degree nodes is higher. It is also less disassortative. In the East and Southeast Asia-Oceania, and the Africa-Middle East-Southern Asia components, the unweighted degree-degree correlation shows a characteristic slightly disassortative. The average weighted degree-degree correlation of these components is relatively constant.

6.2.2 Large Global component

Figure 4.6 reports the average degree-degree correlation as a function of degree for the global component. The unweighted network curve decreases. Therefore, the global component is disassortative. Indeed, the largest hubs ($k > 100$) connect with low degree nodes ($k < 20$). The weighted network curve exhibits a similar evolution. However, $K^w(k) > K(k)$ for most of degree node k. Therefore, even though the hubs connect with small degree nodes, they accumulate more flights.

6.3 Strength Distribution

6.3.1 Large Local components

Figure 4.7 represents the strength distributions of the large local components. We perform a goodness of fit evaluation with the Kolmogorov Smirnov test (KS) using the power-law, truncated power law, log-normal, and stretched exponential distributions. Results reported in Table 4.6 reveal that the log-normal distribution better fits the large local components.

Moreover, they are heavy-tailed, like the degree distribution of the unweighted large local components of the world air transportation network. One can expect this result. Indeed, the higher the node degree, the higher its weight. Figure 4.8 shows this behavior. One can see the relation between the strength and the degree for each large local component. Indeed, the average strength and the degree are linked by the relation ($s(k) = k^\beta$). Even though for all the large local components, β ranges between 2.1 and 2.3, this organization is similar to results reported in [Bagler, 2008] where $\beta \approx 1.43$.

One can differentiate three types of large local components if we focus on the exponent of the strength versus degree curve (β). The Europe-Russia-Central Asia component forms the first category. It includes several hubs with almost the same traffic. Indeed, the traffic increases slowly as a function of degree. The second category consists of the North-Central America-Caribbean and the East-Southeast Asia-Oceania components. These components also contain several hubs. However, the traffic concentrates in a few hubs. The Africa-Middle East-Southern Asia and the South America components are in the last category. There

TABLE 4.6 – Kolmogorov-Smirnov goodness of fit values (KS-test) for the degree distribution. Distributions under test are Power-law, Truncated power law, Log-Normal, and Stretched exponential. The smallest value (bold) corresponds to the best fit.

	Power law	Truncated Power law	Log-Normal	Stretched Exponential
North and Central America-Caribbean	0.31	0.136	**0.044**	0.06
Europe-Russia-Central Asia	0.31	0.27	**0.041**	0.06
East and Southeast Asia	0.32	0.15	**0.032**	0.05
Africa-Middle East-Southern Asia	0.32	0.15	**0.045**	0.065
South America	0.33	0.15	**0.043**	0.18

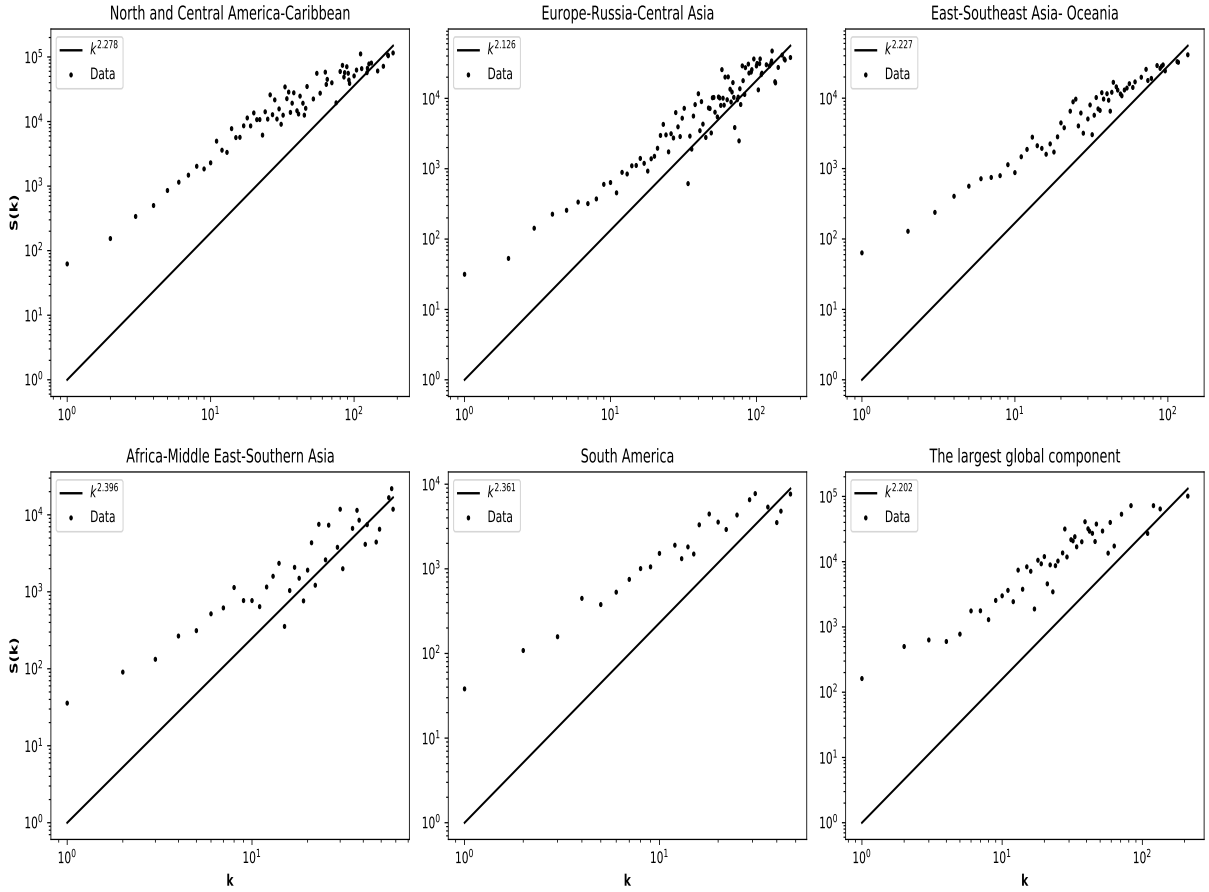


FIGURE 4.8 – Strength as a function of degree of the large local components and the largest global component. Empirical distribution (**dot**), Estimates (**line**).

are fewer hubs and less traffic compared to the other types. The strength as a function of degree increases faster. Consequently, few hubs accumulate most of the traffic.

6.3.2 Large Global component

Figure 4.7 and Figure 4.8 show the strength distribution and the strength as a function of k of the large global component. Like the large local components, the log-normal law better approximates its strength distribution according to the KS-test. One can see also that the strength distribution is heavy-tailed. In addition, the distribution parameters are similar to those of the East-Southeast Asia-Oceania component. As found with the degree-degree correlation, the degree's strength shows that the higher the node degree, the more traffic it accumulates. Airports with fewer connections have at least more than 100 flights. It is not the case in the large local components. The β exponent is comparable to those of North and Central America-Caribbean and the East-Southeast Asia-Oceania components.

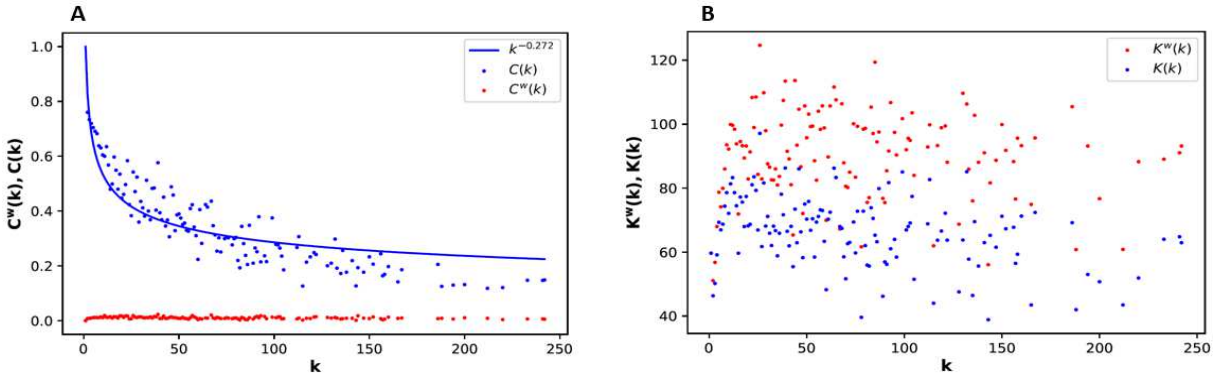


FIGURE 4.9 – A) represents the average clustering coefficient as a function of k. B) represents the average degree-correlation of degree k. For both figures, the red point is the values of the weighted world transportation network, while the blue points are the unweighted world transportation network values.

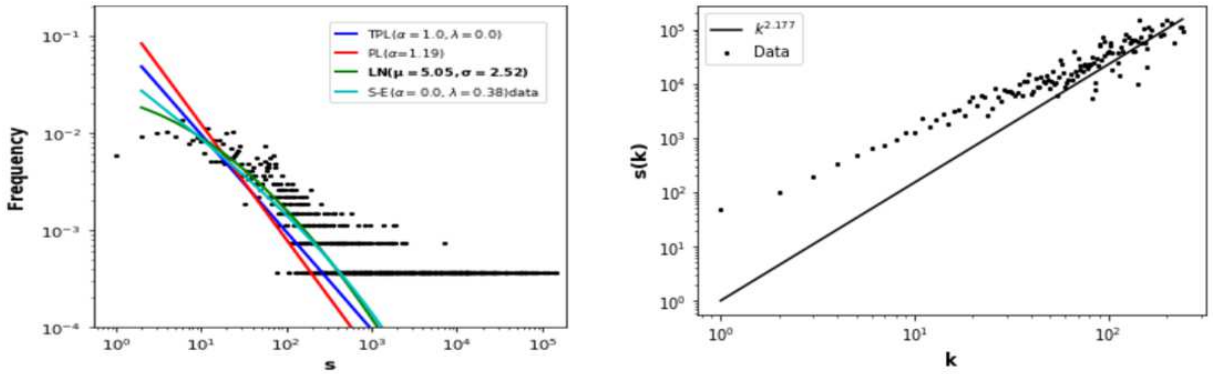


FIGURE 4.10 – The strength distribution (**left**) of the world air transportation network. Empirical distribution (**dot**), Estimates (**line**).The distribution under test are Power Law (**PL**), Truncated Power Law (**TPL**), Log-Normal (**LN**), and Stretched Exponential (**S-EXP**). The values in bold are the best fit parameters according to the Kolmogorov-Smirnov test. The strength as a function of degree (**right**) of the large global component.

7 Topological properties of the world air transportation

This section investigates the topological properties of the world air transportation network. We also perform a comparative analysis with the large components.

7.1 Clustering coefficient

Figure 4.9.A presents the average clustering coefficient as a function of degree k for the weighted and unweighted world air transportation network. Similar to the large components, the unweighted clustering coefficient decreases monotonically as the degree increases. The function $k^{-\gamma}$ (with $\gamma \approx 0.27$) gives a good approximation of their relation. The exponent value estimated is identical to the Europe-Russia-Central Asia component. The weighted average clustering is very low and almost independent of the degree. As it is always below the unweighted average clustering coefficient, one can conclude that edges with low degrees tend to constitute the triplets.

7.2 Degree-degree correlation

Figure 4.9.B illustrates the evolution of the degree-degree correlation as a function of degree k for the weighted world transportation network. For both weighted and unweighted networks, one can see that the distribution of the points is not monotone. Thus, one cannot conclude about the degree-degree correlation of nodes. This result differs from previous results showing an assortative behavior for the weighted world air transportation network [Barrat, 2004]. The Africa-Middle East-Southern Asia component is the only component with similar behavior. Indeed, all the others exhibit disassortative behavior.

7.3 Strength distribution

Like the components, the log-normal distribution best fits the strength distribution represented in Figure 4.10. In addition, the strength as a function of degree exponent ($s(k) = k^{\beta}$ with $\beta \approx 2.177$) is in the range of the local components.

TABLE 4.7 – Strength Centrality in the five large local components. We report the top five hubs for each large local component. The internal Strength of an airport is the number of flights with airports in its large local component. Local Rank is the airport rank in its component by decreasing internal strength. The cumulative fraction of connected airports represents the proportion of airports that the top x local hubs can reach in their components.

Large Local Component	**Airport**	**City**	**Country**	Internal Strength	Local Rank	Unweighted Local Rank
North and central America-Caribbean	Hartsfield J Atlanta	Atlanta	United States	115410	1	1
	Los Angeles	Los Angeles	United States	111415	2	11
	Chicago O'Hare	Chicago	United States	107554	3	3
	Dallas Fort Worth	Dallas-Fort Worth	United States	104321	4	2
	Ronald Reagan Washington	Washington	United States	87953	5	19
Europe	Dublin	Dublin	Ireland	47105	1	8
	Barcelona	Barcelona	Spain	45408	2	4
	Charles de Gaulle	Charles de Gaulle	Paris	41216	3	7
	John F Kennedy	New York	United States	38697	4	15
	Amsterdam Schiphol	Amsterdam	Netherlands	38065	5	1
East and Southeast Asia	Beijing Capital	Beijing	China	41320	1	1
	Singapore Changi	Singapore	Singapore	35556	2	12
	Shanghai Pudong	Shanghai	China	33043	3	3
	Guangzhou Baiyun	Guangzhou	China	32318	4	2
	Hong Kong	Hong Kong	Hong Kong	29712	5	8
Africa-Middle East-Southern Asia	Indira Gandhi	Delhi	India	21874	1	3
	Chhatrapati Shivaji	Mumbai	India	16704	2	4
	Kempegowda	Bangalore	India	11860	3	17
	King Abdudaziz	Jeddah	Saudi Arabia	11829	4	2
	Frankfurt am Main	Frankfurt	Germany	11778	5	4
South America	Tancredo Neves	Belo Horizonte	Brazil	7745	1	7
	G G A F Montoro	Sao Paulo	Brazil	7706	2	1
	El Dorado	Bogota	Colombia	7588	3	2
	Rio G–T Jobim	Rio De Janeiro	Brazil	6553	4	6
	Presidente J Kubistschek	Brasilia	Brazil	5366	5	5

8 Centrality Analysis

Centrality analysis investigate the most influential nodes in a network. There are multiple definitions of centrality that exploit either local or global caracteristics of the networks [LÃŒ, 2016 ; Das, 2018 ; Ibnoulouafi, 2018]. Here, we perform a comparative analysis of the Strenght (Number of flights in an airport) and Degree (Number of routes in an airport) centralities of the various components. This analysis is in line with recent works considering the community structure to define new centrality measures [Ghalmane, 2019b ; Kumar, 2018 ; Chakraborty, 2016 ; Blöcker, 2022].

8.1 Top five Strength Analysis

8.1.1 Regional Analysis

Table 4.7 reports the top five airports in descending order of the number of flights with airports in their local component (internal strength centrality).

The top five airports in the *North and central America-Caribbean component* are in the United States. They are in densely populated states such as Georgia, Illinois, Texas, California, and Washington D.C. Hartsfield J Atlanta Airport, captures the higher number of flights. It is also the most connected. It is one of the densest in terms of passengers. Located in the second-most populous city in the United States, the Los Angeles Airport is the second busiest in terms of flights, although it has a low ranking in terms of the number of routes. Chicago O'Hare Airport ranks third in the number of flights and destinations. Dallas Fort Worth Airport, the second most connected airport, ranks fourth according to the number of flights. All the airports mentioned above have more than 100 connections and 100000 flights within this component.

This is not the case for the fifth important airport, the Ronald Reagan Washington Airport. It is a national airport in the capital city of the United States. It is a hub for America Airlines and receives several million passengers. Altogether, these airports handle almost 15% of the regional flights. London Heathrow ranks ninth. It operates around 74000 flights. Denver and Houston airports' which are in the top five hubs in terms of connections rank respectively eleventh and sixteenth [Diop, 2021] when considering the number of flights.

In the *Europe-Russia-Central Asia-Central Asia component*, four out of five top airports are in Europe and one in the United States. Dublin Airport in Ireland is the first with more than 4500 flights. It is the central hub of the important low-cost carrier, Aer Lingus Airlines. Very connected, the Barcelona Airport is the second with numerous flights. Indeed, It is a tourist city receiving several million passengers per year. The largest hub, Charles de Gaulle Airport in the capital city of France, ranks third with considerable traffic. Unexpectedly, John F Kennedy Airport in the United States ranks fourth with 40 links within this component. This airport has more flights to Europe than several local airports. Indeed, it is the main gate between the United States and Europe. Even though it has a higher number of links in this component, Amsterdam Schiphol Airport ranks fifth. It is the principal airport of the Netherlands. The Munich Airport and the London Stansted Airport, which are in the five highest hubs in terms of routes, [Diop, 2021] rank respectively sixth and sixty-fifth in terms of flights. The Frankfurt Airport, in the same group as the two mentioned above, is now in Africa-Middle East-Southern Asia. The most vital airport in Russia, the Pulkovo Airport, is the fourteenth. The Dubai Airport, which belongs to this component, ranks twenty-fifth, although it is the biggest airport in the Middle East.

In the *East and Southeast Asia-Oceania component*, three out of five most busy airports are in China. The others are in Singapore and Hong Kong. The top airport is Beijing Capital Airport, China's capital city. It is the only one to have more than 40000 flights. Singapore Changi Airport ranks second, even if it is not one of the five most connected airports in this component. It is the primary hub of Singapore

TABLE 4.8 – Strength Centrality in the large global component. We report the top five inter-regional hubs in each region. External Strength is the number of connections of an airport with airports located in the large global component. GRR is the Global Rank in the Region is the airport Rank in its region by decreasing external Strength. GRC is the Global Rank in the Component is the airport rank in the global component by decreasing external Strength.

Region	**Airport**	**City**	**Country**	External Strength	GRR	GRC
North and central America-Caribbean	John F Kennedy	New York	United States	72166	1	3
	Newark Liberty	Newark	United States	40924	2	6
	Chicago O'Hare	Chicago	United States	38539	3	10
	Los Angeles	Los Angeles	United States	35148	4	11
	Lester B Pearson	Toronto	Canada	34818	5	12
Europe	Frankfurt	Frankfurt	Germany	101024	1	1
	Charles de Gaulle	Charles de Gaulle	Paris	72206	2	2
	London Heathrow	London	United Kingdom	64265	3	4
	Amsterdam Schiphol	Amsterdam	Netherlands	53340	4	5
	Munich	Munich Airport	Germany	40785	5	7
East and Southeast Asia	Narita	Tokyo	Japan	40608	1	8
	Beijing Capital	Beijing	China	39915	2	9
	Suvarnabhumi	Bangkok	Thailand	29477	3	16
	Incheon	Seoul	South Korea	29288	4	17
	Hong Kong	Hong Kong	Hong Kong	25989	5	20
Africa-Middle East-Southern Asia	Dubai	Dubai	United Arab Emirates	27074	1	19
	Indira Gandhi	Delhi	India	16882	2	32
	Hamad	Doha	Qatar	1121	3	40
	Abu Dhabi	Abu Dhabi	United Arab Emirates	13512	4	41
	Cairo	Cairo	Egypt	12803	5	43
South America	Guarulhos-G A F M	São Paulo	Brazil	21733	1	24
	El Dorado	Bogotá	Colombia	13616	2	39
	Ministro Pistarini	Buenos Aires	Argentina	9447	3	50
	Rio G–T Jobim	Rio de Janeiro	Brazil	6103	4	64
	Jorge Chávez	Lima	Peru	4575	5	80

Airlines. The third airport is the Shanghai Pudong Airport in Shanghai, a populated city in China. Guangzhou Baiyun Airport is another central hub in China, ranking fourth. Indeed, this airport has the second-largest number of connections and carries millions of passengers. The Hong Kong Airport, with less than 30000 flights, ranks fifth. Chengdu Shuangliu Airport, and Taiwan Taoyuan Airport, in the top five destinations, rank respectively eleventh and ninth [Diop, 2021]. The Sydney K Smith Airport, the most significant Oceania airport, ranks eighth.

The Indian airports dominate the *Africa-Middle East-Southern Asia component*. Indeed, they are in the top three busiest airports. The two others are in Saudi Arabia and Germany. With its numerous links, the Indira Gandhi Airport handles the highest traffic with more than 20000 flights. The second airport with almost 17000 flights is the Chhatrapati Shivaji Airport, located in Bombay, a metropolis of India. One of the primary airports in India, the Kempegowda Airport serving Bangalore, ranks third. It is a hub of Air Asia. The most connected airport, King Abdulaziz, ranks fourth. The Frankfurt am Main Airport is fifth with 30 internal connections and more than 11000 flights in this component. The most connected airport in the region, the Dubai Airport, is in the Europe-Russia-Central Asia component, and the fifth in terms of destinations, the Addis Ababa Bole Airport, ranks nineteenth in terms of flights.

In the *South America component*, among the top five, four airports are in Brazil and one in Colombia. They have less than 8000 flights. Tancredo Neves Airport is the first. Situated in Belo Horizonte metropolitan area, it is a hub of Azul Brazilian Airlines. It doesn't have a high number of connections, but its traffic is very dense. The second, Guarulhos G A F Montoro Airport, is the most crucial airport in Brazil. It serves the largest economic and tourist city, São Paulo. El Dorado Airport, located in the capital of Colombia, ranks third. Rio G–T Jobim Airport in Rio de Janeiro, one of the most populated cities in Brazil, is fourth. It is the less connected among the top five. The fifth airport, the Presidente J Kubistschek Airport, is located in the capital of Brazil.

8.1.2 Inter-Regional Analysis

Table 4.8 lists the top five airports participating in the inter-regional traffic in each large local component. These airports also belong to the large global component.

Four of the five most essential airports for interregional traffic in the North and central American-Caribbean region are in the USA. The other one is in Canada. The first two airports in this ranking are the John F Kennedy Airport and the Newark Liberty Airport. With more connections (120), John F Kennedy airport's (nearly 100000) flight flow is much denser than Newark Liberty airport's (39 flights and almost 70000 flights). These airports are located in New York state. Indeed, this state, located in the Northeastern United States, is one of the most important, with the largest number of residents. Moreover, it is near other regions and includes numerous international institutions and multinational corporations. The Chicago O'Hare Airport and the Los Angeles Airport rank third and fourth. They are in other populated cities, Chicago and Los Angeles, in the United States. These two airports have comparable degrees and traffic. In Canada, the busiest airport, Lester B. Pearson Airport, ranks fifth. It is a gateway to this country and the USA. The top three in this region are in the top 10 of the large global component.

The Frankfurt am Main Airport dominates the inter-regional flow of flights in the Europe-Russia-Central Asia region. Being the most important airport in Germany, it has around 100000 flights with 211 connections. The Charles de Gaulle Airport ranks first with almost 70000 flights. Located in the capital of France, it is the busiest airport in this country. Very connected, the London Heathrow Airport is a third of the list. It serves several flights to different regions of the world. The busiest airport in the Netherlands, the Amsterdam Schiphol Airport ranks fourth. The Munich Airport is the fifth airport with the largest inter-regional traffic. It is the second busiest airport in Germany. All these airports mentioned above are in the top 10 of the global component. One can say that the European airport leads the inter-regional

flights.

The top five busiest interregional airports in the East and Southeast Asia-Oceania region do not include any airport in Oceania. Indeed, the Narita Airport in Japan's capital city is the busiest inter-regional airport with about 40000 flights. The Beijing Airport ranks second. It is the most connected, with traffic comparable to the Narita Airport. The Suvarnabhumi Airport, the most critical airport in Thailand, is the third, with almost 30000 flights. The Incheon Airport, the most significant hub in South Korea, is the fourth. Its traffic is around 30000. The fifth airport with the largest international traffic is the Hong Kong airport, with almost 26000 flights. Only Narita and Beijing airports figure in the top 10 airports in the large global component.

While the Indian airports dominate the regional traffic, the busiest inter-regional airports of the Africa-Middle East-Southern Asia region are scattered in the Middle-east. The Dubai Airport is by far the most dynamic in this region. Its position and big area promote it. Located in the capital of Qatar, the Hamad airport ranks third. The second-largest airport in the United Arab Emirates ranks fourth. Cairo Airport in Egypt is fifth in this ranking. It has a comparable number of flights to the third and fourth airports (between 12000 and 14000). The first inter-regional airport in this region is not in the top 10 of the large global component. It ranks nineteenth.

The inter-regional airports handling the highest traffic in the South America region are in different countries. The Guarulhos Airport in Brazil is the busiest. El Dorado Airport in Colombia follows. These two airports are also regional hubs. The Ministro Pistarini Airport, located in the capital of Argentina, ranks third. The regional hub and the second most important in Brazil, the Rio G–T Jobim Airport, is the fourth inter-regional airport in this area. No airport in this component is in the top 20 of the large global component. Indeed, the Guarulhos Airport ranks twenty-fourth.

8.2 Comparison with the World Transportation network

Table 4.9 lists the 25 busiest airports in the world regarding the number of flights. It also includes their local and global strength rank and their degree rank. It shows that the North and Central America-Caribbean region controls a big part of the world's traffic. Indeed, 19 are in this area. Five are in the Europe-Russia-Central Asia-Central Asia region, and one in the East and Southeast Asia-Oceania. If we compare the airports' strengths with their degree, one detects that European airports usually deserve more destinations worldwide while North American airports have more flights.

Although 19 airports in the North and central America-Caribbean area are in the top 25 worldwide airports for their traffic, only six are in the top 25 inter-regional airports. In contrast, they are all in the top 25 regional airports. These results demonstrate that the North and central America-Caribbean region focuses more on regional traffic. The airports of New York (John F Kennedy and Newark airports) rank respectively 3 and 6 for inter-regional traffic. Therefore, New York City is the US gateway. Los Angeles is the busiest airport in the world, but the traffic is well distributed between local and global destinations. Chicago O'Hare and Hartsfield J Atlanta are in the top 5 airports in the world, but it is mainly due to their regional position. Note that the rank of the regional airports in the world air transportation network can be misleading because they are not the most active in inter-regional traffic.

London Heathrow Airport ranks third worldwide. It is mainly due to its position in the interregional traffic. In contrast, Charles de Gaulle Airport, Frankfurt am Main Airport, Amsterdam Schiphol Airport, and Munich Airport rank 6, 7, 12, and 21 in the top 25 airports of the world. They are all in the top 25 regional and inter-regional airports. However, they generally exert a stronger influence at the regional level. Note that London Heathrow belongs to the American-Caribbean component. Its traffic is very dense compared with the airport this component. One can make a similar remark about the Frankfurt am Main Airport, which belongs to the Africa-Middle East-Southern Asia component.

The Beijing Airport is the only one from the East and Southeast Asia region in the top 25 crucial airports. It is the eighteenth worldwide according to the number of flights. In addition, it is the first regional airport and the most connected to the world.

Dubai Airport is the forty-fifth in the world, the first from Africa-Middle East-Southern Asia, even though it deserves many routes. Unexpectedly, it is the twenty-fifth influential regional airport of the Europe-Russia-Central Asia-Central Asia component and the nineteenth inter-regional airport. The most influential airport in South America, the Guarulhos-Governador André F M Airport in São Paulo, Brazil, is the seventy-fifth airport in the world in terms of traffic. In comparison, it is the first in the South America component and the twenty-fourth important inter-regional airport.

To summarize, North America leads the traffic in the world air transportation network. The component structure shows that most of this traffic is regional. Indeed, the large global component, which captures the inter-regional flights, exhibits numerous airports from different world areas essential to the interregional traffic. In addition, the large local components display the influential regional airports hidden in the world air network.

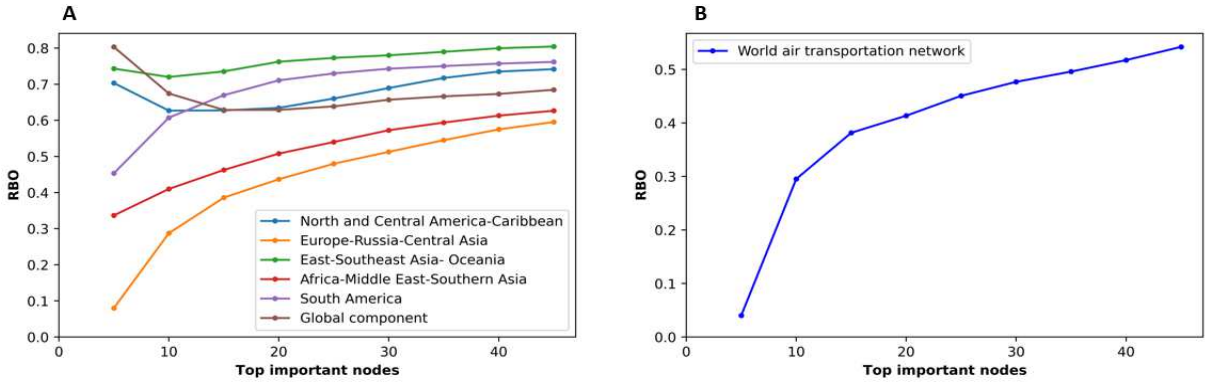


FIGURE 4.11 – The RBO of the large components (A) and the world air transportation network (B). The top 50 hubs of the unweighted (degree) and weighted (strength) are compared by step of 5.

8.3 RBO Analysis

The ranked-biased overlap (RBO) [Webber, 2010] quantifies the similarity of two ranking lists. A parameter allows to weight high ranks more heavily than low and to increase the depth of evaluation. Its value ranges between 0 and 1. The higher its value, the more identical the lists. We compare strength with degree centrality of the airports of the large components[Diop, 2021] . We tune the RBO to give equal importance to all ranks. Figure 4.11 shows the evolution of the RBO of the top-ranked airports for the large components and the world air transportation network. The curves cover the range from top 5 to top 45 sampled with a step of 5.

One can distinguish two types of curves. In the first category, the RBO increases monotonically as the number of airports increases. It includes the Europe-Russia-Central Asia, Africa-Middle East-southern Asia, East-Southeast Asia-Oceania, and South America components. Flights in these components do not concentrate on the main hubs. Indeed, the top 5 degree and strength airports are different. In the second category, the RBO decreases to a minimum, and then it increases monotonically. North and Central America-Caribbean, East -Southeast Asia-Oceania, and the global component belong to this category. Their top 5 airports ranking according to degree or strength is highly similar. There are more differences

between the top 10 to 15 airports, so the RBO decreases. Beyond this value, the strength and degree rankings become more homogeneous, so the RBO increases. Looking at the Top 45 airports, one can rank the components according to the similarity of strength and degree rankings. East-Southeast Asia-Oceania is the component where traffic (strength) and hub size (degree) are the more similar. The Europe-Russia-Central Asia component exhibits the most different ranking. So, in the former, traffic is generated by hubs, while in the latter, airports with few connections can handle a high share of traffic.

The RBO between top strength and degree airports increases monotonically in the world air transportation network. Differences are more pronounced compared to the components. Indeed, the top 5 airports based on degree are concentrated in Europe, while the top 5 based on strength are in the United States. This difference illustrates the different focus in the two regions. Indeed, in the USA, airlines focus on regional traffic, while in Europe, serving many international destinations is a must for companies.



FIGURE 4.12 – The core of the weighted large local components. S is the maximum s-core value, and K is the maximum k-core value. Blue points represent the airports belonging to the maximum s-core and the maximum k-core, the red points are the airports exclusively in the maximum s-core, and the yellow points are the airports belonging only to the maximum k-core. The maximum s-core of the North and central America-Caribbean component contains 26 airports. The maximum s-core of Europe-Russia-Central Asia-Central Asia contains 27 airports. The maximum s-core of East and Southeast Asia-Oceania includes 18 airports. The maximum s-core of Africa-Middle East-Southern Asia contains 9 airports. The maximum s-core of South America contains 10 airports.

TABLE 4.9 – The top 25 nodes in the world air transportation network are in descending order of strength centrality. Worldwide strength rank is the strength centrality rank in the world air transportation network. It is computed using the total strength (Internal and external). WSR is the Worldwide Strength Rank. LSR is the Local Strength rank is the strength centrality rank in the large local components. It is calculated using the Internal strength. GSR is the Global Strength Rank is the strength centrality rank in the global component. It is computed using the External strength. No indication is given when an airport does not belong to the global component. Top inter-regional airports are in bold. Top regional and inter-regional airports are in bold and italic. Others are top regional airports. WDR is the Worlwide Degree Rank.

Region	Airport	City	Country	Strength	WSR	LSR	GSR	WDR
North and central America-Caribbean	Los Angeles	Los Angeles	United States	146563	1	11	11	24
	Chicago O'Hare	Chicago	United States	146093	2	3	10	6
	Hartsfield J Atlanta	Atlanta	United States	135109	4	1	28	5
	Dallas Fort Worth	Dallas	United States	122407	5	4	30	8
	John F Kennedy	***New York***	***United States***	***110863***	***8***	***4***	***3***	***14***
	Newark Liberty	Newark	United States	106850	9	14	6	19
	San Francisco	San Francisco	United States	104325	10	10	15	52
	Seattle Tacoma	Seattle	United States	98539	11	6	35	67
	Charlotte Douglas	Charlotte	United States	89409	13	7	55	26
	Ronald Reagan Washington	Washington	United States	88348	14	5	248	73
	Lester B. Pearson	Toronto	Canada	85631	15	25	12	23
	McCarran	Las Vegas	United States	83054	16	8	76	33
	Philadelphia	Philadelphia	United States	81950	17	15	30	40
	General E L Logan	Boston	United States	78835	19	19	25	57
	George Bush Houston	Houston	United States	77421	20	16	33	13
	Denver	Denver	United States	76329	22	11	77	12
	Montreal/ Pierre E Trudeau	Montreal	Canada	72866	23	18	37	65
	Phoenix Sky HaRBOr	Phoenix	United States	71831	24	12	224	71
	Minneapolis-St Paul	Minneapolis	United States	70461	25	13	87	39
Europe	**London Heathrow**	**London**	**United Kingdom**	**138629**	**3**	**9**	**4**	**10**
	Charles de Gaulle	***Paris***	***France***	***113422***	***6***	***2***	***2***	**3**
	Frankfurt am Main	***Frankfurt***	***Germany***	***112802***	***7***	***5***	***1***	**2**
	Amsterdam Schiphol	***Amsterdam***	***Netherlands***	***91405***	***12***	***5***	***5***	**1**
	Munich	**Munich**	**Germany**	**77154**	**21**	**6**	**7**	**9**
East and Southeast Asia	***Beijing Capital***	***Beijing***	***China***	***81235***	***18***	***1***	***9***	**1**

TABLE 4.10 – 26 Max s-core airports of the North and central America-Caribbean component. Airports not in the top 26 regional airports with the highest Strength airports are in bold. The Core degree of a node is its number of flights in the maximum s-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports are ranked from largest to smallest strength.

Airport	City	Country	Core strength	Local rank	National rank
Seattle Tacoma	Seattle	United States	50753	6	7
London Heathrow	London	United Kingdom	50191	9	1
Ronald Reagan Washington	Washington	United States	45250	5	10
Los Angeles	Los Angeles	United States	45250	2	1
Montreal /Pierre E Trudeau	Montreal	Canada	41570	18	1
San Francisco	San Francisco	United States	37141	10	10
San Diego	San Diego	United States	37061	17	22
McCarran	Las Vegas	United States	35950	8	11
Phoenix Sky HaRBOr	Phoenix	United States	35437	12	16
Charlotte Douglas	Charlotte	United States	33968	7	9
Philadelphia	Philadelphia	United States	31443	15	14
Orlando	Orlando	United States	25965	21	21
Hartsfield J Atlanta	Atlanta	United States	25885	1	1
Portland	Portland	United States	25872	24	24
Denver	Denver	United States	25779	11	15
General E L Logan	Boston	United States	24331	19	13
Dallas Fort Worth	Dallas-Fort Worth	United States	23317	4	4
Salt Lake City	Salt Lake City	United States	23296	22	24
Chicago O'Hare	Chicago	United States	22433	3	2
Minneapolis-St Paul/Wold-Chamberlain	Minneapolis	United States	21173	13	17
Fort Lauderdale Hollywood	Fort Lauderdale	United States	19801	26	26
Raleigh-durham	**Raleigh-durham**	**United States**	**24106**	**34**	**32**
Miami	**Miami**	**United States**	**22423**	**28**	**19**
Bradley	**Windsor Locks**	**United States**	**20296**	**40**	**38**
Pittsburgh	**Pittsburgh**	**United States**	**20125**	**36**	**33**
Cancun	**Cancun**	**Mexico**	**20119**	**27**	**2**

TABLE 4.11 – 27 Max s-core airports of the Europe-Russia-Central Asia component. Airports not in the top 27 regional airports with highest Strength are in bold. The Core strength of a node is its number of flights in the maximum s-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports are ranked from largest to smallest strength.

Airport	City	Country	Core strength	Local rank	National rank
John F Kennedy	New York	United States	23129	4	21
Dublin	Dublin	Ireland	20467	1	1
Manchester	Manchester	United Kingdom	19058	8	2
Amsterdam Schiphol	Amsterdam	Netherlands	62	5	1
Dubai	Dubai	United Arab Emirates	16916	26	1
Geneva Cointrin	Geneva	Switzerland	16810	16	2
Adolfo S Madrid–Barajas	Madrid	Spain	16789	7	1
Humberto Delgado Airport	Lisbon	Portugal	16308	23	1
Nice-Côte d'Azur	Nice	France	16063	21	3
Oslo Lufthavn	Oslo	Norway	16026	27	2
Barcelona	Barcelona	Spain	15153	2	2
Stockholm-Arlanda	Stockholm	Sweden	15002	17	1
Hamburg	Hamburg	Germany	14642	18	5
Edinburgh	Edinburgh	United Kingdom	14029	22	3
Ataturk	Istanbul	Turkey	13352	10	1
Vienna	Vienna	Austria	12872	20	1
Leonardo da V–F	Rome	Italy	11601	11	1
Malpensa	Milano	Italy	11457	25	2
Copenhagen Kastrup	Copenhagen	Denmark	11350	19	1
Pulkovo	St. Petersburg	Russia	11334	15	1
Berlin-Tegel	Berlin	Germany	11312	9	4
Vaclav Havel Prague	Prague	Czech Republic	11219	24	1
Duesseldorf	Duesseldorf	Germany	10881	12	3
Eleftherios Venizelos	**Athens**	**Greece**	**13591**	**28**	**1**
Venice Marco Polo	**Venice**	**Italy**	**15029**	**29**	**3**
Helsinki Vantaa	**Helsinki**	**Finland**	**15523**	**32**	**1**
Ben Gurion	**Tel-aviv**	**Israel**	**10747**	**35**	**1**
Budapest Liszt Ferenc	**Budapest**	**Hungary**	**10840**	**37**	**1**

TABLE 4.12 – 18 Max s-core airports of the East and Southeast Asia-Oceania component. Airports not in the top 18 regional airports with the highest Strength are bold. Core strength of a node is its number of flights in the maximum s-core. Local rank is the airport rank in the large local component. National rank is the airport rank in its country. Airports are ranked from largest to smallest strength.

Airport	City	Country	Core strength	Local rank	National rank
Sydney KS-mith	Sydney	Australia	14412	8	1
Shanghai Pu-dong	Shanghai	China	12981	3	2
Melbourne	Melbourne	Australia	12384	14	2
Hong Kong	Hong Kong	Hong Kong	12202	5	1
Soekarno-Hatta	Jakarta	Indonesia	11662	15	1
Singapore Changi	Singapore	Singapore	10839	2	1
Taiwan Taoyuan	Taipei	Taiwan	9474	9	1
Suvarnabhumi	Bangkok	Thailand	9436	7	1
Incheon	Seoul	South Korea	9410	6	1
Guangzhou Baiyun	Guangzhou	China	8962	4	3
Suvarnabhumi	Kuala Lumpur	Malaysia	8467	10	1
Beijing Capi-tal	Beijing	China	7837	1	1
Brisbane	**Brisbane**	**Australia**	**9241**	**20**	**3**
Tan Son Nhat	**Ho Chi Minh City**	**Vietnam**	**7959**	**21**	**1**
Narita	**Tokyo**	**Japan**	**8313**	**22**	**1**
Auckland	**Auckland**	**Australia**	**8338**	**26**	**1**
Ngurah Rai (Bali)	**Denpasar**	**Indonesia**	**7826**	**36**	**2**
Kansai	**Osaka**	**Japan**	**7687**	**25**	**3**

9 S-core Analysis

The s-core [Eidsaa, 2016 ; Zhou, 2021] analysis is the generalization of the k-core [Kong, 2019a] analysis to weighted graphs. The k-core [Kong, 2019b] of a graph is the subgraph obtained by recursively removing all the vertices of degree smaller than k until the degree of all remaining vertices is larger than or equal to k. By extension, the s-core of a graph is a sub-network in which a node has at least a strength s. One can extract the maximum s-core by removing nodes iteratively from the network. Indeed, the $s_{min(s_i)}$-core, where each node has at least a strength 1, is the whole network. One forms the next level, by removing all the nodes with the minimum strength $s_{min(s_{i+1})}$. The remaining nodes form $s_{min(s_{i+1})}$-core, and so on until one reaches the core number Max s_n-core for which it is impossible to obtain the $s_{min(s_{n+1})}$-core.

This section reports the Max s-core analysis of the large weighted components and a comparative investigation with their corresponding Max k-core. Additionally, it presents a similar analysis of the world transportation network.

TABLE 4.13 – 9 Max s-core airports of the Africa-Middle East-Southern Asia component. Airports not in the top 9 regional airports with the highest Strength are in bold. The Core strength of a node is its number of flights in the maximum s-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports are ranked from largest to smallest strength.

Airport	City	Country	Core strength	local rank	National rank
Indira Gandhi	Delhi	India	8699	1	1
Chhatrapati Shivaji	Mumbai	India	7414	2	2
Kempegowda	Bangalore	India	6455	3	3
Frankfurt am Main	Frankfurt	Germany	5907	6	1
King Abdulaziz	Jeddah	Saudi Arabia	5539	5	1
Chennai	Madras	India	4882	8	4
King Fahd	Dammam	Saudi Arabia	4757	4	2
King Khaled	Riyadh	Saudi Arabia	4658	7	1
Rajiv Gandhi	**Hyderabad**	**India**	**4609**	**10**	**5**

9.1 Regional Analysis

Figure 4.12 reports the maximum s-core values of the five weighted large local components. It also shows the airports it includes.

The Max s-core of the North and central America-Caribbean component contains 26 airports. Fort Lauderdale Airport, with 19801 flights Airport has the lowest traffic. Table 4.10 lists these airports. Only three are outside the United States : 1) Montreal/P E Trudeau Airport in Montreal, the busiest airport in Canada ; 2) Cancun Airport, the second busiest in Mexico ; and 3) London Heathrow Airport in the United Kingdom. Their traffic is so high with the US that they can be considered part of the country. Most of the airports are on the coast of the United States. None deserves New York. Indeed, its airports' main traffic is with other world regions. Four airports (Raleigh-Durham, Pittsburgh, Windsor Locks, and Cancun) in the Max s-core are not among the top 26 airports in terms of traffic. These airports primarily serve other airports in the Max s-core.

The Max s-core contains almost half as many airports as the Max k-core. However, they share 22 airports (84.6% of the Max s-core airports). The traffic and destinations of these airports are therefore closely related. London Heathrow in the United Kingdom, Portland and Windsor Locks in the United

TABLE 4.14 – 10 Max s-core airports of the South America component. Airports not in the top 10 regional airports with the highest Strength are in bold. The Core strength of a node is its number of flights in the maximum s-core. Local rank is the rank in the large local component. National rank is the airport rank in its country. Airports in the top 10 regional hubs and not in the k-core are in red. Airports are ranked from largest to smallest strength.

Airport	City	Country	Core strength	local rank	National rank
Tancredo Neves	Belo Horizonte	Brazil	4305	2	3
Santos Dumont	Rio De Janeiro	Brazil	3564	6	7
Afonso Pena	Curitiba	Brazil	3500	7	8
Congonhas	Sao Paulo	Brazil	3030	8	9
Guarulhos-Governador A F Montoro	Sao Paulo	Brazil	2887	1	1
Deputado Luiz E Magalhaes	Salvador	Brazil	2504	10	8
Presidente J Kubistschek	Brasilia	Brazil	1927	4	4
Rio Galeao–Tom Jobim	Rio De Janeiro	Brazil	1914	3	2
Salgado Filho	**Porto Alegre**	**Brazil**	**2962**	**11**	**10**
Santa Genoveva	**Goiania**	**Brazil**	**2095**	**13**	**12**

States, and Montreal / Pierre Elliott Trudeau have more traffic than destinations with the other airports of the Max s-core.

The Max s-core of the Europe-Russia-Central Asia component has 27 airports, listed in Table4.11. Covering several countries in Europe, they are generally in the capital cities. Pulvoko Airport is the only one in the Russian sub-region. Ben Gurion Airport has the lowest traffic with 10747 flights. John F Kennedy in the USA has the highest number of flights. Indeed, it is the main connection between Europe and the USA. Charles de Gaulle in France, Munich in Germany, Amsterdam in the Netherlands, and Zurich Airport in Switzerland are not in the Max s-core. Indeed, the traffic of these airports is more directed towards inter-regional destinations. Five airports in this Max S-core are not in the 27 top strength airports. These airports are Venice, Marco Polo in Italy, Eleftherios Venizelos in Greece, Budapest Liszt Ferenc in Hungary, Ben Gurion in Israel, and Helsinki Vantaa Airport in Finland.

The Max s-core contains three times fewer airports than the Max k-core. All airports in the Max s-core are in the Max k-core. John F Kennedy Airport in the USA has a lot of traffic and connections to airports in Europe. Otherwise, no country dominates the Max s-core, while the Max k-core airports are mainly in Germany, the United Kingdom, France, Italy, and Spain.

TABLE 4.15 – 22 Max s-core airports of the global component. Airports not in the top 22 inter-regional airports with the highest Strength are in bold. The Core strength of a node is its number of links in the maximum s-core. Global rank is the rank in the largest global component. Airports are ranked from largest to smallest strength.

Airport	City	Country	Core strength	global rank
Frankfurt am Main	Frankfurt	Germany	27927	1
Amsterdam Schiphol	Amsterdam	Netherlands	22890	4
San Francisco	San Francisco	United States	22449	15
Munich	Munich	Germany	22103	7
Chicago O'Hare	Chicago	United States	21893	10
Los Angeles	Los Angeles	United States	21427	11
Charles de Gaulle	Paris	France	20908	2
Newark Liberty	Newark	United States	19757	6
Zurich	Zurich	Switzerland	19578	13
John F Kennedy	New York	United States	18260	3
Narita	Tokyo	Japan	18844	8
Hong Kong	Hong Kong	Hong Kong	17970	20
Beijing Capital	Beijing	China	17445	9
Leonardo da Vinci–Fiumicino	Rome	Italy	15071	14
Shanghai Pudong	Shanghai	China	14494	21
Incheon	Seoul	South Korea	14068	17
Adolfo S Madrid–Barajas	Madrid	Spain	13680	18
Lester B. Pearson	Toronto	Canada	13532	12
Suvarnabhumi	Bangkok	Thailand	13196	16
General E L Logan	**Boston**	**United States**	**17800**	**25**
Dallas Fort Worth	**Dallas-Fort Worth**	**United States**	**13588**	**30**
Seattle Tacoma	**Seattle**	**United States**	**13522**	**35**

Table 4.12 contains the 18 airports of the Max s-core of the East and Southeast Asia-Oceania component. These airports are in different countries, mainly in Capitals and megacities. Only four are in Oceania. Kansai Airport in Japan, with its 7687 flights, has the lowest traffic. Note that the airport of

TABLE 4.16 – 11 airports of the weighted world air transportation network which are not in the Max s-core of the North and Central America-Caribbean component. Airports not in the top 33 worldwide airports with the highest Strength are in bold. The Core Strength of a node is its number of flights in the maximum s-core. Global rank is the rank in the large local component.

Airport	City	Country	Core Strength	worldwide rank	local rank	global rank
Frankfurt am Main	Frankfurt	Germany	43819	7	5	1
Charles de Gaulle	Paris	France	38975	6		2
Newark Liberty	Newark	United States	33703	9		6
George Bush Houston	Houston	United States	29454	20		33
Detroit M Wayne County	Houston	United States	28779	26		46
Amsterdam Schiphol	Amsterdam	Netherlands	27762	12	5	5
Narita	**Tokyo**	**Japan**	**25972**	**38**		**8**
Vancouver	**Vancouver**	**Canada**	**25092**	**49**		**31**
Licenciado Benito Juarez	**Mexico City**	**Mexico**	**25035**	**52**		**53**
John F Kennedy	New York	United States	24839	8	4	3
Munich	Munich	Germany	24458	21		5 7

Beijing, with the most significant traffic of this component, is the third airport with the weakest traffic in this Max s-core. Indeed, it has many flights with low-strength airports. There are five airports (Brisbane Airport in Australia, Auckland Airport in New Zealand, Tan Son Nhat Airport in Vietnam, Ngurah Rai Airport in Indonesia, and Kansai Airport in Japan) in the Max s-core that are absent from the 18 top strength airports. The Max s-core contains half as many airports as the k-core. Nevertheless, China dominates the Max k-core, it is not the case for the Max s-core, which includes airports from Oceania.

Table 4.13 lists the nine airports in Max s-core of the Africa-Middle East-Southern Asia component. They are in three countries (5 in India, 2 in Saudi Arabia, and 1 in Germany). Indeed, there is a lot of traffic between Saudi Arabia and India because these many Indians work in Saudi Arabia. With 4609 flights, Rajiv Gandhi International Airport is the airport with the lowest traffic in the Max s-core. These airports form a complete graph. They concentrate a large part of the traffic of the component. Indeed, except for Rajiv Gandhi Airport, they are all in the top 9 strength airports, The Max k-core is twice larger, and it includes the airports of the Max s-core. Thus, Frankfurt Airport is very connected in terms of traffic and destination with the hubs of this component. Nevertheless, India dominates the Max s-core and the Max k-core.

Table 4.14 gives the ten airports in the Max s-core of the South America component. They are all in Brazil and mainly on the east coast. Rio Galeão - Tom Jobim Airport has the fewest number of flights. Only two airports, Santa Genoveva Airport in Goiania and the Deputado L E Magalhães Airport, are not present in the top 10 strength airports. All the airports in the Max s-core are in the Max k-core, which contains almost twice as many airports. Furthermore, Brazil dominates the Max s-core as it does with the Max k-core.

To summarize, we observe two typical behavior for the Max s-core. In the first one, the airports forming the Max s-core are mainly in a single country. Indeed, the United States, India, and Brazil dominate the Max s-core of their component. In the second case, the airports in the Max s-core are more evenly distributed in the component. It is the case for the East and Southeast Asia-Oceania and Europe-Russia-Central Asia components.

The Max s-core always contains far fewer airports than the Max k-core, typically one-half to one-third. In addition, most airports in the Max s-core are in the Max k-core. Thus, airports with high traffic tend to have also many connections.

9.2 Interregional Analysis

Figure 4.13.A shows the Max s-core of the large global component, and the table 4.15 lists its 22 airports. These airports are in North America, Europe, and East and South East Asia. They are located in capital cities and are the main airports of their country. Suvarnabhumi Airport in Thailand has the lowest traffic with 13196 flights among the Max s-core airports. Three US airports in the Max s-core are not in the top 22 Strength airports (General Edward Lawrence Logan Airport, Dallas Fort Worth Airport, and Seattle Tacoma Airport). Thus, the airports that concentrate the majority of inter-regional traffic share many flights. The Max s-core and the Max k-core have comparable sizes. However, the Max k-core includes more countries. Three airports (General E L Logan Airport, Dallas Fort Worth Airport, and Seattle Tacoma Airport) are in the Max s-core and not in the Max k-core. These airports have many flights and a moderate number of destinations compared with the other airports in the Max s-core.

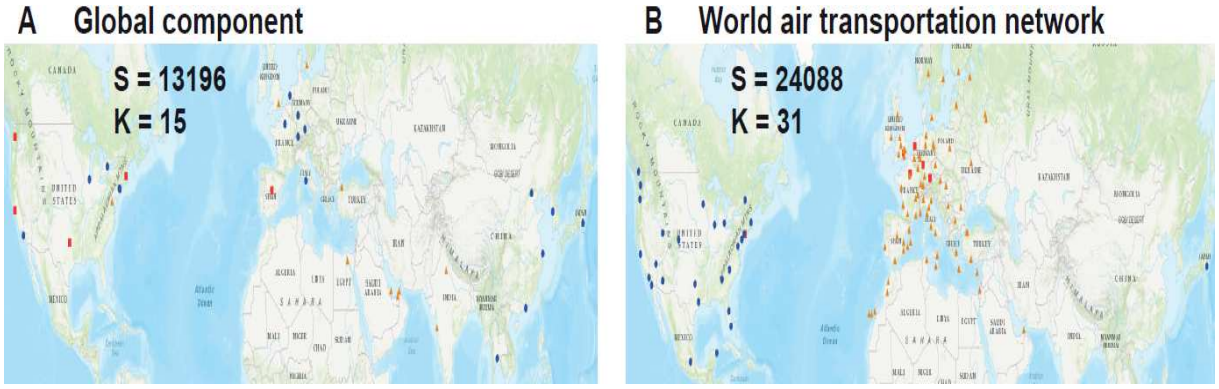


FIGURE 4.13 – A) The 33 airports in the maximum s-core of the weighted world air transportation network. B) The 22 airports in the maximum s-core of the weighted large global component. S is the maximum s-core value, and K is the maximum k-core value. Blue points represent the airports belonging to the maximum s-core and the maximum k-core, the red points are the airports exclusively in the maximum s-core, and the yellow points are the airports belonging only to the maximum k-core.

TABLE 4.17 – Summary table of the main results

		Weighted	**Unweighted**
Community structure	Algorithms	Louvain	Louvain
	Communities	17 communities	27 communities
Component structure	Local components	5 large components	7 large components
		North and Central America-caribbean	North and Central America-caribbean
		Africa-Middle East-Southern Asia	Africa-Middle East-Southern Asia
		South America	South America
		Europe-Russia-Central Asia	Europe
			Russia
		East and Southest Asia-Oceania	East and Southeast Asia
			Oceania
		12 small components	20 small components
	Global components	1 large	1 large
		11 small components	9 small components
Average Clustering coefficient versus Degree	Local component	No correlation with the degree	Decrease as a function of degree
	Global component	No correlation with the degree	Decrease as a function of degree
	Worldwide	No correlation with the degree	Decrease as a function of degree

Table 4.17 – Summary table of the main results

		Weighted	**Unweighted**
Average Degree-degree correlation versus Degree	Local components	North and Central America-Caribbean, South America, Europe-Russia-Central Asia, and the East and Southeast Asia-Oceania are more and more disassortative as the degree increases	
		Independent of the Degree for the Africa-Middle East-Southern Asia component	
	Global component	Disassortativy increases with the degree	Disassortativy increases with the degree
	Worldwide	Independent of the Degree	Independent of the Degree
Degree Strength Distribution	Local components	Log-Normal	The Africa-Middle East-Southern Asia, the Oceania, and the Russia-Central Asia-Transcaucasia components follow a Log-Normal law
			The North America-Caribbean, and the East and Southeast Asia components follow a Stretched Exponential
			Europe and the South America components follow a Truncated Power Law
	Global component	Log-Normal	Truncated Power-Law
	Worldwide	Log-Normal	Truncated Power-Law
Centrality	Local components	Top strength airports and top degree airports of the North and Central America-Caribbean and the East and Southeast Asia-Oceania components are identical	
		Top strength airports and top degree airports of the Europe-Russia-Central Asia, Africa-Middle East-Southern Asia and the South America components are different	
	Global component	Top strength airports and top degree airports of The components are identical	
	Worldwide	Top strength airports and top degree airports of the components are different	
Core	Local components	Some airports in the USA and Brazil cover the s-core and the k-core of the North and Central Asia-America and the South America	

Table 4.17 – Summary table of the main results

		Weighted	**Unweighted**
		The s-core and the k-core of Europe-Russia-Central Asia, Africa-Middle East-Southern Asia is distributed between several countries	
Core		The s-core the East and Southeast Asia-Oceania is distributed between the two sub-regions	Mainly the China airports cover the k-core of the East and Southeast Asia-Oceania
	Global component	3 regions (North and Central Asia-America, Europe-Russia-Central Asia, East and Southeast Asia) cover the s-core	Covering by all the components
	Worldwide	North and Central Asia-America cover mainly the s-core	Europe-Russia-Central Asia cover mainly the k-core

9.3 Comparison with the world air transportation network

Figure 4.13.B represents the Max s-core of the global air transportation network. With 27 airports, North America dominates. There are also five airports in Europe and one in Japan. There are 11 airports listed in Table 4.16 in the Max s-core of the North and Central America-Caribbean region. Minneapolis-St Paul Airport, with its 24088 flights, has the lowest traffic. Seven airports of the Max s-core of the world air network are not in the top 33 strength airports. The Max s-core global air network has half as many airports as the Max k-core. Moreover, the airports in the Max s-core are mostly absent in the Max k-core. Indeed, the Max k-core is more concentrated in Europe. These results confirm the orientation for high traffic in the USA and high number of destinations in Europe. Overall, with their high traffic, North American airports lead the weighted world air transportation network, while European airports dominate the unweighted world air transportation network with their high number of destinations. The component structure eliminates this disparity and reveals other important airports worldwide.

10 Discussion and Conclusion

This paper investigates the relationship between the weighted and unweighted worldwide air transport network and its impact on its component structure. Table 7.4 summarize the main findings. Overall, the weighted network contains fewer components. In both cases, the large local components cover distinct geographical areas. However, their geographical repartition differs slightly. Indeed, the weighted network has five large local components, while the unweighted network contains 7. Three components are quite similar (North and Central America-Caribbean, Africa-Middle East-South Asia, and South America). The two other components (Europe-Russia-Central Asia, East-South-East Asia-Oceania) group neighbor components of the unweighted network. Russia-Central Asia is attached to Europe, while Oceania joins East-South-East Asia. It is due to the substantial traffic between these regions linked to their economic integration. One can observe the same type of phenomenon more locally. Indeed, some major airports integrate regions with which they share a high proportion of their flights. For example, John F Kennedy

Airport in the United States belongs to the Europe-Russia-Central Asia component, London Heathrow is in the North and Central America-Caribbean component, and Frankfurt airports in Europe in the Africa-Middle East-South Asia component.

Although there are more global components in the weighted network than in the unweighted, there is also a single large global component in both cases. Although they are of comparable size (around 20 % of the world's airports), their content is quite different. Indeed, when components merge, many airports that were linking them disappear from the global component. Additionally, new airports emerge when an airport does not belong to its natural geographical area. For example, most of the airports linked to John F Kennedy located outside the Europe-Russia-Central Asia component integrate the global component.

Analysis of the average weighted clustering coefficient reveals it is independent of the degree for the large local components. Furthermore, their values are always lower than their unweighted equivalent. Low-traffic airports form triplets in the components and the global air network. In contrast, the average unweighted clustering coefficient decreases monotonically with the degree. Indeed, low-degree nodes tend to have a higher clustering coefficient than hubs. It reflects the lack of adequacy of the Hub and Spoke configuration for rerouting passengers when necessary. The Africa-Middle East-Southern Asia and the South America components have fewer interconnected hubs than the other local components. Consequently, their configuration is less hub and spoke. The global component exhibits similar behavior. However, the hub and spoke effect is more pronounced. Indeed, as it contains more hubs, it also has fewer triplets accentuating the rerouting inefficiency of the component.

According to the evolution of the average weighted degree-degree correlation as a function of degree, one can classify the components into two categories. The first category includes the Africa-Middle East-Southern Asia Component. In that case, the degree of the nodes does not exert a significant influence. The second category contains the four other local components and the global component. In these components, the weighted degree-degree correlation tends to decrease as the degree increases. The high-degree nodes link more and more with small-strength nodes. In other words, as the degree increases, these components' Hub and Spoke Configuration are more pronounced. Africa-Middle East-Southern Asia Component departs from this behavior because air transportation is less mature in this region. The world air transportation network exhibits similar behavior to the Africa-Middle East-Southern Asia Component. The same observations are also valid for the unweighted network suggesting that weights are not essential for this property. Additionally, For all the local components and the global air network, it appears that $K_w(i) \geq K(i)$. Consequently, airports with high degrees manage a higher number of flights.

The strength of the components follows a Log-Normal distribution with heavy-tailed characteristics. The strength as a function of degree shows that numerous hubs in the Europe-Russia-Central Asia component manage a comparable number of flights. Some hubs have a lot of traffic in the North-Central America-Caribbean and East-Southeast Asia-Oceania components. Compared to the other components, Africa-Middle East-Southern Asia and South America have fewer hubs and less traffic. However, these hubs handle a high fraction of flights. The strength distribution of the global component and the world air transportation network is also Log-Normal. Moreover, their hubs also manage most of the traffic.

The centrality analysis shows that the top five high-strength airports are usually in the leading countries of their region (the USA in North and central America-Caribbean, China in East and Southeast Asia-Oceania, India in Africa-Middle East-Southern Asia, Brazil in South America). It is more homogeneous in the Europe-Russia-Central Asia component. Furthermore, airports included in components far from their geographical areas are in the top five high-strength airports. The top high-degree airports serve the higher number of flights in the North and central America-Caribbean and the East and Southeast Asia components. It is not the case for the other local components.

The top high-strength inter-regional airports are in several countries except for the USA, which dominates the inter-regional traffic in the North and central America-Caribbean area. The top five strength

inter-regional airports differ from the leading regional airports. It shows that airports are more or less specialized in regional or inter-regional destinations. Furthermore, with a high fraction of the top inter-regional airports, Europe leads the inter-regional traffic. In addition, the top high-degree airports and high-strength airports are similar. Centrality analysis of the world air transportation network shows that North American airports dominate the traffic while Europe-Russia-Central Asia dominates the destinations.

Two categories appear in the max s-core analysis of the local components. The first includes the Europe-Russia-Central Asia and the East and Southeast Asia components, with the maximum s-core covering several countries. In the second category, a few countries (one to three) concentrate on the high traffic of the component. The Max s-core of the local components contains far fewer airports (half or one-third) than the Max k-core. Thus, few airports share numerous flights among the airports in the local components with several routes.

The maximum s-core of the global component covers several countries from the North and central America-Caribbean, Europe-Russia-Central Asia, and the East and Southeast Asia-Oceania regions. These three regions serve the most inter-regional flights. In contrast to the local components, the maximum s-core and k-core of the global component have comparable sizes. Consequently, the inter-regional hubs of these components concentrate several numbers of flights.

The Max s-core of the global air network is mainly located in the USA, while the Max k-core airports are in Europe. Then, the most important airports in the s-core and the k-core are blurred when considering the global air network instead of the component structure.

This comparative analysis illustrates the essential contribution of the component structure representation for uncovering the regional and inter-regional similarities and differences of the world air transportation network. Indeed, typical network properties have been designed for networks with a homogeneous density. Because of their local density variations, some characteristics can be blurred hence the importance of decoupling local from global analysis. This representation opens multiple research directions.

Now, we exploit the component structure to gain a better understanding of the robustness of the air transportation network against targeted attacks in the following chapters. Indeed, one can design tailored attacks on the components and inspect their global, regional or inter-regional impact.

Chapter 5

Targeted attacks on the world air transportation network : impact on its regional structure [1]

Summary

Perturbations of the air transport network have a tremendous impact on many sectors of activity. Therefore, a better understanding of its robustness to targeted attacks is essential. The literature reports numerous investigations at different levels (world, regional, airline) considering various targeted attack strategies. However, few works consider the mesoscopic organization of the network. To fill this gap, we rely on the component structure recently introduced in the network literature. Indeed, the world air transportation network possesses seven local components capturing the regional flights in localized areas. Its global component, distributed worldwide, capture the interregional routes. We investigate the impact of two influential attacks (Degree, Betweenness) on the world air transportation network at the regional and inter-regional levels. Results show that the seven regions are isolated one after the other from the world air transportation network. Additionally, although the Betweenness attack effectively splits the network, its impact on regional routes is less pronounced.

1. The results contained in this chapter are published in the paper :
[Diop, 2022c] Issa Moussa Diop, Chantal Cherifi, Cherif Diallo, Hocine Cherifi. "Targeted attacks on the world air transportation network : impact on its regional structure". Complex Networks XIII : Proceedings of the 13th Conference on Complex Networks, CompleNet Live 2022, Exeter, UK, May 30-June 1, 2022. Springer Nature. https ://doi.org/10.1007/978-3-031-17658-6$_1$.

1 Introduction

Every day, millions of passengers and frets transit by air transport. This infrastructure is essential in people's lives for economic, social, and health purposes. A disturbance can have significant consequences in different sectors of activity. Therefore, it is vital to study its disruption to limit the damage. The complex network paradigm attempts to provide solutions in this sense. Indeed, representing the airports as nodes and the flights between two airports as links, the robustness of the air transportation network has been extensively studied. One uses a random attack to account for accidental service disruption in an airport. In this case, one evaluates the impact of removing nodes at random on the network topology. Robustness studies also consider targeted attacks. In this case, the goal is to elaborate a node removal strategy to damage as much as possible the network. Besides, one can distinguish studies either linked to geographical areas (worldwide, regional, national) or airlines in the literature. To contextualize our work, we briefly describe important related works. For more details about the robustness of the air transportation network, the reader can refer to [Lordan, 2014a ; Sun, 2021].

In [Sun, 2017a], the authors conduct an extensive analysis of the robustness of the world air transportation network. They use six centrality measures (Betweenness, Closeness, Eigenvector, Bonacich, and Damage) to quantify the importance of airports. They consider two attack strategies. A strong attack removes the nodes in descending order of the centrality measure, while a weak attack uses the inverse order. They perform the experiments on three networks : 1) the unweighted network, 2) the network weighted by the number of passengers, 3) the network weighted by the inverse distance between airports. They compare three robustness metrics when removing a node : 1) the size of the giant component, 2) the number of survived links, 3) the number of unaffected passengers with rerouting. Results show that the size of the giant component is not appropriate for evaluating weak attacks' robustness. Furthermore, it overestimates the robustness of strong attacks. In comparison, survived links fix these two anomalies. In addition, according to the unaffected passenger with rerouting, Degree centrality is the most effective attack when 7% of nodes are disconnected. Bonacich takes the lead when more than 7% are disconnected. The robustness is sensitive to the type of weight of the network. In [Lordan, 2019], the authors investigate the robustness of the seven regional unweighted networks defined by the OAG (Africa, Asia, Europe, Latin America, Middle East, North America, and South Pacific). They perform strong attacks based on Degree, Betweenness, and Damage. The size of the giant component evaluates the robustness. Results show that Damage is the most effective attack for a small fraction of removed nodes. However, Betweenness performs better when the number of removed nodes grows. Results also show evidence that differences across regions are related to the size of the value of the k-core. Areas with a large core of densely connected airports, such as Europe and the Middle East, are more resilient than regions with a smaller core. In [Zhang, 2021], the authors investigate the robustness and weighted (number of flights per week) and unweighted Belt and Road region network. Strong attacks on the unweighted network use Degree, Betweenness, Closeness, and Eigenvector centrality. For the weighted network, they investigate strong attacks based on recursive power and recursive centrality. They test four robustness metrics : clustering coefficient, average shortest path length, graph diversity, and global efficiency. Results show that the Betweenness centrality is the least robust for the clustering coefficient, average shortest path, and global efficiency for the unweighted network. The Degree is less resilient for the graph diversity. Recursive power is more sensitive to all the network properties for the weighted network. By comparing the weighted and unweighted network's robustness, no targeted attack dominates entirely. Indeed, the Betweenness centrality is less robust for the average shortest path and the graph diversity. The Degree is sensitive to graph diversity. Finally, the recursive power is less resilient for the clustering coefficient. In [Lordan, 2016], the authors investigate the topology and robustness of unweighted networks of airlines. They study 10 Full-Service Carriers belonging to three airline alliances (Star Alliance, OneWorld,

and SkyTeam) and 3 Low-Cost Carriers (Ryanair, EasyJet, and Southwest Airlines). Random and targeted strong attacks based on Degree and Betweenness centrality are performed. The size of the giant component is the robustness metric. Low-Cost Carriers are more resilient than Full-Service Carriers to random attacks, but the difference is not significant. The attack based on the Degree is more resilient than the attack based on the Betweenness. But the difference is small. The hybrid model airlines tend to be more robust than the Low-Cost Carriers airlines, which resist better targeted attacks than the Full-Service Carriers airlines.

Our work departs from previous studies. It concerns the interactions between the unweighted world air transportation network and its regional components. In earlier work, we introduced a network decomposition called the component structure of a network [Diop, 2021]. It decomposes a network into its local components and global components. Local components are localized dense areas of the original network. The links joining the local components and their associated nodes form the global components. Based on this representation, air world transportation comprises several regional components corresponding to natural geographic and cultural areas. The inter-regional components reveal the main airports and routes between these regions. Based on this representation, we explore the impact of targeted attacks on the world air transportation network on its regional and inter-regional components. We consider Degree and Betweenness centrality measures to remove nodes in descending order. These experiments give new insight into the interactions between international and regional routes exposed to disruption.

The rest of the chapter is organized as follows. Section 2 introduces the background. Section 3 presents the data and methods. Section 4 reports the main findings of our analysis. Section 5 discusses the results and then gives conclusions.

2 Background

2.1 Component structure

The density of real-world networks is generally not uniform. One usually captures this phenomenon using two mesoscopic features : 1) the community structure, 2) the core-periphery structure. Although there is no consensus on a universal definition of these representations, they share the fact that the network contains groups of nodes tightly connected called core or communities. They are supposed to be loosely related to other groups when considering the community structure. Peripherical nodes sharing few connections surround these core groups in the multi-core-periphery structure. The component structure builds in these representations. It splits the networks into dense groups and their interactions. One obtains the local components by isolating the dense parts of the networks. Links and nodes connecting the local components form the global components. To build the component structure one proceeds as follows :

1. Uncover the dense parts of the network.
2. Remove the links between the dense parts to extract the local components.
3. Remove the links within the dense parts and the subsequently isolated nodes to extract the global components.

Note that this representation is redundant. Indeed, a node can simultaneously belong to a local and a global component. One can use community detection or multi-core-periphery algorithms to extract the dense parts of the network. We consider an approach based on the community structure to extract the components in this work. Fig. 5.1.A describes the extraction process of the component structure. In this example, one uses a non-overlapping community detection algorithm to extract the dense parts of the network. Then, we form the local components by removing the inter-community links. Removing the intra-community links and the isolated nodes extracts the global components.

2.2 Targeted Attack

Targeted attacks aim to remove the most vital nodes for network connectivity[Chakraborty, 2016]. Centrality measures generally describe the importance of nodes [Ibnoulouafi, 2018]. In a strong attack strategy, one removes nodes in the network in descending order of magnitude of the chosen centrality. This work uses the most popular measures : Degree and Betweenness.

Degree centrality of a node is the number of its first-order neighbors. Given a graph $G(V, E)$, such as V is the set of nodes and E the set of links, the Degree k_i of node i is defined as :

$$k_i = \sum_{j \in V, i \neq j} a_{ij}$$

a_{ij} is an element of the binary adjacency matrix of G such as $a_{ij} = 1$ if i and j are connected, otherwise, $a_{ij} = 0$.

Betweenness centrality of a node is the fraction of the shortest path passing through it. When it is normalized, the Betweenness of the node i is defined as :

$$b(i) = \frac{2}{(n-1)(n-2)} \sum_{i \neq j} \frac{\sigma_{jk(i)}}{\sigma_{jk}}$$

σ_{jk} is the number of the shortest path between j and k. $\sigma_{jk}(i)$ is the number of the shortest path from j to k passing in i.

2.3 Evaluation measures

The size of the giant component is the most popular metric to assess the robustness of a network. It refers to the size of the largest set of interconnected nodes when the network brakes into several parts due to node removal. The higher the size of the giant component, the most resilient the network is to the attack.

3 Data and Methods

3.1 Data

We consider an unweighted and undirected network originating from FlightAware[Flightaware,]. Flight information has been collected for six days (between May 17, 2018, and May 22, 2018). Nodes represent airports, and links represent direct flights between airports during the period. Table 7.1 reports its basic topological properties.

3.2 Methods

Our goal is to evaluate the impact of a targeted attack on the world route network on its regional and inter-regional constituents. Therefore, once the component structure is extracted, the robustness evaluation process proceeds as follows :

1. Disconnect the node from the world air transportation network according to an attack strategy.
2. Disconnect the same node from its local component.
3. Disconnect the same node from the global component if it also belongs.
4. Extract the giant component from the world air transportation network.
5. Extract the giant component from the concerned local component.

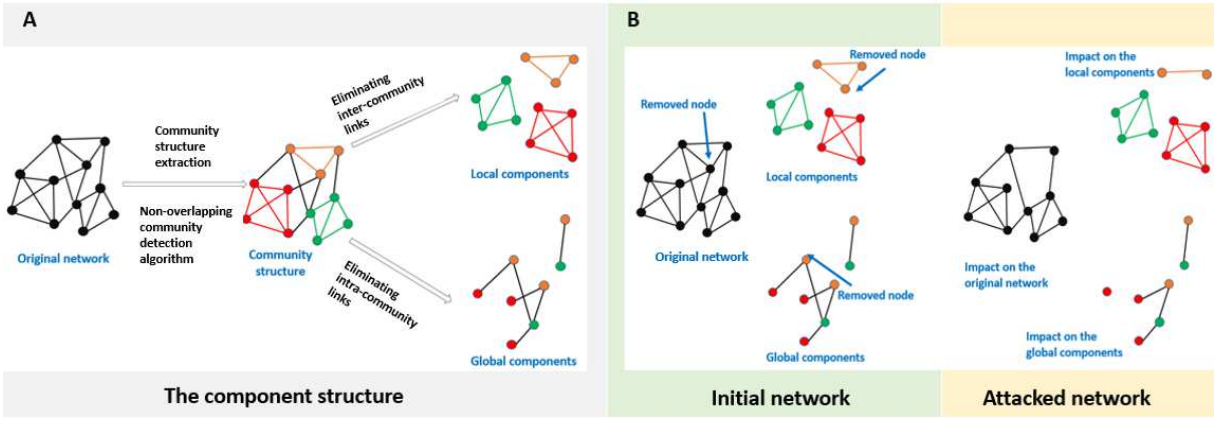


FIGURE 5.1 – (A) Process to uncover the component structure. The community structure is used for example, to extract the dense parts (B) Attack on the network.

6. Extract the giant component from the global component.

This approach allows us to visualize the impact of removing a critical airport in the world air transportation network on the robustness of the regional and inter-regional networks. Fig. 5.1.B reports a toy example illustrating an attack on an airport of the world network and the disruptions induced in the local and global components of the network.

4 Experimental results

4.1 Component structure

We rely on the Louvain community detection algorithm to uncover the dense parts of the network. Indeed, communities are tightly connected nodes. It reveals 27 communities. Therefore, we consider that there are 27 local components. There are seven large and twenty small components localized in various geographical areas. The largest local components cover the following regions : 1) North America-Caribbean, 2) Europe, 3) East and Southeast Asia, 4) Africa-Middle East-Southern Asia, 5) Oceania, 6) South America, 7) Russia-Central Asia-Transcaucasia. Note that this subdivision is different than the partition proposed by the OAG (Official Airline Guide) used in [Lordan, 2019].

There are fifteen global components. The largest one regroups more than 96% of the airports, and it is distributed over the world [Diop, 2021]. Fig. 6.3 represents the airports included in the largest local and global components. We restrict our attention to the seven largest local components and the main global component in the following robustness analysis. Table 7.1 reports their basic topological properties.

4.2 Degree targeted attack

Fig. 5.3.A illustrates the evolution of the LCC with the fraction of top Degree nodes removed from the world air transportation network in descending order of the airports' Degree centrality. We also plot the induced evolution of the LCC of the various large components. All the curves exhibit similar behavior. As the fraction of nodes removed from the world air transportation network grows, the LCC of the components decreases almost linearly. Overall, the European component is the most resilient. The East and Southeast Asia components and the World Air Transportation Network follow. They behave similarly when the fraction of removed nodes is below 5%. Beyond this value, the gap between the evolution of their LCC with the European component widens. Africa-Middle East-Southern, Russia-Central Asia-

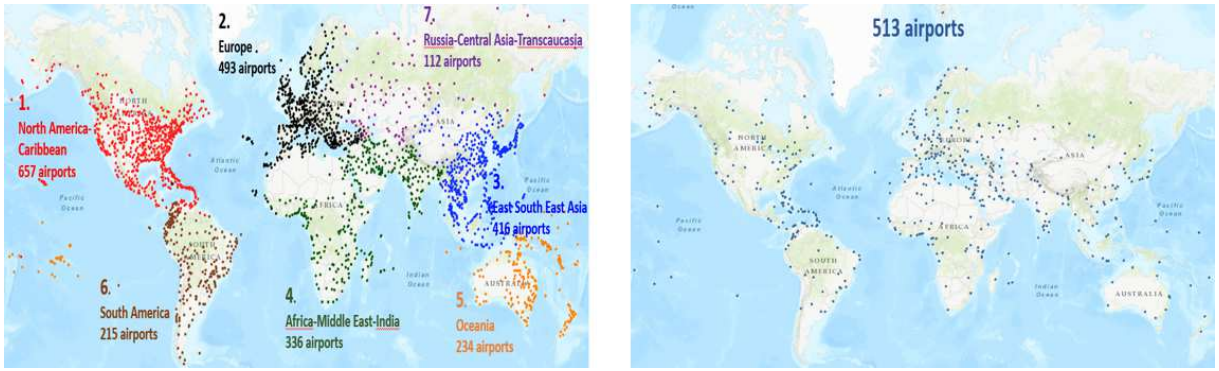


FIGURE 5.2 – Left figure :The airports in the seven large local components. Each color represent a component. Right figure : The largest global component.

TABLE 5.1 – Basic topological properties of the world air transportation network, the 7 large local components, and the largest global component. N is the network size. $|E|$ is the number of edges. $diam$ is the diameter. l is the average shortest path length. ν is the density. ζ is the transitivity also called global clustering coefficient. $k_{nn}(k)$ is the assortativity also called Degree correlation coefficient. η is the hub dominance.

Components	N	$\|E\|$	$diam$	l	ν	ζ	$k_{nn}(k)$	η
World air transportation network	2734	16665	12	3,86	0,004	0,26	-0,05	0.09
North America-Caribbean	657	3828	7	2.88	0.018	0.28	-0.325	0.29
Europe	493	5181	6	2.58	0.042	0.32	-0.2	0.33
East and Southeast Asia	416	2495	7	2.9	0.029	0.34	-0.22	0.32
Africa-Middle East-India	337	1197	6	3.25	0.021	0.28	-0.15	0.23
Oceania	234	464	9	3.5	0.017	0.18	-0.22	0.24
South America	215	527	6	3.15	0.023	0.23	-0.35	0.22
Russia-Central Asia-Transcaucasia	112	427	4	2.23	0.07	0.23	-0.39	0.73
Large global component	513	2194	8	3.28	0.017	0.13	-0.25	0.2

Transcaucasia, and the largest global components form a group of overlapping curves when the fraction of removed nodes is below 7.5%. Beyond this value, they diverge. The North America-Caribbean component follows. Finally, Oceania and South America components are the most sensitive to the attack. Indeed, their LCC breaks down to 33% and 28% respectively, after removing around 4.7% of the airports. A more detailed observation shows two types of behavior. In the first case, the LCC exhibits piecewise linear variations. It concerns South America, Oceania, and Russia-Central Asia-Transcaucasia. Removing a particular node produces a sharp drop in the LCC because an entire subnetwork broke away from the primary component, causing considerable damage. In contrast, in the other components and the world air transportation network, the size of the LCC varies almost linearly with the fraction of removed nodes. The same proportion of nodes leaves the LCC, whatever the node removed in this situation.

We now look at when the components become isolated from the world air transportation network. We identify the critical airport disconnecting the component. We explore the topological properties of the remaining regional network to evaluate the impact on the regional traffic. Table 5.2 displays the basic topological properties of the components after isolation. Oceania is the first component to separate from the world air transportation network. It occurs after removing 9% of the top Degree nodes from the world air transportation network. The proportion of top Degree nodes of this component subtracted before the split equals 4.7%. Christchurch, in New Zealand, is the last top-Degree airport removed, provoking the

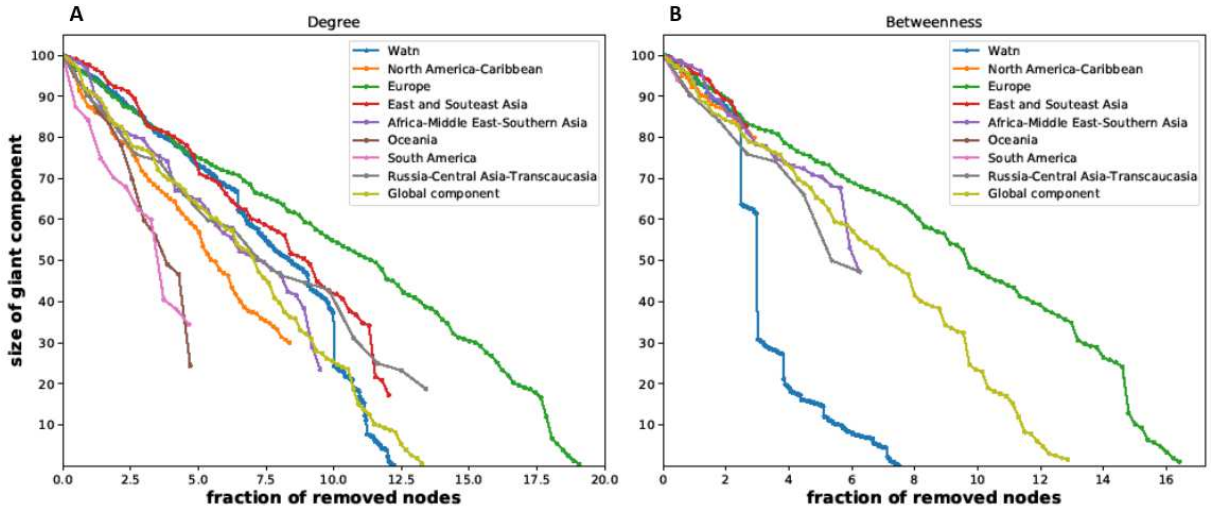


FIGURE 5.3 – The size of the giant component of the world air transportation network, and the large components as a function of removed nodes under targeted attacks on the world air transportation network.

TABLE 5.2 – Basic topological properties of the 7 large isolated local components after the attack based on the Degree. LCC is the Largest Connected Component. $|E|$ is the number of edges. $diam$ is the diameter. l is the average shortest path length. ν is the density. ζ is the transitivity also called global clustering coefficient. $k_{nn}(k)$ is the assortativity also called Degree correlation coefficient.

Components	LCC	$\|E\|$	$diam$	l	ν	ζ	$k_{nn}(k)$	$LCC(\%)$
North America-Caribbean	196	301	21	7.22	0.015	0.21	-0.14	29.8
Europe	136	177	13	5.75	0.019	0.07	-0.13	27.5
East and Southeast Asia	72	104	11	4.51	0.04	0.18	-0.15	17.2
Africa-Middle East-India	79	110	17	6.9	0.035	0.26	-0.08	23.5
Oceania	57	65	15	5.1	0.04	0.12	-0.2	24.3
South America	74	98	13	5.7	0.036	0.21	-0.43	34.4
Russia-Central Asia-Transcaucasia	21	28	6	2.85	0.13	0.08	-0.11	18.7

separation. Only 24% of the airports remain in the LCC of this component. They are all in Australia. All the other countries are unreachable. Although the LCC is denser than the initial component, traveling in this region becomes more challenging. Indeed, Table 5.2 shows that the diameter and the average shortest path increase, and the transitivity decrease. The North America-Caribbean local component is the second to break away after removing 9.7% of the top Degree nodes from the world air transportation network. When it happens, 8.4% top Degree airports in this component have already been removed. Winnipeg /James A Richardson Airport in Canada is the last airport connecting the region to the world before the split. In contrast to Oceania, airports in the LCC are distributed in the different countries of the component. However, Alaska is isolated, and in Canada, there are only a few airports in Quebec and Ontario. The remaining LCC contains 30% of the airports of the initial component. Table 5.2 shows that travels in the region become uneasy. Indeed, eight hops on average and 21 at maximum are required to join any two airports. In addition, density and transitivity decrease. Disconnecting 9.8% of the airports in the global air network is sufficient to isolate the South America component. A proportion of 4.6% top Degree nodes are removed from the component before separation. Salgado Filho Airport (Brazil) is the

last link with the rest of the world before the break. The LCC contains 34% of the airports of the initial component. Most of its airports are in Brazil and its neighboring countries. Argentina is unreachable. Even though it is denser than the initial component, the shortest path and the diameter reveal that this network doesn't facilitate the traffic (see Table 5.2). Note that the transitivity doesn't fluctuate much. The next component that becomes unreachable is East and Southeast Asia. It breaks away after removing almost 11% of the top hubs from the overall network. Before the split, 12% of the top Degree nodes of this component have been removed. Gimhae Airport in South Korea is the last airport reachable before separation. The LCC contains 17% of the airports from the initial component. They are mainly in China and its satellite countries. Compared to the other isolated component, it is the easiest to travel, even though its transitivity is low. One needs to eliminate 11.2% of the high Degree airports of the world air network to isolate the Africa-Middle East-Southern Asia region. The targeted hubs include 9.5% of the airports of this component. The Benazir Bhutto Airport in Pakistan is the last connected to other areas in the world. The LCC contains 23.5% of the airports of the initial component. They are located mainly in West-Central Africa and India. Traveling into the LCC is difficult even though it is denser than the original component. Almost all the Middle East, South Africa, and East of Africa airports are unreachable. Moreover, a maximum of 17 hops are required to join any two airports, while on average, it is seven hops. Russia-Central Asia-Transcaucasia is the next component to separate from the world air transportation network. One needs to remove 11.4% of the top hubs of the world, including 13.4% of this component's hubs. The last removed airport, the Heydar Aliyev Airport, is in Azerbaijan. Only 18.7% of the airports initially in the component remains in the LCC. Except for two airports located in Uzbekistan and Tajikistan, All the others are in Russia. Compared to the initial component, the topological properties do not change much. Traveling in what remains of the component is as easy. Once Russia-Central Asia-Transcaucasia is isolated, all the airports remaining in the world air transportation are in Europe. It occurs after removing 19% of the top Degree nodes of the Europe component. The LCC contains 27.5% of the airports of the original component located in various countries. However, countries like Spain, Norway, Sweden, and Finland are almost unreachable. It is more difficult to travel within the LCC than in the initial component. Indeed, the diameter and the average shortest path values double. In addition, the density and the transitivity decrease significantly. Finally, there are no more routes in the world air transportation network after removing 12.2% of its most influential airports.

4.3 Betweenness targeted attack

Fig. 5.3.B displays the variation of the LLC as a function of the fraction of top Betweenness centrality airports removed from the world air transportation network. We also plot the corresponding evolution for the components. Removing up to 3% of top airports does not significantly differ in the various components. Indeed, the curves are very close. Nevertheless, Africa-Middle East-Southern Asia, East and Southeast Asia, and Europe are slightly more resilient. Above this value, Europe appears clearly as the most robust component. One can still distinguish the two types of behavior observed in the previous experiment. The LCC varies almost linearly in the Europe component and to a lesser extent in the global component. Piecewise linear variation with sharp drops characterizes the others and, more particularly, the World air transportation network.

South America is the first isolated area after removing 1.6% of the world's major airports. The last link with the World is Comodoro A M Benítez airport in Chili. Only two airports in this region are in the top 1.6% world airports (Guarulhos-Governador, André F Montoro in Brazil, and Comodoro A M Benítez airport in Chili). The LCC contains 93.5% of the airports of the initial component scattered in all the countries. The topological properties of the LCC do not change significantly as compared to the initial component, and the regional traffic keeps its efficiency. Removing 2.1% of the top airports disconnect

TABLE 5.3 – Basic topological properties of the 7 large isolated local components after the attack based on the Betweenness. LCC is the Largest Connected Component. $|E|$ is the number of edges. $diam$ is the diameter. l is the average shortest path length. ν is the density. ζ is the transitivity also called global clustering coefficient. $k_{nn}(k)$ is the assortativity also called Degree correlation coefficient.

Components	LCC	$\|E\|$	$diam$	l	ν	ζ	$k_{nn}(k)$	$LCC(\%)$
North America-Caribbean	523	1918	8	3.23	0.14	0.2	-0.27	79.6
Europe	305	1252	8	3.39	0.027	0.16	0.014	61.8
East and Southeast Asia	343	1421	8	3.51	0.24	0.32	-0.11	82.4
Africa-Middle East-India	160	290	14	5.29	0.022	0.26	-0.25	47.6
Oceania	211	383	9	3.6	0.017	0.17	-0.21	90.2
South America	201	449	7	3.32	0.022	0.21	-0.38	93.5
Russia-Central Asia-Transcaucasia	53	105	5	2.84	0.07	0.2	-0.32	47.32

Oceania from the world transportation network. Perth airport is the last connection before isolation. Only one airport (Sydney K Smith) before Perth has been targeted before isolation. The LCC retains 90.2% of the airports from the initial component across all its countries. The topological properties do not change much, and the internal travels in this component are not much disturbed. East and Southeast Asia is the next component to become unreachable from the other parts of the world after removing 2.3% of the world's most important airports. Among them, there are 2.6% of the airports in this region. The last connection is through the Kunming Changshui Airport in China. The LCC contains 82.4% of the airports initially in the component. They are located in the various countries of this component. Compared to the initial component, the topological properties of the LCC are comparable. Therefore, the impact of the isolation on regional routes is limited. North America-Caribbean is the fourth isolated area. It becomes unreachable after removing 2.9% of the top central world airports. Among these, there are the top 2.9% of this component. Simón Bolívar Airport in Venezuela is the last liaison to the world before isolation. The LCC contains 79.6% of the airports still interconnected. Note that the Alaska subregion is inaccessible. The LCC is a bit less efficient than the component. Indeed, it keeps some large hubs, except in Canada. The topological properties of the LCC indicate that it is more challenging to travel than in the initial network. The next isolated area is the Africa-Middle East-Southern Asia, when one removes 3.9% of the central airports from the world network. Around 5.6% of the airports from this component are the target of the attack. As for the Degree centrality, Benazir Bhutto Airport in Pakistan is the last removed. The LCC maintains 47.6% of the airports in this region connected. Once again, the Middle East and the Horn of Africa are the most impacted. They are almost unreachable. Traveling within this region becomes difficult. Indeed, the diameter (14 hops) and the average shortest path (6 hops) of the LCC are two times higher. The density and the transitivity are comparable with the initial component values. Russia-Central Asia-Transcaucasia is the next region to break away from the world network after removing the top 5% central airports of the global air transportation network. The attack involves 6.2% major airports of this component before the separation. They are mainly in Russia. The last removed airport is Krasnodar Pashkovsky Airport. The size of the LCC reduces to 47.3% of the initial component. Comparing its topological properties to the initial component allows us to conclude that the impact on regional easiness to travel is limited. After splitting with Russia-Central Asia-Transcaucasia, Europe preserves 61.8% of its airports in the LCC. They are scattered in all the countries of the initial component. Compared to the latter, the LCC is less dense. In addition, one requires more hops to join any two airports.

5 Discussion and Conclusion

In this work, we leverage the component structure of the world air transportation network to analyze the impact of targeted attacks on the regional and inter-regional routes. We consider attacks based on two influential centrality measures. The first one, based on Degree, removes nodes according to their number of connections in descending order. The second one uses the proportion of shortest paths transiting through a node to rank the nodes. We perform the experiments on an undirected and unweighted network to focus on routes rather than flights or passengers.

Results show that whatever the attack on the world air transportation network, one can link the way the network disintegrates to the component structure. Indeed, components get isolated one after the other when removing a certain proportion of the top centrality nodes from the world air transportation network. The main differences between the two types of attack are the fraction of top removed nodes needed for a component to break away from the world network and the size of the remaining isolated component. Globally, one needs to target more nodes before isolation in the Degree attack than the Betweenness attack. Furthermore, the isolated components' size is smaller in the Degree of attack. Indeed, it prioritizes the top internal hubs in the components, reducing their size. One needs to remove many internal high Degree nodes before reaching the inter-regional airports tying the components together. In contrast, Betweenness centrality focuses on inter-regional airports, splitting earlier the components with minor damage to their internal structure. One after the other, South America, Oceania, East, Southeast Asia, and North America are isolated regions by the Betweenness attack. Oceania, North America, South America, and East and Southeast Asia leave the world network in this order when targeting nodes by Degree centrality. Then Africa-Middle East-Southern Asia, followed by Russia-Central Asia-Transcaucasia separate from Europe in the two attack strategies. Note that economic and geographic ties connect these two groups of regions.

The component structure allows us to study the robustness of the world air transportation network from a new perspective. Indeed, it is a mixture of subnetworks with various internal densities geographically well identified. Looking at its robustness through its component structure allows us to highlight the impact of targeted attacks in different world areas. These results will enable us to tailor the protection strategies to maintain inter-regional routes and minimize the effect of disruption at the regional level.

In the next chapter, we investigate the impact, on the components, of the attacks based on the weighted centrality on the global air network.

Chapter 6

Robustness of the Weighted Air Transport Network [1]

Summary

The robustness of a system indicates its ability to withstand disturbances while maintaining its properties, performance, and efficiency. There are plenty of studies on the robustness of air transport networks in the literature. However, few works consider its mesoscopic organization. Building on the recently introduced component structure, we explore the impact of targeted attacks on the weighted world air transportation network on its components. Indeed, it contains five local components covering different regions (North America-Caribbean, Europe-Russia, East and Southeast Asia-Oceania, Africa-Middle East-Southern Asia, South America) and one global component linking these regions. We investigate targeted attacks based on influential weighted centrality measures (Strength, Betweenness, PageRank). Results show that the local components gradually separate from the world air transportation network as the fraction of removed airports grows. The Weighted Betweenness attack removes fewer top airports to isolate the regions compared to its alternatives. Furthermore, it is still convenient to travel locally in the separated areas. In contrast, Strength and PageRank attacks need to target more airports to split the network. However, they are more disruptive. Indeed, the size of the isolated local components reduces drastically, so it becomes more challenging to travel locally. Looking at the world air transportation network through its component structure reveals a new viewpoint on its resilience. It opens new perspectives to design more efficient attacks.

1. The results contained in this chapter are published in the papers :
[Diop, 2022a] Issa Moussa Diop, Chantal Cherifi, Cherif Diallo, Hocine Cherifi. "Robustness of the Weighted World Air Transportation Network Components". 2022 IEEE Workshop on Complexity in Engineering (COMPENG), 18-20 July 2022, Florence, Italy, pp. 1-6.
[Diop, 2022e] Issa Moussa Diop, Cherif Diallo, Chantal Cherifi, Hocine Cherifi. "Robustness Analysis of the Regional and Inter-Regional Components of the Weighted World Air Transportation Network". Complexity, vol. 2022, Article ID 6595314, 17 pages, 2022. https ://doi.org/10.1155/2022/6595314.

1 Introduction

The air transportation network is essential for the development and integration of countries. Indeed, millions of people and tons of frets from different world regions pass through the air. Ensuring secure and effective air transport infrastructure and traffic is a vital issue. The complex network paradigm allows for a better understanding of an extensive range of complex interconnected systems such as infrastructure, economics, and social networks [Jusup, 2022]. Considering the air transportation networks where nodes represent airports and links are flights between two airports, network robustness is a popular tool to study the response to possible disruptions. Disruptions can be accidents or intentional attacks on airports or flights. Numerous investigations on the robustness of air transport networks (worldwide, regional, and national airlines) have been conducted [Lordan, 2014a ; Sun, 2021 ; Aparicio, 2022 ; Zhang, 2019 ; Verma, 2014 ; Wandelt, 2021]. In the following, we concentrate on targeted attacks based on centrality measures. It consists in ranking the airports according to a given centrality measure and removing the nodes in descending order of their centrality value. The following paragraph summarizes some important related works.

In [Lordan, 2014b], the authors explore the robustness of the global air transportation network. They investigate five centrality measures (Degree, Betweenness, Modal Analysis, Damage, and Bonachich Power). They measure the effectiveness of the attacks using the size of the Largest Connected Component (LCC). Indeed, the attacks split the network into multiple components. Results show that Damage is the most effective when a small proportion of airports are attacked ($< 2.5\%$). Beyond this value, Betweenness takes the lead. The other strategies are less effective. Indeed, Modal Analysis is good when more than 10% of airports are isolated. For larger fractions of isolated airports, Bonachich Power outperforms Degree. In [Sun, 2017a], the authors investigate the robustness of the directed and weighted world air transportation network. Besides the unweighted network, they consider two weighting schemes : the geographical distances and the number of passengers traveling between two airports. They evaluate six targeted attacks based on various centralities ordered in descending and ascending order (Strength, Betweenness, Closeness, Eigenvector, Damage, and Bonacich). Moreover, three metrics (unaffected passenger with rerouting, survival links, size of the largest connected component) quantify the robustness. The unaffected passenger with rerouting is a baseline measure. The survival links measure introduced in this chapter is the proportion of remaining links. The correlation analysis between these metrics shows that the unaffected passengers with rerouting and survival links are well correlated. The world air transportation network is more sensitive to attacks based on Bonacich and Degree. Attacks are more effective when one takes into account the number of passengers.

Numerous airports and routes have been disrupted during the COVID-19 pandemic. In [Bao, 2021], the authors explore the impact of the first wave on the structure of the global air network and on the epidemic spreading. They consider a network where nodes are cities and links are flights between the cities. They observe that the average distance between airports increases and long-distance flights decrease. The average betweenness and the number of triangles also drop significantly. They show that one can simulate this situation using a mixture of random and degree-based attacks. Furthermore, they show that politically motivated shutdowns lead to high variations in airport centrality and have consequences on national and international economies. Finally, the global air transport network started recovering its initial structure when the situation returned to normal.

In [Zhou, 2019], the authors investigate the resilience of eight weighted domestic air transportation networks (Russia, Brazil, Australia, Canada, India, China, The US, and European). The number of routes is the weight of each of these networks. The attacks target the airports according to the decreasing order of their Local Weighted Efficiency measure. The global weighted efficiency measures the robustness of the networks. They show that airports with high Weighted Efficiency have a large Degree and Strength.

Moreover, two groups of domestic air transportation networks appear. The first, including Australia, Canada, India, and the US, are the most vulnerable. Indeed, removing 5% of the most critical airports reduce the weighted efficiency to 70%. The second groups, including the other air transportation networks, are more resilient. Indeed, removing the same proportion of airports in these networks reduces only 40% of their weighted efficiency.

In [Kim, 2019], the authors explore the robustness of the Northeast Asian network (China, Japan, and South Korea). They use three types of networks, one unweighted and two weighted (distance, link-wise air traffic). Targeted attacks rely on six centrality measures (Degree, Betweenness, Closeness, Strength, Weighted Betweenness, and Weighted Closeness). Evaluation includes two robustness metrics (the size of the largest connected component and the number of operable flights with optional rerouting). The Northeast Asia network is seriously affected when the Weighted Betweenness attack concerns 0.8% of airports. Indeed, its size decreases to around 50%. According to Betweenness, the same damage is detected when targeting the top 2% of airports. There is a considerable gap in operable flights if one considers rerouting or not for all the weighted attacks. In [Zhang, 2019], the authors study the robustness of the air transportation network of the Belt and Road regions. Nodes are cities that contain an airport with more than 25 airlines and the country capitals. The number of flights between two destinations weights the network. Targeted attack strategies use Recursive Power and Recursive Centrality. The Efficiency measure assesses the impact of the attacks. In addition, the authors define an edge addition strategy to improve the network Efficiency. Results show that attacks based on Recursive power are the most effective. In addition, adding links with top Recursive Power nodes increase the network robustness.

Our work departs from these studies. Indeed, we analyze the resilience of the weighted world air transportation network from a new viewpoint centered on the component structure. Previous work shows that the local components correspond to delimited geographical, economic, and cultural areas. In contrast, the global components represent their interactions and cover the world [Diop, 2021]. Therefore, this study also concerns various local regions of the world. However, the attacks are not local. They target the entire world air transportation network independently of its component structure. It allows investigating the impact of targeted attacks worldwide on the different regional infrastructures. It shows that a high level of perturbations on the world air transportation network does not necessarily translate to poor travel experience at the regional level.

The rest of the chapter is organized as follows. Section 2 briefly introduces the basic elements of this study. Section 3 presents the data and methods. The section 4 analyzes the component structure. Section 5 reports the main findings on the Strength attack. Section 6 evaluates the Weighted betweenness attack. Section 7 assesses the Weighted PageRank attack. The Section 8 discusses the results. Finally, section 9 concludes the study.

2 Background

2.1 Component structure

The component structure attempts to capture the dense parts of the network. Indeed, in real-world networks, the density is inhomogeneous. The community [Jin, 2021b ; Cherifi, 2019b ; Orman, 2013] and core-periphery [Tang, 2019 ; Kojaku, 2017b] structures are good features to capture this phenomenon. Indeed, communities are dense clusters of nodes loosely connected. Cores are also dense parts of the multi-core-periphery structure [Kojaku, 2017a]. Peripheral nodes with few connections surround cores. The component structure is an alternative mesoscopic representation. It states that a network consists of local and global components. Local components are the dense parts. The global components regroup the nodes and links connecting the local components. Uncovering the component structure proceeds as

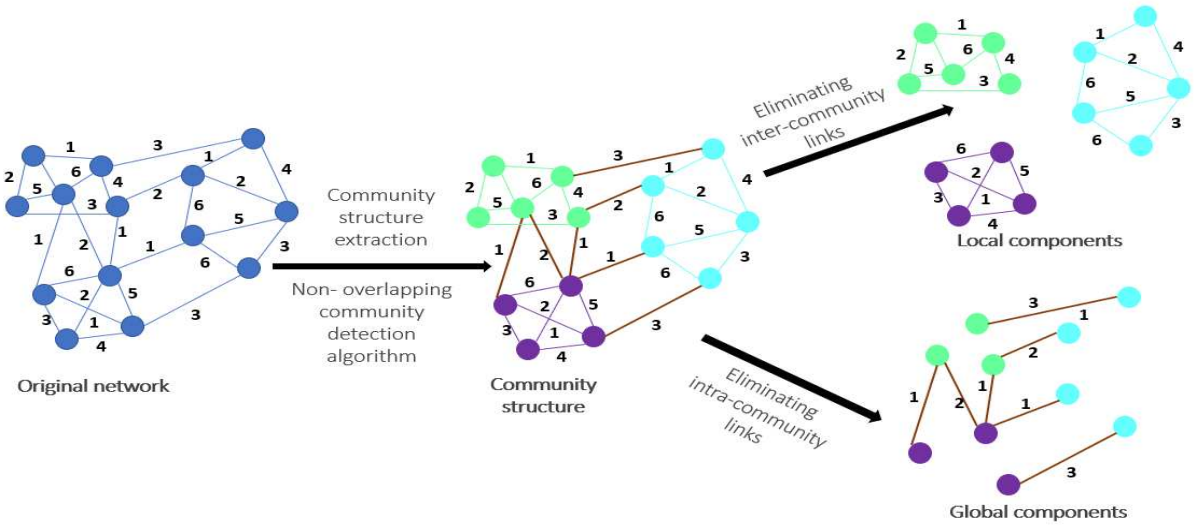


FIGURE 6.1 – This figure illustrates the process of uncovering the component structure using a community detection algorithm to extract the dense parts forming the local components. After revealing the local components, one builds the global components with the nodes and links shared by the local components. In this example, the network has three communities colored green, light blue, and dark blue. Therefore, it has three local components. The links and nodes joining these local components form three global components.

follows :

1. Extract the dense parts of the network.
2. Eliminate the links between the dense parts to extract the local components.
3. Eliminate the links within the local components and the subsequently isolated nodes to extract the global components.

Fig. 6.1 illustrates the extraction process of the component structure. In this example, one uses a weighted non-overlapping community detection algorithm to uncover the dense parts of the network. Then, we form the local components by removing the inter-community links. Removing the intra-community links and the isolated nodes extracts the global components. One can see that this representation is redundant. Indeed, a node can simultaneously belong to a local and a global component. However, it disentangles the local interactions from the global interactions.

2.2 Targeted Attack

One can consider two types of attacks in a network : random and targeted attacks. A random attack removes nodes randomly. It allows for studying the robustness of the network when an unexpected failure occurs. In contrast, targeted attacks aim to remove the most vital nodes for network connectivity [Wan, 2021 ; Wandelt, 2018a]. Centrality measures generally describe the importance of nodes [Lü, 2016 ; Ghalmane, 2019a ; Rajeh, 2020 ; Chakraborty, 2016 ; Kumar, 2018 ; Ibnoulouafi, 2018]. In a strong attack strategy, one removes nodes in the network in descending order of magnitude of a centrality measure. Weak attacks use the reverse order. This work evaluates the impact of strong attacks based on influential centrality measures for weighted networks : Strength, Betweenness, and PageRank

2.2.1 Weighted Strength Centrality

A node's strength centrality [Lü, 2016] is the sum of weights of its first-order neighbors. Given a graph $G(V, E, W)$, such as V is the number of nodes, E the number of links, and W is the set of weights, the Strength s_i of node i is defined as :

$$s_i = \sum_{j \in V, i \neq j} w_{ij} \tag{6.1}$$

w_{ij} is an element of the adjacency matrix of G such as $w_{ij} = \alpha$ if i and j are connected, otherwise, $w_{ij} = 0$, $\alpha \in \mathbb{R}_+^*$.

2.2.2 Weighted Betweenness Centrality

The Betweenness [Brandes, 2001] $b(i)$ of a node i is the fraction of the shortest path passing through it. When it is normalized, the Betweenness of the node i is defined as follows :

$$b(i) = \frac{2}{(n-1)(n-2)} \sum_{i \neq j} \frac{\sigma^w_{jk(i)}}{\sigma^w_{jk}} \tag{6.2}$$

σ^w_{jk} is the number of the shortest path between j and k. $\sigma_{jk}(i)^w$ is the number of the shortest path from j to k passing in i.

2.2.3 Weighted PageRank centrality

Initially introduced to rank websites, PageRank [Ding, 2011] is an iterative refinement measure based on a random walk process. Although defined for directed networks, one can use two directed links for undirected networks. For weighted network, it is defined as follows :

$$wpr_i(t) = \sum_{j=1}^{n} w_{ji} \frac{wpr_j(t-1)}{s_j^{out}} \tag{6.3}$$

s_j^{out} is the out-strength of node j.

2.3 Evaluation Measures

2.3.1 Largest Connected Component

The Largest Connected Component (LCC) is the most popular metric to assess the robustness of a network [Oehlers, 2021 ; Wandelt, 2018b]. It refers to the largest connected component after removing a node and its links. Indeed, this operation can dislocate the network into several independent components. It denotes the cohesiveness of the network and its ability to diffuse information. The higher the Largest Connected Component, the most resilient the network is to the attack.

2.3.2 Weighted Efficiency

The efficiency [Crucitti, 2003] evaluates the exchange of information between the nodes of a network. Zhou et al. introduce the weighted efficiency in [Zhou, 2019] to consider the influence of the link weights. In contrast to the unweighted efficiency, the weighted efficiency can be greater than 1. It is defined as follows :

$$WE(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in V} \frac{1}{\sum_{l \in L_{ij}} \frac{1}{w_l}} \tag{6.4}$$

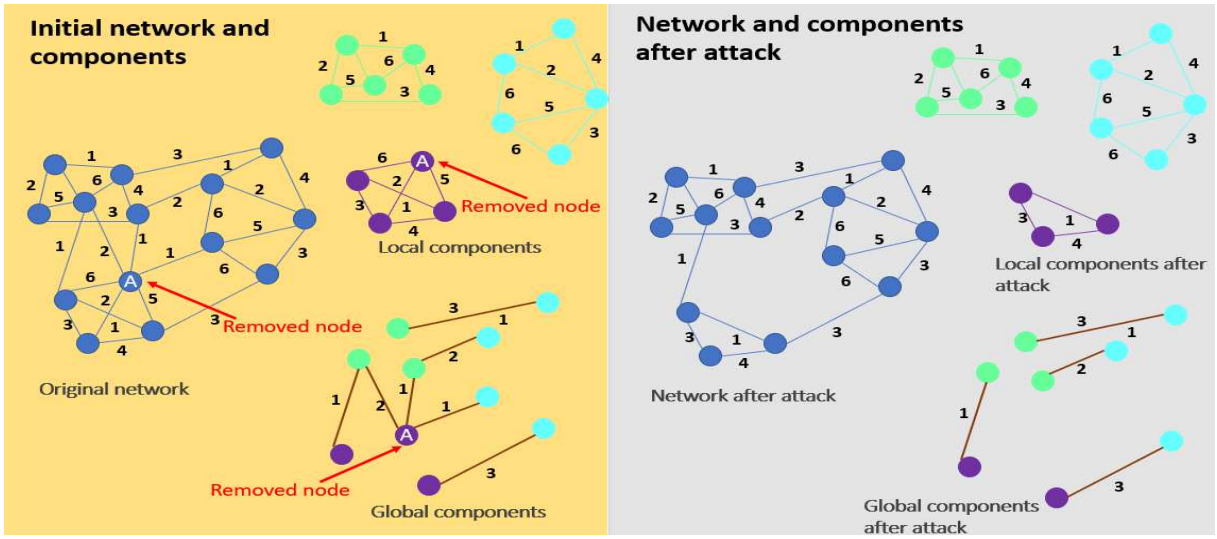


FIGURE 6.2 – This toy example illustrates the influence of an attack on the network component structure. Removing node A in the original network impacts the dark blue local component (One node and three links are removed). It also affects the largest global component, which splits into two components with two nodes and an isolated node. Indeed, A has a global and local role.

L_{ij} is the set of links on the shortest path between i and j. w_l is the weight of link l.

To assess the robustness of a network, this measure is recalculated after a node is removed from the network.

2.4 Jaccard Index

A Jaccard Index [Costa, 2021] is used to compare the similarity of two sets. It is defined as follows :

$$J(P,Q) = \frac{|P \cap Q|}{|P \cup Q|} \tag{6.5}$$

In the case of identical sets, the Jaccard index equals 1. In the absence of a shared element between two sets, it equals zero.

3 Data and Methods

3.1 Data

We use a weighted and undirected network. It represents the flight information collected during six days from FlightAware[Flightaware,] (between May 17, 2018, and May 22, 2018). Nodes represent airports, and link weights represent the number of direct flights between airports. Table 7.1 presents its basic topological properties. One can see that the global air network is sparse. It is disassortative. In other words, large degree nodes tend to connect with low degree nodes. It has a large diameter. Indeed, one needs 12 flights to reach the most distant airports. The transitivity indicates that there are not many triplets. Consequently, rerouting is uneasy.

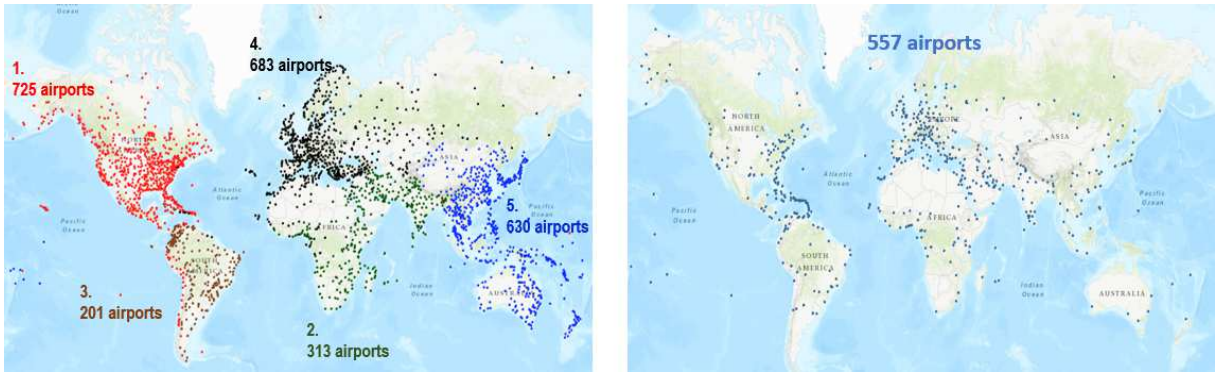


FIGURE 6.3 – The figure at the left represents the airports in the uncovered large local components. They are in well-delimited geographic areas. Each color represents a component. 1) North America-Caribbean, 2) Europe-Russia, 3) East and Southeast Asia-Oceania, 4) Africa-Middle East-Southern Asia, 5) South America. The figure at the right represents the airports of the largest global component. In contrast to local components, they are scattered all over the world.

TABLE 6.1 – Basic topological properties of the world air transportation network, the five large local components, and the largest global component. N is the network size. $|E|$ is the number of edges. δ is the diameter. l is the average shortest path length. μ is the density. ζ is the transitivity, also called the global clustering coefficient. $k_{nn}(k)$ is the assortativity, also called the Degree correlation coefficient.

Components	N	$\|E\|$	δ	l	μ	ζ	$k_{nn}(k)$
World air transportation network	2734	16665	12	3,86	0,004	0,26	-0,05
North America-Caribbean	725	3794	7	3.14	0.014	0.28	-0.325
Europe-Russia	683	6016	6	2.85	0.025	0.32	-0.2
East and Southeast Asia-Oceania	630	3012	9	3.35	0.015	0.34	-0.22
Africa-Middle East-Southern Asia	313	989	7	3.37	0.02	0.28	-0.15
South America	201	494	6	3.2	0.024	0.23	-0.35
Large global component	557	2108	8	3	0.013	0.13	-0.25

3.2 Methods

We extract the component structure of the world transportation network to evaluate the impact of a targeted attack on its regional and inter-regional constituents. The robustness evaluation process proceeds as follows :

1. Disconnect a node from the world air transportation network according to an attack strategy.
2. Disconnect the same node from its local component.
3. Disconnect the same node from the global component if possible.
4. Extract the Largest Connected Component from the world air transportation network.
5. Extract the Largest Connected Component of the local components.
6. Extract the Largest Connected Component of the global component.

This approach allows us to evaluate the impact of removing a critical airport in the world air transportation network on the regional and inter-regional networks. Fig. 6.2 presents a toy example illustrating an attack on an airport of the world network and the effect on its components.

4 Component structure

We use the weighted Louvain community detection algorithm [Blondel, 2008 ; Jin, 2021a] to extract the dense parts of the network. It uncovers 17 communities. Therefore, there are 17 local components. Among them, there are five large components covering the following regions : 1) North America-Caribbean, 2) Europe-Russia, 3) East and Southeast Asia-Oceania, 4) Africa-Middle East-Southern Asia, 5) South America. There are eight global components. The largest one includes more than 95% of the airports distributed worldwide [Diop, 2021]. Fig. 6.3 represents the airports included in the large local and global components. We restrict our attention to these components in the robustness analysis.

Table 7.1 reports their basic topological properties. The diameter of the components quantifies the maximum number of links between two destinations in the component. Its value is around six and seven, except for the East and Southeast Asia-Oceania component, which requires a maximum of nine jumps to join two destinations. The global component's diameter is eight. It suggests that interregional relations are not very different than regional ones. The average path length confirms these results. On average, the Europe-Russia component requires less than three hops to travel between two airports. It takes more than three hops on average for all the other local components. With an average value of 3 for the large local component, inter-regional travels appear well organized compared to regional trips. Concerning the density of the components, one can distinguish two categories. The first class regroups the densest components : Europe-Russia, Africa-Middle East-Southern Asia, and South America. Indeed, their density is around 0.2. The second category contains North America-Caribbean, East and Southeast Asia- Oceania, and the large global component with a much lower density (about 0.13). With a transitivity value of 0.3, Europe-Russia and East and Southeast Asia components contain more interconnected triplets than the other components. Therefore, rerouting is more straightforward in these components. Note that the global component is less transitive. Assorsativity measures the ability to connect nodes sharing similar degrees. All the components present various levels of disassortativity. In other words, airports with a high degree tend to connect with low-degree ones. The North America-Caribbean and the South America components are the most disassortative. This behavior is in line with the hub-and-spoke organizational model used in the airline industry. The Africa-Middle East-Southern Asia component is the less disassortative.

5 Strength targeted attack

5.1 Evolution of the Largest Connected Component

Fig. 6.4.A gives the LCC as a function of the fraction of top Strength nodes removed from the world air transportation network. It also reports the corresponding LCC values of the large components. All the curves reflect a similar behavior. As the fraction of top Strength nodes increases, the size of the various LCC decrease almost linearly. The LCC of the Europe-Russia component decreases very slowly. Consequently, it is the most resilient region. The East-Southeast Asia-Oceania region follows. The Africa-Middle East-Southern Asia component is slightly below. One can distinguish two cases for the global component. Its size varies in high proportions below a threshold of around 8% of removed nodes. In contrast, above this threshold, it resists better to attacks and tends to be as resilient as the Europe-Russia component. The North America-Caribbean component is the most sensitive to the attack. The South America component behaves slightly better. Note that the LCC of the East-Southeast Asia-Oceania and North America-Caribbean components break down to around 20% when around 9% of their top airports are removed.

One can observe that the size of the LLC of South America and the Africa-Middle East-Southern Asia components decreases similarly. Isolating a single node produces an abrupt reduction of the LCC.

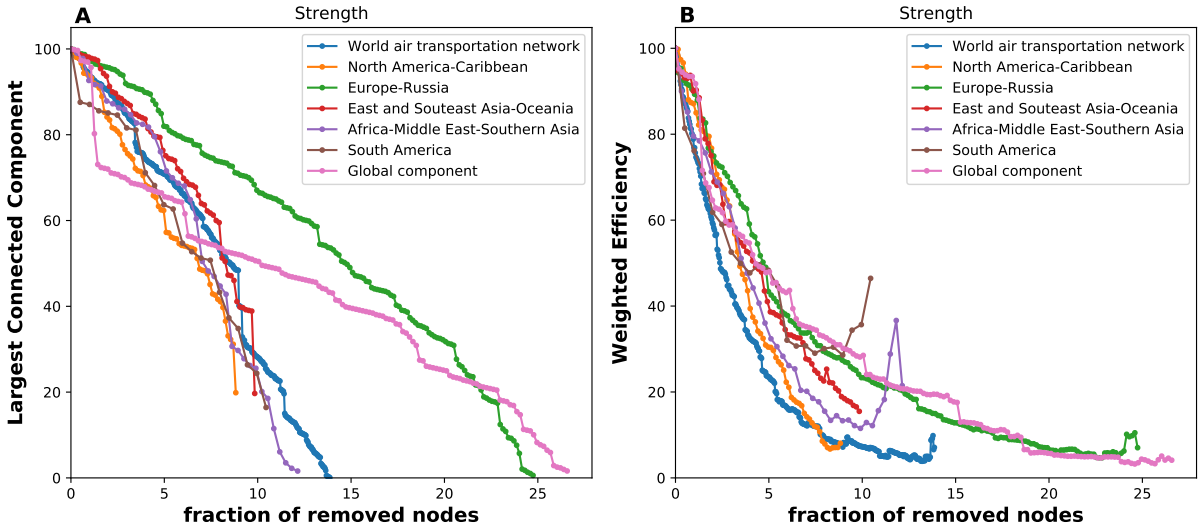


FIGURE 6.4 – The figure at the left represents the fraction of nodes in the LCC of the large local and global components as a function of the fraction of removed nodes under a Strength attack on the world air transportation network. We also represent the LCC of the world air transportation network in this situation. The Europe-Russia component and the global component are less sensitive to the attack. The figure at the right shows the weighted efficiency of the large local and global components as a function of the fraction of removed nodes under a Strength attack on the world air transportation network. We also represent the weighted efficiency of the world air transportation network in this situation. Efficiency decreases at a high rate as the fraction of top Strength nodes removed increases. Furthermore, The North America-Caribbean and Africa-Middle East-Southern Asia components appear to be the most vulnerable.

TABLE 6.2 – Basic topological properties of the five local components isolated by the Strength attack. λ is the size of the isolated component. ρ is the fraction of top Strength nodes removed from the world air transportation network to isolate the component. δ is the diameter. l is the average shortest path length. μ is the density.

Components	$\rho(\%)$	$\lambda(\%)$	δ	l	μ
North America-Caribbean	9	19.86	14	5.7	0.021
Europe-Russia	13.5	21.5	18	7.96	0.019
East and Southeast Asia-Oceania	9.2	19.6	16	4.84	0.04
Africa-Middle East-Southern Asia	13.5	16.61	4	2	0,4
South America	11.4	16.4	7	3.69	0.063

Indeed, removing a node separates an entire subnetwork from the component in this situation. For the other components and the world air transportation network, the variations of the LCC are more regular. The Europe-Russia component is a good illustration of this behavior. Indeed, it is almost a straight line. Nevertheless, removing a single node can also cause considerable damage to these networks. The world air transportation network is a typical example of brutal variations followed by quasi-linear behavior.

As the fraction of removed nodes increases, the components break away from the world air transportation network one after the other. Table 7.2 displays information about the local components. We specify the fraction of nodes needed to isolate it from the world air transportation network. It also reports their giant component's size and basic topological properties after isolation.

North America-Caribbean is the first region isolated after removing 9% of the top Strength airports

of the overall network. Indeed, this component contains several airports with high traffic in the world. That is why it is heavily targeted. The top 9% of airports include 9.1% of the airports of this component. Removing Piarco Airport in Trinidad and Tobago provoked the disconnection. The remaining airports form a network that contains 19.86% of the airport from the initial component. These airports are in the United States and Canada. The Alaska subregion is also isolated from the LCC. One can see that the diameter and the shortest path values double. Therefore, even though the LCC is denser than the initial component, it is more challenging to travel in the LCC.

East-Southeast-Asia-Oceania is the next separated component after disconnecting 9.2% of the top Strength airports from the world air transportation. A proportion of 10% of the top airports belongs to this component. Beijing Nanyuan Airport in China was the last airport removed before isolation. The LCC includes 19.6% of the airports of the initial component, mainly in the Oceania area. Note that this component is cut in two. Indeed, the Beijing Nanyuan Airport is the last interaction between East-Southeast-Asia and Oceania. Traveling within this LCC is more complicated than the initial component. Indeed, although denser than the original component, its diameter almost doubled, and the average shortest path increases by one hop.

Eliminating 11.4% of airports with the highest Strength of the world air transportation network isolates South America. There are 11.4% of the airports belonging to this component among these airports. The separation occurs when removing Ministro Pistarini Airport in Argentina. Its LCC contains 16.4% of the airports of the initial component situated in Brazil and Venezuela. Traveling in the LCC is almost as easy compared to the initial component. Indeed, the diameter increases only by one. On average, it requires the same number of hops. In addition, the LCC is denser than the initial component.

Africa-Middle East-Southern Asia is the next region detaching from the world air transportation network. It happens after removing 13.5% of the top airports in the world. Among them, 17.8% of the airports belong to this component. India's Sardar Vallabhbhai Patel Airport is the last airport connecting it to the world. Once separated, The LCC of this component includes five airports distributed across five countries (Nepal, Bhutan, United Arab Emirates, India, and Bangladesh). Except for the United Arab Emirates, these countries are neighbors.

Once these four components are separated, European-Russia remains the only component in the world air transportation network. It contains 21.5% of the airports of the initial component. They are in various countries. However, countries like Portugal, Norway, Sweden, Turkey, and Spain are almost unreachable. In addition, except for Tajikistan, Central Asia and Transcaucasia countries are disconnected. The airports of Russia and Europe are still interconnected. Nevertheless, it isn't easy to travel within this LCC. Indeed, the density is low, and the diameter and average shortest path are three times higher.

5.2 Weighted Efficiency

Fig. 6.4.B illustrates the evolution of the normalized weighted efficiency for the components and the world air transportation network under targeted strength attack. For the components, the curves report their values until they are isolated. Indeed, after isolation, their weighted efficiency does not change anymore. Indeed, removing a node in the world air transportation network does not affect the isolated component topology. One can observe a similar behavior on all the curves. Efficiency decreases at a high rate as the fraction of top Strength nodes removed increases. One can notice a slightly different behavior for the Africa-Middle East-Southern Asia and the South America components. Before their isolation, the efficiency of these components increases. It is because airports located in a few countries compose their LCC. Overall, the European-Russia and the global components are less sensitive to Strength targeted attacks in terms of efficiency. Conversely, the North America-Caribbean and Africa-Middle East-Southern Asia components appear most vulnerable. Let's now compare the component efficiencies' after isolation

TABLE 6.3 – Evolution of the weighted efficiency, the LCC of the large local components after their isolation due to attacks based on Strength. The order of isolation of the components is also presented. ω_i is the Weighted Efficiency before attacks based on Strength. ω is the weighted efficiency after the separation. λ is the Largest Connected Component size after the separation. ρ is the fraction of top Strength nodes removed from the world air transportation network to isolate the component. γ is the order in which the components are isolated. After their isolation, the components lose much of their efficiency.

Components	ω_i	ω	$\frac{\omega}{\omega_i}(\%)$	$\lambda(\%)$	$\rho(\%)$	γ
North America-Caribbean	57.42	4	7	19.86	9	1
Europe-Russia	26.4	1.52	5.75	17.8	13.3	4
East and Southeast Asia-Oceania	31.31	4.85	15.4	19.6	9.6	2
Africa-Middle East-India	18.5	3.97	21.45	16.61	13.3	4
South America	23	10.7	46.5	16.4	11.4	3

from the world transportation networks.

Table 6.3 displays the initial value and the value after isolation of the Weighted Efficiency for the local components. One can see that the initial values are very different. Initially, the North America-Caribbean component is by far the most efficient. The East and Southeast Asia, Europe-Russia, and South America components follow. Finally, Africa-Middle East-Southern Asia is the less efficient local component.

After isolation, the South American local component is the most efficient. Indeed, few of its airports rank high in the world air transportation network. In addition, except for four airports in Venezuela, it includes only Brazilian airports. The Africa-Middle East-Southern Asia is the following component with the most significant efficiency. It contains only six airports located in different countries after isolation. East and Southern Asia-Oceania is next. It contains only airports from Oceania, located in several countries. Here again, an attack on the global air network has few consequences on the regional traffic in Oceania. North America-Caribbean and the Europe-Russia components are the less efficient. Indeed, they include the top-rank hubs in the world air transportation network. Consequently, they lose numerous hubs. Note that the North America-Caribbean component is more efficient than Europe-Russia. Indeed, it includes several airports, but they are in only two countries. In contrast, airports are located in several European countries and Russia in the European-Russia component. One can notice that there seems to be no relationship between the order of isolation of the components and their relative efficiency after an attack.

6 Weighted Betweenness targeted attack

6.1 Evolution of the Largest Connected Component

Fig. 6.5.A reports the variation of the LLC while removing top Betweenness centrality airports from the world air transportation network. One also plots the corresponding evolutions of the LCC for the components. The curves overlap until the fraction of top nodes removed reaches around 2%. Beyond this value, they diverge. One can see that South America is the less resilient component, followed by the East and Southeast Asia-Oceania, the North America-Caribbean, and the Africa-Middle East-Southern Asia components. Europe-Russia and the global components exhibit similar behavior. Above 12% of removed airports, Europe-Russia takes the lead and becomes more resilient. One can see that the size of the LCC of North America-Caribbean, Europe-Russia, and the East-Southeast Asia-Oceania components decrease slowly. Piecewise linear variation with sharp drops characterizes the others, particularly the World air

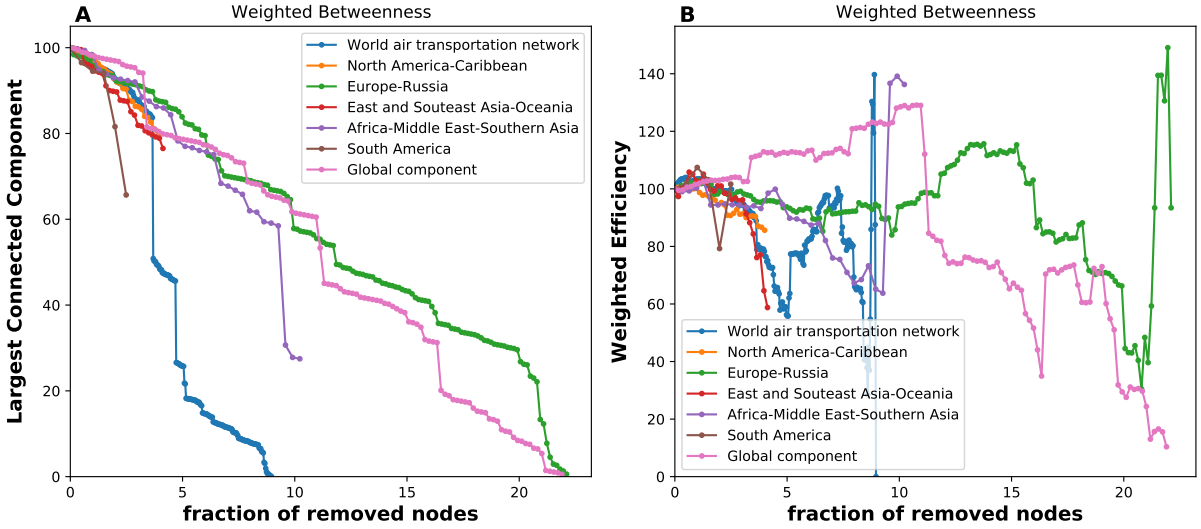


FIGURE 6.5 – The figure at the left represents the fraction of nodes in the LCC of the large local and global components as a function of the fraction of removed nodes under a Weighted Betweenness attack on the world air transportation network. We also represent the LCC of the world air transportation network in this situation. The Europe-Russia component and the global component are less sensitive to the attack. The figure at the right shows the weighted efficiency of the large local and global components as a function of the fraction of removed nodes under a Strength attack on the world air transportation network. We also represent the weighted efficiency of the world air transportation network in this situation. The weighted efficiency of the world air transportation network and the large components are very sensitive to attacks based on the betweenness centrality.

TABLE 6.4 – Basic topological properties of the five local components isolated by the Betweenness attack. λ is the Largest Connected Component size. ρ is the fraction of top Strength nodes removed from the world air transportation network to isolate the component. δ is the diameter. l is the average shortest path length. μ is the density.

Components	$\rho(\%)$	$\lambda(\%)$	δ	l	μ
North America-Caribbean	4.1	80	12	3.75	0.01
Europe-Russia	5.1	70	7	3.1	0.028
East and Southeast Asia-Oceania	4.7	76.5	17	4.77	0.012
Africa-Middle East-Southern Asia	5.1	27.4	6	3.28	0.045
South America	3.4	66.6	10	3.32	0.036

transportation network. Indeed, removing an airport can provoke significant damage. Table 7.3 reports the topological properties of the components once they are isolated from the network.

After removing 3.4% of the top Betweenness airports from the world air transportation network, South America is the first isolated component. The attack contains only five airports (2.5%) of this component. They are in Venezuela, Brazil, Cuba, Colombia, and Argentina. Ministro Pistarini Airport in Argentina is the last liaison to the world. The LCC includes 65.6% of the airports of the initial component, distributed between all countries. The diameter and the density of the LCC are higher than in the initial component. In contrast, the average shortest path is comparable. Therefore, traveling within the LCC is almost as easy as traveling in the original component (Table7.3).

The North America-Caribbean component is the second region that becomes unreachable after removing 3.7% of the top central airports of the world air transportation network. Among the 3.7% of top world airports, 4.1% belong to this component. The last airport connected to the other region is the Orlando Airport in the United States. The LCC contains 80% of the airport of the initial component. These airports are scattered across all the countries. Even though the diameter is almost twice higher, it is always easy to travel between the remaining airports. Indeed, the average shortest path and the density increase slightly.

Removing 4.7% of the top central airports of the world air network isolates the East and Southeast Asia region. The attack targets 4.1% of the airports of this component. Tokyo Haneda Airport in Japan is the last link to the world before the isolation. The size of the LCC decreases to 76.5% of the original component. Its airports are scattered across all the countries of the component. Nevertheless, air travel is more uneasy. Indeed, the LCC is less dense and its diameter and average shortest path increase compared to the initial component.

Africa-Middle East-Southern Asia is the next isolated area. One needs to remove 5.1% of the top Betweenness airport of the world air transportation network. There are 10.2% of the airports of this component among these top 5.1% airports. Disconnecting the Borg El Arab Airport in Egypt isolates this region. 27.4% of the airports of the original component remain on the LCC. These airports are located only in the Middle East, Southern Asia, and Seychelles. Nevertheless, the United Arab Emirates, Yemen, and Oman are unreachable. Air travel in the LCC is more straightforward than in the initial component. Indeed, the diameter, the average shortest path, and the density are smaller.

Finally, only the Europe-Russia component remains on the world air transportation network. It includes 70% of airports of the initial component. All the countries of this component are well represented. Joining two airports of the LCC is slightly more complicated. Indeed, the diameter and the average shortest path increase. Note that the density also increases.

TABLE 6.5 – Evolution of the weighted efficiency, the LCC of the large local components after their isolation due to attacks based on Weighted Betweenness. We also report the order of isolation of the components. ω_i is the Weighted Efficiency before attack based on Weighted Betweenness. ω is the weighted efficiency after the separation. λ is the Largest Connected Component size after the split. ρ is the fraction of top Weighted Betweenness nodes removed from the world air transportation network to isolate the component. $\gamma$is the order in which the components are isolated. Globally, after their isolation, the efficiency of the LCC does not change much. Some LCCs are even more efficient.

Components	ω_i	ω	$\frac{\omega}{\omega_i}(\%)$	$\lambda(\%)$	$\rho(\%)$	γ
North America-Caribbean	57.42	49.17	85.63	80	4.1	2
Europe-Russia	26.4	24.13	91.4	70	5.1	4
East and Southeast Asia-Oceania	31.31	18.4	58.76	76.5	4.7	3
Africa-Middle East-India	18.5	25.19	136.16	27.4	5.1	4
South America	23	23.37	101.6	66.6	3.4	1

6.2 Weighted Efficiency

The weighted efficiency of the world air transportation network and the large components are very sensitive to the attacks based on the betweenness centrality, as shown in Figure 6.5.B. One can distinguish three typical behavior for the efficiency variation versus the fraction of removed nodes. The first category includes the North America-Caribbean and East and Southeast Asia components. Their relative efficiency decreases almost monotonically as the fraction of removed nodes increases. This behavior is similar to the Strength attack. Nevertheless, the efficiency of these components after the Weighted Betweenness attack is less impacted. The second category concerns the Africa-Middle East-Southern Asia and the South America components. In this case, the efficiency decreases as in the first category. However, before the component splits from the network, its efficiency increases. The third category regroups the European-Russia component, the global component, and the world air transportation network. Large variations of efficiency characterize it. Indeed, the Weighted Betweenness attack isolates the peripheral sub-regions successively so that the weighted efficiency can surpass its initial value.

Table 6.5 displays the initial value of the Weighted Efficiency for the local components and the value after their isolation. After isolation, the Africa-Middle East-Southern Asia component is much more efficient. Indeed, while it initially covers many countries, the isolated component exclusively contains airports in the Middle East and Southern Asia (except for two airports in Seychelles). As These two regions are very well connected, the component efficiency increases even though it loses numerous hubs.

The South America component efficiency increases slightly compared to its initial value. Indeed, most of the removed airports by the attack are in Brazil. Overall, other countries still have essential airports to maintain travel efficiency. The North America-Caribbean component is the second isolated component. It looses 15% of its efficiency while preserving 80% of its airports. The United States and Canada airports are the most vulnerable to the attack. Nevertheless, some essential airports in different countries remain in this component. Travel becomes more challenging but well guaranteed.

The European-Russia component is the next most efficient. Despite being heavily attacked, this component loses less than 10% of its efficiency while keeping 70% of its initial nodes. As the isolated part preserves several hubs in Europe, traffic within European airports is still efficient. However, reaching and traveling in the Russian area is more complicated. Indeed, air transport in Russia depends on a few hubs. Once attacked, traveling in and to the region is more challenging.

The East-Southeast-Asia - Oceania component is the more impacted by the attack. It loses more than 40% of its efficiency while keeping more than 70% of its airports. Indeed, a strong Hub and Spoke

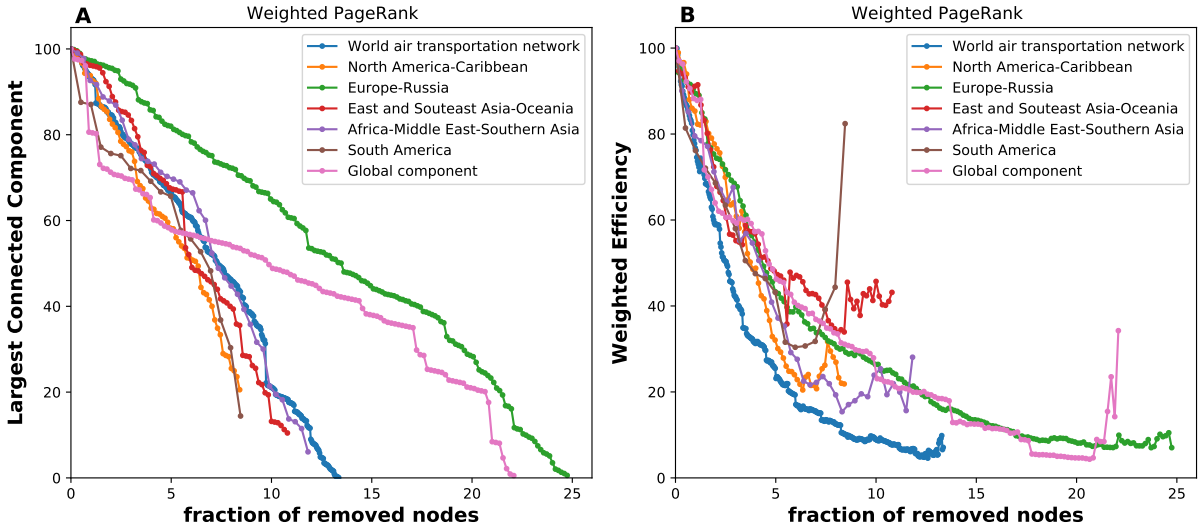


FIGURE 6.6 – The figure at the left represents the fraction of nodes in the LCC of the large local and global components as a function of the fraction of removed nodes under a Weighted PageRank attack on the world air transportation network. We also represent the LCC of the world air transportation network in this situation. The Europe-Russia component and the global component are less sensitive to the attack. The figure at the right shows the weighted efficiency of the large local and global components as a function of the fraction of removed nodes under a Strength attack on the world air transportation network. We also represent the weighted efficiency of the world air transportation network in this situation. The world air transportation network appears as the most vulnerable.

TABLE 6.6 – Basic topological properties of the local components isolated by the Weighted PageRank attack. λ is the Largest Connected Component size. ρ is the fraction of top Strength nodes removed from the world air transportation network to isolate the component. δ is the diameter. l is the average shortest path length. μ is the density.

Components	$\rho(\%)$	$\lambda(\%)$	δ	l	μ
North America-Caribbean	8.1	20.4	17	6.11	0.021
Europe-Russia	13.2	16.7	13	6.10	0.022
East and Southeast Asia-Oceania	12	10.47	8	3.82	0.045
Africa-Middle East-Southern Asia	13.2	6	8	4	0,13
South America	9.7	14.9	6	3	0.09

topology characterize this region. The attack removes important Hubs in China and Oceania involved in the traffic between these two regions. Consequently, the global efficiency of the component decreases drastically.

7 Weighted PageRank Targeted Attack

7.1 Evolution of the Largest connected component

Fig. 6.6.A presents the variation of the LCC for the five local components and the world air transportation network when removing the top Weighted PageRank nodes in the world air transportation network. One can see that the European-Russia component is the most resilient. Indeed, its curve is

above all the others, whatever the fraction of removed nodes. The other local components exhibit similar behavior. Their sizes decrease almost linearly in equal proportions. Nevertheless, one can notice that globally, the North America-Caribbean and the South America components are the most sensitive to the Weighted PageRank attacks. For the global component, we can distinguish two situations. It behaves like the mainstream local components until it reaches a fraction of 7% removal of the top weighted PageRank airports. Above this value, it tends to follow the European-Russia component.

Removing 8.11% of the top-ranked airports in the world air transportation network Isolates the North America-Caribbean component. The attack concerns only 8.4% of the airports in this component. Southwest Florida Airport in the United States was the last airport allowing to reach the rest of the world. Its LCC includes 20.4% of the airports of the initial component. They cover the initial component area, except for Alaska, Chile, and a large part of Canada. Comparing its topological properties with the initial component, one can say that it is more difficult to travel in the isolated component. Indeed, as shown in Table6.6, even though the density increases, the diameter, and the average shortest path values are two times higher.

South America separates next from the world air network after removing 9.7% of airports. Among these airports, only 8.4% belong to this component. Simon Bolivar Airport in Venezuela is the last airport connected to the world. Once isolated, 14.9% of the airports in the initial component remain in this LCC. Except for four airports in Venezuela, they are all in Brazil. Note that the topological properties of the LCC are similar to the initial component. Therefore, the regional traffic is not affected much.

Removing 12% of the airports from the global air network isolates the East and Southeast Asia-Oceania component. At this point, the attack concerns only 10.8% of the airports in this component. Lhasa Gonggar Airport in China was the last link to the rest of the world. The LCC reduces to 10.47% of the initial component size, with its airports scattered in nine countries. However, they are not well distributed. Indeed, there are no airports in Oceania, and half of them are in China airports. Although the component shrinks a lot, covering a more concentrated area makes travel more straightforward in what remains. Indeed, the airports of the LCC are well connected compared to the initial component.

The Africa-Middle East-Southern Asia component disconnects from the whole network when the attack involves 13.2% of the top airports. Among these airports, only 12.21% belong to this component. Khartoum Airport in Sudan is the last link to the world. Only 6% of the airports remain connected in the LCC. They are in East Africa and islands such as Madagascar, La Reunion, etc. However, the LCC covers a more compact and denser area than the initial component. In addition, the diameter increases by one hop, while the average shortest path does not change. Therefore the local travel experience in the LCC is comparable to the initial component.

Once all the components are isolated from the world air network, only the Europe-Russia component remains. Moreover, the LCC connects 16.7% of the initial airports in this component. It includes airports from various countries except for Portugal, Finland, Kazakhstan, and Turkey, which have become unreachable. The diameter and the average shortest path length values are two times higher than in the initial component. Additionally, the density decreases. Consequently, traveling in what remains of the component is much more difficult.

7.2 Weighted Efficiency

Fig. 6.6.B displays the variation of the Weighted Efficiency of the components and the world air transportation network when the top Weighted PageRank airports are removed iteratively from the world air transportation network. The world air transportation network appears as the most vulnerable. Results for the components are more mixed. Until the attack involves 5% of the airports, curves are very close. After this threshold, Europe-Russia and the global component are the most efficient. Moreover, we

TABLE 6.7 – Evolution of the weighted efficiency, the LCC of the large local components after their isolation due to attacks based on Weighted PageRank. We also report the order of isolation of the components. ω_i is the Weighted Efficiency before attack based on PageRank. ω is the weighted efficiency after the separation. λ is the Largest Connected Component size after the split. ρ is the fraction of top PageRank nodes removed from the world air transportation network to isolate the component. γ is the order in which the components are isolated. Isolating the components reduces their efficiency significantly.

Components	ω_i	ω	$\frac{\omega}{\omega_i}(\%)$	$\lambda(\%)$	$\rho(\%)$	γ
North America-Caribbean	57.42	12.55	21.85	20.4	8.4	1
Europe-Russia	26.4	1.86	7	17.8	13.4	4
East and Southeast Asia-Oceania	31.31	13.51	43.14	10.47	12	3
Africa-Middle East-India	18.5	5.19	28	6	13.4	4
South America	23	18.95	82.4	14.4	9.7	2

distinguish two types of curves.

In the first type, we observe that the Weighted Efficiency increases before the isolation. It is the case of the East and Southeast Asia-Oceania, Africa-Middle East-Southern Asia, South America, and the global components. Indeed, the attack splits the component into multiple sub-regions. Therefore, the airports in the LCC cover a small area facilitating the regional exchanges well, even though one cannot reach some top PageRank airports. The second type concerns the North-America, Europe-Russia components. Indeed, these components lose several top PageRank airports, making regional travel more challenging.

Table 6.7 presents the weighted efficiency of the components and the world air transportation network before the attack and after the isolation of the LCC. In that case, the attack always translates into a lower efficiency. The South America component is the less vulnerable, with an efficiency decreasing by less than 20%. Indeed, the isolated component contains essentially Brazilian airports. Furthermore, it covers mainly the northern parts of Brazil, near Venezuela. This reduced coverage compared to the initial component makes it quite efficient.

The East and Southeast Asia-Oceania component follows. Its efficiency after the attack drops to around 40% of its initial value. There are no airports from Oceania in the LCC. Most of them are in the East of China and its satellite countries. Therefore, relative preservation of the efficiency also comes at the cost of a smaller geographical coverage. The Africa-Middle East-Southern Asia reduces its efficiency to 28% of its initial value. In this case, also the airports are concentrated in a reduced area. Indeed, the isolated component includes only airports in East Africa and Madagascar. The North America-Caribbean component is susceptible to the attack. Once isolated, its efficiency reduces around 20% of its initial value. However, the isolated component by PageRank covers almost all the original regions. Europe-Russia is the most affected component by the weighted PageRank attack. It loses several hubs. In addition, many peripheral subregions, including secondary airports, remain in the isolated component. It is why its weighted efficiency reduces to 7% of its initial value.

8 Comparing the attacks

8.1 Evolution of the Largest Connected Component

Table 6.8 displays the order of isolation of the local components for the various attacks. Although it changes with the attack, one can observe some trends. The North America-Caribbean region tends to leave first the world air transportation network. South America, which is very connected to the North America-

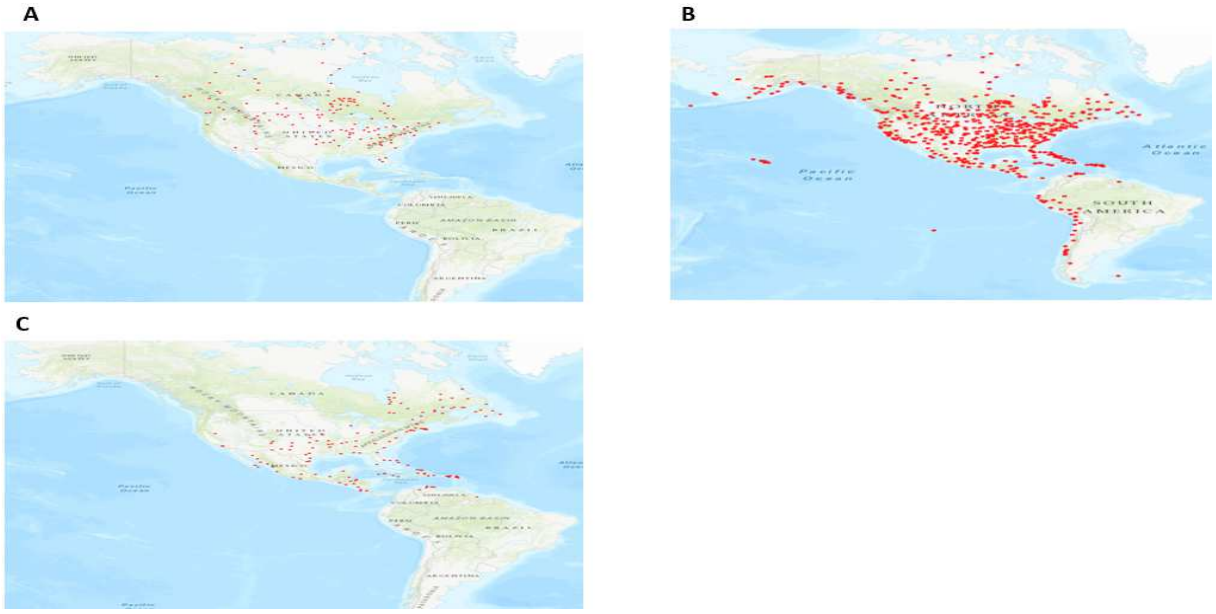


FIGURE 6.7 – LCC of the North America-Caribbean component after its isolation from the network. The red dots are the remaining airports after attacks on the world air transportation network (A. Strength attack, B. Weighted Betweenness attack. C. Weighted PageRank attack). The airports in the LCC isolated by the Strength and Weighted PageRank attacks are concentrated in the US and a small part of Canada. In contrast, airports are scattered in all the areas covered by the initial component after the Weighted Betweenness attack.

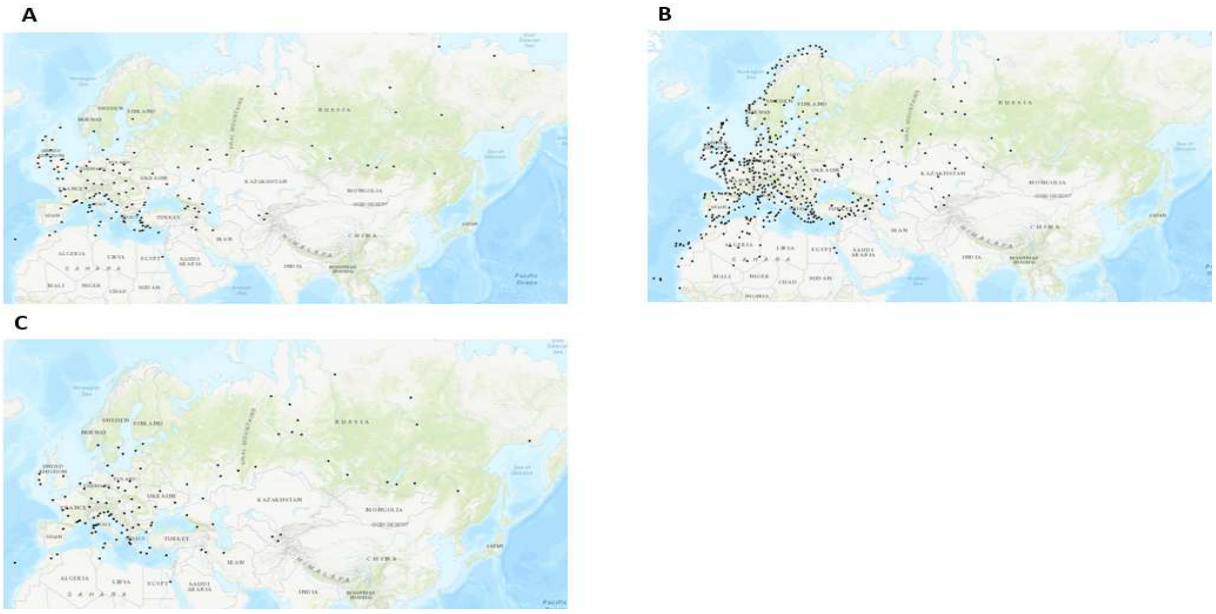


FIGURE 6.8 – LCC of the Europe-Russia component after its isolation from the network. The black dots are the remaining airports after attacks on the world air transportation network (A. Strength attack, B. Weighted Betweenness attack. C. Weighted PageRank attack). The LCC isolated by the Strength and Weighted PageRank attacks covers almost all countries sparsely. The LCC is well distributed in all the initial component areas for the Weighted Betweenness attack.

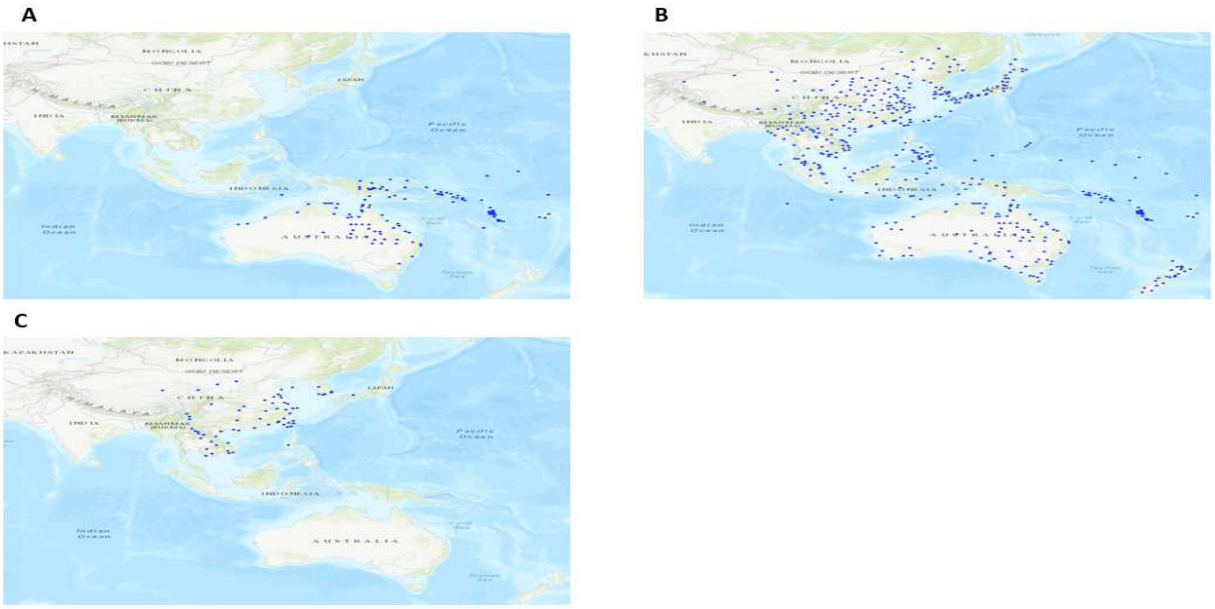


FIGURE 6.9 – LCC of the East and Southeast Asia-Oceania component after isolation from the network. The blue dots are the remaining airports after attacks on the world air transportation network (A. Strength attack, B. Weighted Betweenness attack. C. Weighted PageRank attack). The LCC isolated by the Strength attack covers the countries in Oceania. The LCC separated by Weighted PageRank attack, covers East and Southeast Asia major countries. In contrast, the LCC isolated by the Weighted Betweenness attack covers all the regions of the initial component.

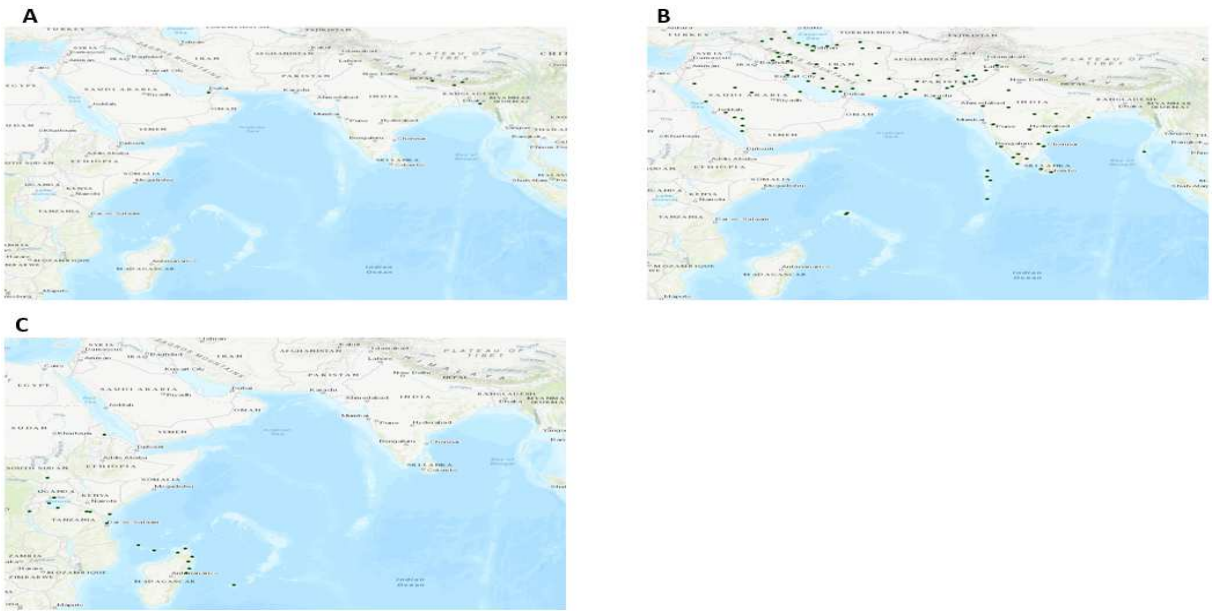


FIGURE 6.10 – LCC of the Africa-Middle East-Southern Asia component after isolation from the network. The green dots are the remaining airports after attacks on the world air transportation network (A. Strength attack, B. Weighted Betweenness attack. C. Weighted PageRank attack). The LCC isolated by the Strength attack covers five countries (Nepal, Bhutan, United Arab Emirates, India, and Bangladesh). The LCC isolated by Weighted PageRank attack covers East Africa and Madagascar countries. The LCC separated by the Weighted Betweenness attack, covers mainly the Middle East and Southern Asia countries.

TABLE 6.8 – This table summarizes the results for the attacks under evaluation (Strength, Weighted Betweenness, and Weighted PageRank). We report the order of isolation, the proportion of airports removed, and the LCC of the local component. The Jaccard Index between the LCCs for various couples of attacks is reported for each large component. γ_s, γ_b, γ_p are respectively the order of isolation of the LCC from Strength, Weighted Betweenness and Weighted PageRank. ρ_s, ρ_b, ρ_p are respectively the proportion of attacked airports from the world air transportation network according to the Strength, Weighted Betweenness and Weighted PageRank. λ_s, λ_b, λ_p are respectively the LCC from Strength, Weighted Betweenness and Weighted PageRank. J_{sb}, J_{sp}, J_{bp} represent respectively the Jaccard index of the LCCs isolated by Strength and Weighted Betweenness, Strength and Weighted PageRank, Weighted Betweenness and Weighted PageRank.

Components	γ_s	γ_b	γ_p	$\rho_s(\%)$	$\rho_b(\%)$	$\rho_p(\%)$	$\lambda_s(\%)$	$\lambda_b(\%)$	$\lambda_p(\%)$	J_{sb}	J_{sp}	J_{bp}
North America-Caribbean	1	2	1	9.1	4.1	8.1	19. 86	80	20.4	0.23	0.017	0.25
Europe-Russia	4	4	4	13.2	5.3	13.4	21.5	70	16.7	0.25	0.69	0.2
East and Southeast Asia-Oceania	2	3	3	9.2	4.7	12	19.6	76.5	10.47	0.2	0	0.13
Africa-Middle East-Southern Asia	4	4	4	13.2	5.3	13.4	16.61	27.4	6	0	0	0
South America	3	1	2	11.4	3.4	9.7	16.4	66.6	14.9	0.25	0.46	0.22

TABLE 6.9 – Variation of the Weighted Efficiency of the different attacks once the component splits from the overall network. ω_s, ω_b, ω_p, are respectively the fraction of the Weighted Efficiency of the LCCs for Strength, Weighted Betweenness, and Weighted PageRank attacks. The Weighted Betweenness attack does not impact much the efficiency of the components compared to the other attacks.

Components	$\frac{\omega_i^s}{\omega}(\%)$	$\frac{\omega_i^b}{\omega}(\%)$	$\frac{\omega_i^p}{\omega}(\%)$
North America-Caribbean	7	85.6	21.85
Europe-Russia	5	91.4	7
East and Southeast Asia-Oceania	15	58.7	43.14
Africa-Middle East-Southern Asia	21.45	113.6	28
South America	46.5	101	82.5

Caribbean component, generally follows it. Indeed, once the North America-Caribbean component leaves the global air network, South America has fewer connections with the rest of the world. Furthermore, these connections are with the worldwide most important airports easily targeted. Then East-Southeast Asia-Oceania component gets isolated. Finally, the attack strategy separates Africa-Middle East-Southern Asia from the Europe-Russia component. Indeed, due to their geographic and political relations, these two components are inextricably linked.

Table 6.8 also reports the fraction of removed airports to isolate the components. The Weighted Betweenness attack is the most effective. Indeed, it needs to remove less than half as many airports compared to its alternatives to separate the components. Strength and PageRank need similar resources to reach the same goal.



FIGURE 6.11 – LCC of the South America component after isolation from the network. The brown dots represent the remaining airports after attacks on the world air transportation network (A. Strength attack, B. Weighted Betweenness attack. C. Weighted PageRank attack). The LCC isolated by the Strength and Weighted PageRank attacks includes Brazil and Venezuela. The LCC separated by the Weighted Betweenness attack, covers all the initial component regions.

Table 6.8 presents the fraction of airports in the LCC of the components after breaking away from the world transportation network. Globally, the Weighted Betweenness attack tends to maintain a large fraction of the airports included in the initial component (around 70%). It is not the case for the Africa-Middle East-Southern Asia component. Indeed, the last region to be isolated is divided into multiple sub-regions so that the LCC contains only 28% of the initial airports. Strength and PageRank seem comparable with a slight advantage to Strength. They preserve respectively around 20% and 15% of the size of the initial components.

The Jaccard index between the couples of attacks reported in Table 6.8 shows that globally the largest connected components have few airports in common. The highest score of 0.69 concerns the Europe-Russia component and the Strength and PageRank attacks. The lowest score is zero. It concerns the three couples of attack for the Africa-Middle-East-Southern Asia component and PageRank and Strength for the East and Southeast Asia component. In all these cases, the isolated components have no airport in common.

Figures 6.7 to 6.11 represent the airports in the LCC after the attack for the five local components. They demonstrate the lower regional impact of Weighted Betweenness on the isolated largest connected components. One can also note that, except for the North America-Caribbean component, the isolated components by the Weighted PageRank are smaller than the Strength attack. Additionally, it shows that the isolated components in the same region do not cover the same areas. Indeed, although of similar size, the LCC of the Weighted PageRank attack covers more than 20 countries. In contrast, the LCC derived from the Strength attack covers only the United States and Canada. It confirms the low value of the Jaccard Index coefficient. Indeed, the two LCCs have only five airports in common, and even airports in the same country are different.

The LCCs derived from the Strength and PageRank attacks of the South America component are comparable in size. They cover Brazil and Venezuela. In Venezuela, they include the same four airports. With a Jaccard Index near 0.5, the difference is in Brazil. Nevertheless, they have 10% of the airports in

common. The LCCs from the Strength and Weighted PageRank attacks of Europe-Russia cover the same geographical area. As reflected by the Jaccard Index value (0.7), they have several airports in common (15.6%). The LCCs of the East and South East Asia-Oceania regions have no airports in common. Indeed, for the Strength attack, the LCC includes only the airports in the Oceania component, whereas the PageRank attack includes airports in East and Southeast Asia. The same observation holds for the Africa-Middle-East-Southern Asia component. Indeed, the LCC of the Strength attack is in Southern Asia while it is in Africa for PageRank. To summarize, PageRank and Strength attacks have a comparable impact on the robustness of the local components. However, qualitatively they are very different. Indeed, the isolated components cover different regions.

The Weighted Betweenness attack isolates the components earlier compared to its alternatives. Consequently, the separated LCCs contain 2 to 3 times more airports covering larger geographical areas. Therefore, its Jaccard Index with the other attacks is small. However, the largest isolated components by the Strength attack share more than 80% of its airports with the Weighted Betweenness LCC in the North America-Caribbean, Europe-Russia, and East and Southeast Asia components. The PageRank LCCs share more than 90% of their airports with the Weighted Betweenness LCCs in these components. The LCC of the Weighted Betweenness includes all the airports in the LCCs of the two other attacks in South America. In contrast, all the LCCs are disjoint in the Africa-Middle East-Southern Asia

8.2 Weighted Efficiency

Table 6.9 displays the variation of the Weighted Efficiency of the local components subject to the different attacks. With an average Weighted Efficiency equal to 90 % of its original value, the Weighted Betweenness attack is less disturbing for regional trips. Note that in the Africa-Middle-East-Southern Asia and South America components, it is more efficient to travel in the separated components than in its initial version. Indeed, they are more compact. The PageRank attack is more disturbing, with an average of 36 % compared to the initial values. Strength is the most disruptive attack. Indeed, the average efficiency fraction drops to 18 %. Going into more details, we observe that the LCC of the South America component keeps a high Weighted Efficiency whatever the attack. Its value ranges from 46.5 % to 82.5 %. As the LCCs include mainly airports in Brazil, and Venezuela, traveling in what remains from the original component is still manageable. We observe similar behavior for the Africa-Middle East-Southern Asia component. Its Weighted Efficiency ranges from 21.45 % to 131.6 %. In any case, the LCCs of this component corresponds to different well-interconnected sub-regions. The East and Southeast Asia-Oceania preserve a relatively good efficiency, notably for the Betweenness and PageRank attacks. The LCC is in Oceania for the Strength attack and the East and Southeast Asia sub-region for the PageRank attack. The North America-Caribbean region exhibits a low efficiency after a Strength attack. Despite its LCC covering only two countries, its efficiency is low because it loses several hubs. It is three times more efficient after a PageRank attack. Indeed, the LCC covers several countries while maintaining more hubs. Thus, the regional traffic is less degraded. Europe-Russia has the lowest efficiency for Strength and PageRank attacks. This component loses several hubs in Europe and Russia despite covering all countries. Thus, traffic between the European and Russian regions becomes very challenging. Overall, one can say that the more reduced area the LCC of the components cover, the more efficient they are. Except when the attack targets many hubs, such as North America-Caribbean and Europe-Russia components.

9 Conclusion

The component structure represents the air transportation network as an aggregate of interconnected regional networks. It allows investigating the impact of targeted attacks on the air transport network on

its regional and inter-regional constituents. This study investigates the Strength, Weighted Betweenness, and Weighted PageRank attacks on networks weighted by the number of flights.

In the three attack strategies, the different areas break away successively when removing a given fraction of critical airports from the world air transportation network. Nevertheless, the order of isolation of the regions differs. After the Strength attack, North America-Caribbean, East- Southeast Asia-Oceania, and South America leave the world air transportation network. South America, the North America-Caribbean, and the East and Southeast Asia-Oceania regions disconnect in this order after removing top Weighted Betweenness airports from the world air network. Attacking the top Weighted PageRank airports isolates in sequence North America-Caribbean, South America, East, and Southeast Asia-Oceania. Finally, in any case, Africa-Middle East-Southern Asia and Europe-Russia are the last to split.

The weighted world air transportation network is more vulnerable to Weighted Betweenness attacks than its alternatives under test. Indeed, one needs to remove fewer vital airports in the Weighted Betweenness attack to isolate the different regions. Nevertheless, in this case, the LCCs of the isolated components contain a high fraction of the initial component structure. It allows them to keep and even improve their regional efficiency. The LCC of the components obtained with a budget of removed airports in the same order of magnitude using the Strength and PageRank attacks are much smaller. At first glance, they may seem comparable with a slight advantage for Strength at preserving a higher fraction of the original component airports. However, their content is quite different. Most often, they cover different sub-regions.

The impact of the different attacks in terms of Weighted Efficiency in what remains of the five initial components is not uniform. Globally, Weighted Betweenness is the less disruptive, followed by PageRank and Strength. The proportion of hubs involved in the attack, the order of separation of the components, and the geographical compactness of the LCC exert a significant influence on the local efficiency.

The component structure allows a better understanding of the resilience of the global air transportation network. The weighted Betweenness attack rapidly dismantles the world air transportation network by isolating large geographic areas. However, this vulnerability is relative since travel in these areas remains seamless.

In the next chapter, we will focus on developing attack strategies tailored to the component structure.

Chapter 7

Targeted Attack of the Air Transportation Network Global Component [1]

Summary

Targeted attacks aim to disintegrate a network by removing nodes or links. Classically, in node attacks, one removes nodes in descending order of a given centrality measure. Although most real-world networks exhibit a mesoscopic structure, strategies proposed in the literature do not exploit this ubiquitous property. We introduce an attack strategy based on the network component structure to overcome these drawbacks. The component structure of a network consists of local components and global components. The local components are the dense parts of the network, and the global components are the nodes and links connecting them. We investigate this attack on the world air transportation network using Degree and Betweenness centrality measures. Results show that the degree-based attack on the global component is more effective than the classical attack on the entire network. In contrast, the classical Betweenness attack slightly outperforms the Betweenness attack on the global component. However, the latter is more efficient.

1. The results contained in this chapter are published in the paper :
[Diop, 2022b] Issa Moussa Diop, Chantal Cherifi, Cherif Diallo, Hocine Cherifi. "Targeted Attack of the Air Transportation Network Global Component". The 11th International Conference on Computational Data and Social Networks. CSoNet 2022, December 5-7, 2022 – Virtual (Tampa, FL), Springer Nature, 2022.

1 Introduction

The robustness analysis of networks is very active, especially for critical systems such as transportation, internet, and power-grid systems [Sun, 2021 ; Wandelt, 2018a]. It investigates how the network resists an attack on the nodes or links of a network. The goal is to design a strategy to dismantle the network by deleting the least number of nodes or links. Generally, one relies on centrality measures to extract essential nodes. Nodes are then removed according to their centrality rank. In this sense, a lot of work has been done. However, most papers do not consider the mesoscopic structure of complex networks in developing attack strategies unless the centrality [Ghalmane, 2019a] measure exploits the community structure. We present some related papers on targeted attacks based on the mesoscopic structure. In [Wandelt, 2021], the authors propose algorithms to dismantle a network based on the community structure. They target the inter-community nodes and the inter-community links. Adding intercommunity links, they also introduce a strategy for improving network resilience. The algorithms are evaluated using twelve networks from different regions (air, bus, rail, road). For node attacks, they compare eight attack strategies (4 attacks based on centrality measures and four based on alternative network dismantling). The paper explores five algorithms for edge attacks. Results show that, for node attacks, their algorithm is an excellent candidate to dismantle a network with a linear execution speed. The attacks based on the Betweenness centrality of the entire network outperform their strategy. Nevertheless, the latter requires more time, especially for large networks. For attacks, their method outperforms all the other methods. The approach to improve the robustness is more effective than its classical alternatives.

The authors of [Requião da Cunha, 2015], investigate an algorithm Module-Based Attack (MBA). It relies on the nodes or edges interconnecting the communities to fragment a network. Indeed, once the communities are extracted, these nodes or edges are removed in descending order of their Betweenness centrality (no sequential attacks). Ten real networks (4 infrastructural, three biological, and three social networks) are investigated. For attack comparisons, the Betweenness-based attack, Degree-based attack, and Longest Path attack are explored. Results show that the MBA is more efficient in disintegrating a network than the centrality measures. Nevertheless, this efficiency depends on the number of communities and the community size.

This chapter introduces a strategy to dismantle a network based on its global component. The global component is a part of the component structure, a new mesoscopic structure introduced in [Diop, 2021]. We investigate the world air transportation network in which the nodes represent the airports, and the links are the flights. In addition, one analyzes the impact of the attacks on world air transportation and the local components

2 Data and methods

2.1 Data

The undirected and unweighted network we consider originates from FlightAware[Flightaware,]. The extracted data covered six days of flight information, from May 17, 2018, to May 22, 2018. Nodes represent airports, and links represent direct flights between airports. Table 7.1 summarize its basic topological properties. The global air network is sparse and requires twelve hops to join at maximum any two airports. It is disassortative, and its clustering coefficient is small.

2.2 Component Structure

The component structure splits the network into local and global components. The local components capture the dense part of the network. The global components derive from the nodes and links that

TABLE 7.1 – Basic topological properties of the world air transportation network, the 7 large local components, and the largest global component. N is the network size. $|E|$ is the number of edges. *diam* is the diameter. l is the average shortest path length. ν is the density. ζ is the transitivity, also called the global clustering coefficient. $k_{nn}(k)$ is the assortativity, also called the Degree correlation coefficient. η is the hub dominance.

Components	N	$\|E\|$	*diam*	l	ν	ζ	$k_{nn}(k)$	η
World air transportation network	2734	16665	12	3,86	0,004	0,26	-0,05	0.09
North America-Caribbean	657	3828	7	2.88	0.018	0.28	-0.325	0.29
Europe	493	5181	6	2.58	0.042	0.32	-0.2	0.33
East and Southeast Asia	416	2495	7	2.9	0.029	0.34	-0.22	0.32
Africa-Middle East-India	337	1197	6	3.25	0.021	0.28	-0.15	0.23
Oceania	234	464	9	3.5	0.017	0.18	-0.22	0.24
South America	215	527	6	3.15	0.023	0.23	-0.35	0.22
Russia-Central Asia-Transcaucasia	112	427	4	2.23	0.07	0.23	-0.39	0.73
Large global component	513	2194	8	3.28	0.017	0.13	-0.25	0.2

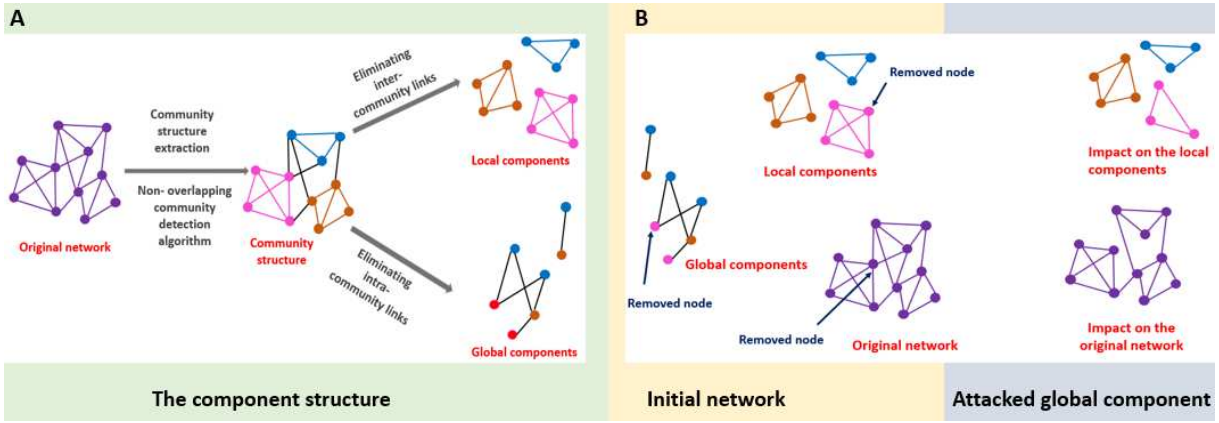


FIGURE 7.1 – A)The component structure B) Attack on the global component.

connect the local components. One can use a community structure or multiple core-periphery algorithms to uncover the dense groups. Then, one removes the connections between the dense groups to form the local components. Finally, one removes the links within the dense groups. The remained non-isolated nodes and links form the global components. Fig. 7.1.A illustrates the process of uncovering the component structure. This example relies on the non-overlapping community detection algorithm to extract the dense parts. Then, by eliminating the inter-community links (black lines), we form the local components. The global components are formed by eliminating the intra-community links (blue, yellow, and pink colors).

2.3 Targeted Attack on the Largest Global Component

We aim to disconnect the network while removing as few airports as possible. Centrality measures play an important role in extracting those vital nodes. A strong attack strategy eliminates the nodes in descending order of a given centrality measure. This chapter investigates the Degree and Betweenness centrality measures.

We propose a global component-based attack strategy. Then, we evaluate the impact of this attack on the local components and the whole network. The robustness evaluation process proceeds as follows :

1. Disconnect a node from the largest global component according to an attack strategy.

2. Disconnect the same node from its local component.
3. Disconnect the same node from the world air transportation network.
4. Extract the Largest Connected Component of the world air transportation network.
5. Extract the Largest Connected Component of the local components.

Note that the attack strategy based on the centrality is computed only once before the first attack. Fig. 7.1.B displays an example of attack on the global components.

2.4 Performance Measure

The Largest Connected Component (LCC) is widely used to measure the robustness of a network. It is the size of the largest connected component when the network is broken down into multiple components after an attack. The smallest the LCC the most effective the attack with the same budget (number of deleted airports). Let $LCC(i)$ be the size of the LCC when i nodes are removed from a network of size N, we call $\lambda(i) = \frac{LLC(i)*100}{N}$ the fraction of the LCC remaining after attack.

3 Experimental results

3.1 Component structure

We use the Louvain community detection algorithm to extract the dense parts of the network. It uncovers 27 communities that correspond to the local components. There are 20 small components located in a single country and seven large components covering different regions of the world [Diop, 2021]. The large components split the world into the following areas : 1) North America-Caribbean, 2) Europe, 3) East and Southeast Asia, 4) Africa-Middle East-Southern Asia, 5) Oceania, 6) South America, and 7) Russia-Central Asia- Transcaucasia.

The global components include nine small global components, and one large global component [Diop, 2021]. The airports of the largest global component are distributed over the world. In the following, we focus on the large local and global components. Table 7.1 presents their basic topological properties.

3.2 Degree attacks

The **Degree** of a node is the number of its connections. Fig. 7.2.A gives the fraction of the LCC of the large local components and the world air transportation network[Diop, 2022c] as a function of the fraction of top Degree nodes removed from the global component. One can see that the South America and Africa-Middle East-Southern Asia components are the most sensitive to attacks. The most resilient components seem to be Europe and East and Southeast Asia. The LLC of the North America-Caribbean, Europe, and East and Southeast Asia components decrease similarly. Indeed, when an airport is removed, the LCC decreases slowly. Removing an airport from the Oceania, South America, and Russia-Central Asia-Transcaucasia components affects their LCC significantly. The LCC of the Africa-Middle East-Southern Asia component and the world air transportation network combines the two schemes already mentioned when an airport is removed. Indeed, alternatively, their LCC decreases slowly and sharply.

The North America-Caribbean and the South America components are isolated from the world air transportation network after 5% of the airports are removed. Thus, these two components are more connected than the other components. This detached block represents 26% of the airports in the global network. Traveling between airports in this block is not easy. Indeed, the network is not very dense, and it takes a maximum of 19 hops and an average of 6 hops to connect two airports. The world air transportation network is more sensitive to degree attacks based on the global component than degree

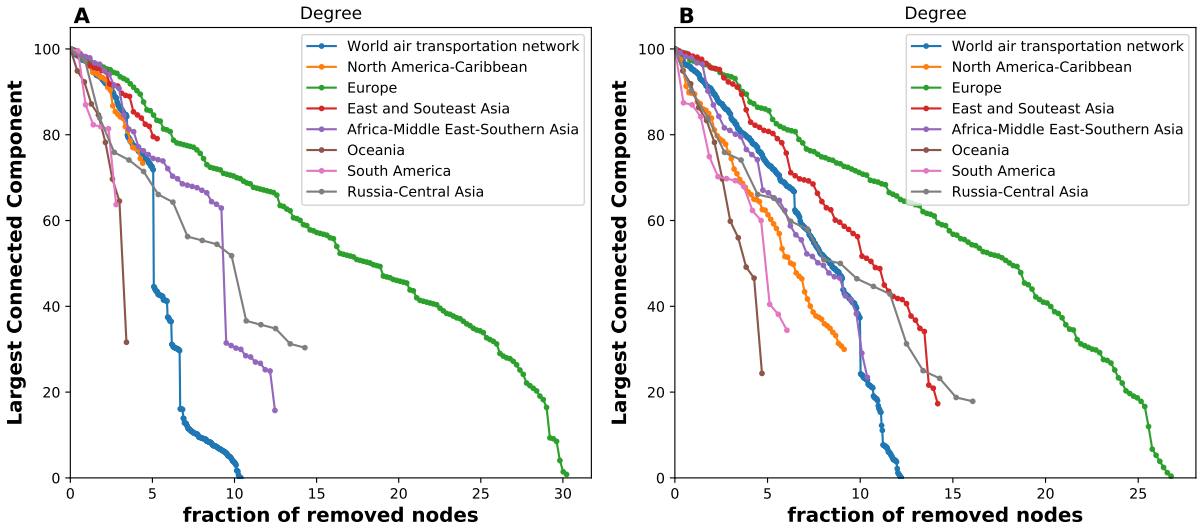


FIGURE 7.2 – A) Attack from the largest global component.B) Attack from the world air transportation network.

attacks based on the world air transportation network. Taking out 10% of the airports is enough to dismantle the overall network for attacks on the global component. While for the attacks on the world air transportation network, removing 12% is necessary to dismantle the network.

By removing more nodes from the global component, the local components become disconnected from the global air transportation network. Table 7.2 displays information about the local components, once they are isolated. Specifically, we determine how many nodes are required to isolate it from the global air transportation network. Additionally, it reports their giant components' size and basic topological properties.

The North America-Caribbean component is isolated from the network very early. 5% of the overall network's airports must be removed from the top degree global component airports to separate it. Removing the Calgary Airport in Canada provokes the isolation of this component. The LCC of this component contains 69% of the airport of the initial component. These airports cover all the countries of this component. It becomes challenging to travel in this LCC compared to the initial component. Indeed, several hubs are removed by the attacks. Thus, it requires more hops at maximum and, on average, to join any two airports in this LCC. In addition, this LCC is less dense than the initial component.

The South America component is simultaneously isolated from the global air network as the North America-Caribbean component. The attacks directly remove only 3% of the airports of this component. The Comodoro A M Benítez Airport in Chili is the last removed airport before the separation. The LCC of this component keeps 68.16% of the airports of the initial component. These airports are distributed over the countries of this component. On average, it requires the same number of hops to connect two airports in this LCC as in the initial component. Nevertheless, the diameter and the density of the LCC increase. Traveling is no more difficult than in the initial component.

The third isolated component is Oceania when removing 6.1% of the airports. Among these airports, 3.4% belong to this component. The Nadi Airport in Fiji is the last airport connecting this component to the world. Thus, the initial component separates into two components. The LCC includes 35% of the airports of the initial component. These airports cover only the North of Australia and New Zealand. The second largest connected component, which contains 30% of the airport, covers the countries such as Vanuatu, French Polynesia, and the Solomon Islands. Traveling in the LCC becomes more challenging

TABLE 7.2 – **Basic topological properties of the five local components isolated by the Degree attack.** λ is the Largest Connected Component size. ρ is the fraction of top Strength nodes removed from the world air transportation network to isolate the component. δ is the diameter. l is the average shortest path length. μ is the density.

Components	$\rho(\%)$	$\lambda(\%)$	δ	l	μ
North America-Caribbean	5	69	9	3.44	0.014
Europe	6.87	76.8	14	3.85	0.024
East and Southeast Asia	4.25	76.5	9	3.7	0.023
Africa-Middle East-Southern Asia	6.87	27.4	9	3.85	0.06
Oceania	6.1	27.4	15	5	0.037
South America	5	66.6	10	3.72	0.032
Russia-Central Asia-Transcaucasia	6.32	27.4	7	3.15	0.01

than in the initial component. Indeed, more hops are required to connect at maximum and, on average, two airports in the LCC.

The East and Southeast Asia component is the next region to separate. This isolation occurs when removing 6.25% of the top hubs. In these 6.25%, there are 5.2% of the airports of this component. The Chubu Centrair Airport in Japan is the last connected airport to the world before isolation. The LCC contains 75.4% of the airports of the initial component. The LCC covers all countries of this component. Nevertheless, traveling is more challenging in the LCC than in the initial component. Indeed, the LCC's diameter and the average shortest path length increase while the density decreases.

Removing 6.32% of the top degree airports from the global component, isolates the Russia-Central Asia-Transcaucasia component from the world. Among these airports, there are 14.2% belong to this component. The Kurumoch Airport in Russia is the last connected airport before the separation. The LCC includes 30.35% of the airports of the initial component. These airports are mainly in Russia, except for one airport in Tajikistan and another in Uzbekistan. The diameter almost double ; on average, one more jump is required to join two airports in this LCC. Moreover, the diameter decreases profoundly. Thus, it is more challenging to get from one airport to another.

The Africa-Middle East-Southern Asia is the following separated component when 6.87% of the airports are removed from the global component. 12.5% of the airports of this component are included in the 6.87%. The Kigali Airport is the last connected airport to Europe before the isolation. The LCC consists of 15.8% of the airports of the initial component. These airports are in the South of Africa. The diameter increase, whereas the average shortest path is the same. In addition, the LCC is denser than the initial component. One can say that traveling in the LCC is almost the same as in the initial component.

Once the Africa-Middle East-Southern Asia is isolated from the overall network, only the European airports remain. Although its LCC keeps 76.8% of the airport of the initial component, traveling within this LCC becomes more challenging. Indeed, this component is the most attacked. Several hubs are removed. Thus, the diameter value double, the average shortest path increases, and the density decrease.

3.3 Betweenness attacks

The **Betweenness** of a node is the fraction of the shortest path passing through it. Fig. 7.3.A represents the evolution of the LCC of the large component and the world air transportation network when removing the top Betweenness airports from the largest global component. One can see that South America and Oceania components are the most sensitive to the consecutive suppression of top Betweenness airports. The other components decrease in the same way until 3% of the airports are removed. Beyond

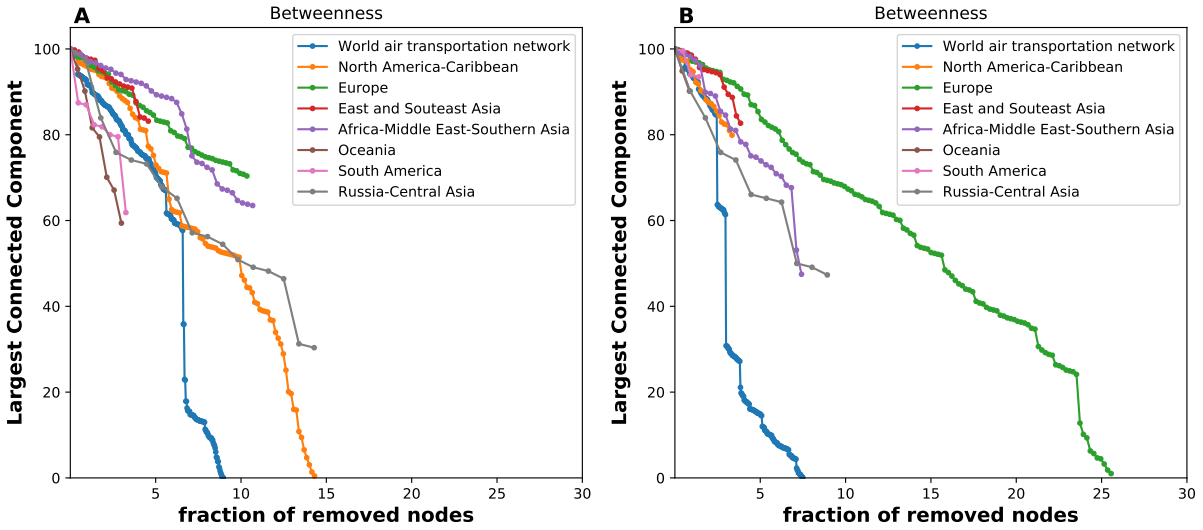


FIGURE 7.3 – A) Attack from the largest global component.B) Attack from the world air transportation network.

this value, the Africa-Middle East-Southern Asia and East and Southeast Asia components are the most resilient. The European component becomes more robust 8% of the airports are removed.

The same tendencies of curves that we see in the attack of the top degrees are present in the Betweenness attacks. The evolution of the LCC of the components is the same except for the Africa-Middle East-Southern Asia component. Indeed, removing an airport slightly impacts this component. Table 7.3 reports some topological properties of the LCCs once they are isolated.

East and Southeast Asia, the Africa-Middle East-Southern Asia, and a part of the Russia-Central Asia-Transcaucasia component are isolated from the world air transportation network after 6.14% of the airports are removed from the largest global component. Removing the Irkutsk Airport in Russia isolates these three components simultaneously. This LCC includes 21% of the airports of the global air network. Nevertheless, traveling within these three regions is more challenging. Indeed, the diameter and the average shortest path length of this LCC have almost doubled the ones of the global air network, even though one can easily find a stopover to join two airports. The global air transport network is more resistant to Betweenness-based attacks on the global component than the global air network. Nevertheless, the difference is not very big. These two attacks are almost comparable. Indeed, for the attack based on the global component, it is necessary to remove 7.8% of the airports to dismantle the global network,whereas, removing 7.4% on the global network is enough.

The Oceania component is the first isolated part of the global air transportation network. Removing 4.6% of the airports from the global component provokes this separation. Among these airports, only 3% belonging to this component are directly removed. This component becomes unreachable when removing the Brisbane Airport. The LCC contains 59.4% of the airports of the initial component. The LCC covers the West of Australia and the other countries of this component. Traveling within this LCC is very challenging, despite increasing density. Indeed, a maximum of 19 jumps are necessary to join any two airports. On average, one more hop is required to join two airports compared to the initial component.

Removing 4.9% of the airports separates South America from the global air network. 3.25% of the airports of this component belongs to the 4.9%. The Comodoro Arturo M B Airport is the last airport that ensures the connection between this region and the world. After the separation, the LCC contains 61.8% of its airports, which are distributed over this component, except Venezuela, Colombia, and Ecuador. On

average, the same number of jumps is required to travel. Nevertheless, long-distance travel requires more hops and the density increase.

East and Southeast Asia is the next region to separate from the world. 6.14% of the airports are sufficient to isolate it. 4.5% of the airports of this component are included in the 6.14%. The Tan Son Nhat Airport in Vietnam is the world's last connected airport. 83.17% of the airports of the initial component, covering all the countries, belong to the LCC. Traveling within this LCC is a bit more challenging. Indeed, the diameter increases by more than two jumps, and the average shortest path increase by more than one hop. In addition, the density decreases.

TABLE 7.3 – **Basic topological properties of the five local components isolated by the Betweenness attack.** λ is the Largest Connected Component size. ρ is the fraction of top Strength nodes removed from the world air transportation network to isolate the component. δ is the diameter. l is the average shortest path length. μ is the density.

Components	$\rho(\%)$	$\lambda(\%)$	δ	l	μ
North America-Caribbean	6.65	71.38	12	3.57	0.013
Europe	6.65	71	10	3	0.032
East and Southeast Asia	6.14	83.17	9	3.5	0.023
Africa-Middle East-Southern Asia	6.14	63.5	11	5.64	0.02
Oceania	4.9	61.8	18	4.66	0.021
South America	5	66.6	10	3.72	0.033
Russia-Central Asia-Transcaucasia	6.14	30.35	7	3.17	0.09

The Africa-Middle East-Southern Asia component is the next region to split. The Bahrain Airport in Bahrain is the last airport connected to the other region. 10% of the airports of this component is directly eliminated. The LCC keeps 63.5% of the airports of the initial component. These airports cover the region. Nevertheless, taking off from one airport to another requires almost twice as many hops on average and at maximum.

The Russia-Central Asia-Transcaucasia component is separated from the world air transportation network when the Irkutsk Airport in Russia is removed. Among the removed airports, there 14% of the airports of this component. The LCC contains 30.35% of the airports of the initial component. It covers only Russia, except Tajikistan and Uzbekistan. The diameter has almost doubled compared to the one of the initial component, and the average shortest path increases by one. But, The density increases. Thus, it becomes difficult to travel between any two airports.

Removing 6.65% of the airport of the global air network isolates the European component. Among these airports, there are 10.14% of the European airports. Eliminating the Nice-Côte d'Azur Airport in France separates the Europe component from the North America-Caribbean component. The LCC contains 71% of the airports of the initial component. These airports are located in all the countries of this component. On average, the same number of hops are required to join two airports. Nevertheless, the diameter increases by three jumps, and the density decreases. Thus, traveling within this LCC is a bit challenging compared to the initial component.

Once all these components are separated, the North America-Caribbean component remains in the global air network. Its LCC includes 71.38% of the airport of the initial component. These airports cover all the components except Alaska in the United States. Traveling within the LCC becomes difficult. Indeed, the diameter and the average shortest have almost doubled, while the density decreases.

TABLE 7.4 – γ, ρ, λ are the order of isolation of local components, the proportion of airports removed from the global air network before isolation, and LCCs of the local components after the separation.

Components	Degree						Betweenness					
	Global component			World air transportation network			Global component			World air transportation network		
	γ	$\rho(\%)$	$\lambda(\%)$	γ	$\rho(\%)$	$\lambda(\%)$	γ	$\rho(\%)$	$\lambda(\%)$	γ	$\rho(\%)$	$\lambda(\%)$
North America-Caribbean	1	5	69	2	9.7	29.8	7	6.65	71.38	4	2.9	79.6
Europe	7	6.87	76.8	7	11.4	27.5	6	6.65	71	7	5	61.8
East and Southeast Asia	4	4.25	76.5	4	11	17.2	3	6.14	83.17	3	2.3	82.4
Africa-Middle East-Southern Asia	6	6.87	27.4	5	11.2	23.5	3	6.14	63.5	5	3.9	47.6
Oceania	3	6.1	27.4	1	9	24.3	1	4.9	61.8	2	2.1	90.2
South America	1	5	66.5	3	9.8	34.4	2	5	66.6	1	1.6	93.5
Russia-Central Asia-Transcaucasia	5	6.32	27.4	6	11.4	18.7	3	6.14	30.35	6	5	47.32

4 Discussion and Conclusion

This chapter investigates the vulnerability of air transport to targeted attacks on the global component of the network using two centrality measures (Degree and Betweenness). In addition, we explore their impact on the various local components. The global air transportation network is more vulnerable to attacks on the global component based on Betweenness than Degree. For the Degree, the attacks on the global component cause more damage than the attacks on the world air transportation network. For the Betweenness, removing an airport from the world air transportation network is more efficient than the global component. Nevertheless, the difference in efficiency is not large. Indeed, eliminating 7.5% of the airports from the world air transportation network is sufficient to dismantle it. In contrast, one needs to remove 7.9% of the airports from the global component to fragment the overall network. Consequently, the strategy proposed by this study can be an alternative to the Betweenness-based attacks on the global air transport network. Indeed, the computation of the Betweenness on large networks is less efficient.

Table 7.4 reports the order of isolation of the local components, the proportion of airports removed from the global air network before isolation, and the LCCs of the local components after separation. It allows us to compare the attacks of the global component and the impacts on the local components and the world's air transportation. The local components progressively separate from the world network as airports are removed from the global component. Nevertheless, the order of isolation of the components differs with the attacks. For the Degree, the North America-Caribbean and the South America components

separate first from the overall network. Then Oceania, East and Southeast Asia, Russia-Central Asia-Transcaucasia, Africa-Middle East-Southern Asia, and Europe follow in that order. For Betweenness attacks, Oceania and the South America components are first isolated from the overall network. Then, the East and Southeast Asia, Africa-Middle East-Southern Asia, and Russia-Central Asia-Transcaucasia components are unreachable simultaneously from the overall network. The Europe and North America-Caribbean components are the last isolated components. Thus, Oceania and South America are the most vulnerable components. The European component is the most robust for Degree attacks, while the North America-Caribbean component is the most resilient for Betweenness attacks.

The local components separate faster from the global air network with the attacks on the global component than on the overall network for the Degree attack. Indeed, it is necessary to remove half as many airports on the global component for the first local component to be isolated. The LCCs from the attacks on the global component are larger and connect more countries. However, two categories emerge. At first, one can see that the LCC of North America-Caribbean, Europe, East, and Southeast Asia, and South America, from the attacks on the global component, are twice as large. Traveling within its LCCs is more accessible than the LCCs from the attacks on the overall network. The size LCCs of the remained components are comparable. Nevertheless, the LCC from the attack on the global component of Africa-Middle East-Southern Asia is usable. In the second category, it is the opposite. Indeed, with much fewer airports, three times less, the attack on the world air network isolates a first local component. In addition, the LCCs from this attack tend to include more airports, except the LCCs of Europe, East and Southeast Asia, and Africa-Middle East-Southern Asia components. Except for the Africa-Middle East-Southern Asia component, traveling within the LCCs from the attack on the overall network is less challenging. This chapter explores exclusively a targeted attack on the global component of the world air transportation network with two centrality measure (Degree and Betweenness). Future work will consider the impact of various attacks on the local components. Additionally, we plan to investigate other transportation networks.

Conclusion and future works

This thesis introduces a new mesoscopic structure called the component structure, which includes local components and global components. The local components are the dense parts of a network. The global components represent the interactions between the local components. We can rely on the community structure or the multiple core-periphery structure to extract the components. Thus, throughout this manuscript, we have shown the advantages of the component structure using the global air transportation network as a case study. Thus, several contributions have been made, particularly in finding the most important nodes, in the extraction of k-core or s-core, and the robustness.

First, we analyzed the component structure of the global air network, the most important airports, and the k-core. The analysis of the component structure shows the heterogeneity of the network. The large components share some properties with the global air network. This shows, in a way, the fractal nature of this network. The component structure allows us to see a node as a vector. A node belongs to a local component, global component, or both. In the air transport network context, an airport may have a regional or inter-regional influence or both. Concerning the k-core analysis, we have seen that a network can have several k-cores instead of one.

Then, we add the weight to understand the global air network based on the component structure. Thus, we consider the weighted centrality and the weighted k-core, which is called s-core. This comparative study shows that the results are partial by examining the network as a whole, and the conclusions can change radically. Whereas with the component structure, the results are more consistent.

In the third part, we study the robustness of the global network using classical attacks. In this part, we evaluate how the attacks on the global network influence the components. Both weighted and unweighted classical exchanges are considered attacks. Thus, the components become separated from the network as the nodes are removed from the network. This experiment shows us, beyond the effectiveness of the attacks, the resilience of the components.

Finally, we propose a new attack strategy. Indeed, we remove the global component airports while evaluating the robustness of the global air network and of its local components. The classical centrality measures are exploited. This attack strategy proves to be very efficient and requires less computation time since the global component is smaller than the entire network.

In future work, we plan to explore multiple directions. The most straightforward is to consider directed networks and other applications such as co-authorship networks analysis. One can also look at the hierarchical structure of the representation. Indeed, components are networks that can also be split into components and so on. It can be an excellent way to gain information about the fractal nature of real-world networks. The interplay between the core-periphery and the component structure is another direction of research. The component structure can be represented as a multilayer network in which each local component is a layer, and the links of the global component are the links between the layers. Furthermore, another perspective is to consider the impact of various attacks on the local components.

Additionally, we plan to investigate other transportation networks.

List of publications

Journals

Issa Moussa Diop, Cherif Diallo, Chantal Cherifi, Hocine Cherifi. "*Robustness Analysis of the Regional and Inter-Regional Components of the Weighted World Air Transportation Network*". Complexity, vol. 2022, Article ID 6595314, 17 pages, 2022. https ://doi.org/10.1155/2022/6595314.

Issa Moussa Diop, Chantal Cherifi, Cherif Diallo, Hocine Cherifi. "*Revealing the component structure of the world air transportation network*". https ://doi.org/10.1007/s41109-021-00430-2, Applied Network Science, 2021, vol. 6, no 1, p. 1-50. Springer Nature, 2021.

Conferences

Issa Moussa Diop, Chantal Cherifi, Cherif Diallo, Hocine Cherifi. "*Targeted Attack of the Air Transportation Network Global Component*". The 11th International Conference on Computational Data and Social Networks. CSoNet 2022, December 5-7, 2022 – Virtual (Tampa, FL), Springer Nature, 2022.

Issa Moussa Diop, Chantal Cherifi, Cherif Diallo, Hocine Cherifi. "*Robustness of the Weighted World Air Transportation Network Components*". 2022 IEEE Workshop on Complexity in Engineering (COMPENG), 18-20 July 2022, Florence, Italy, pp. 1-6.

Issa Moussa Diop, Chantal Cherifi, Cherif Diallo, Hocine Cherifi. "*Targeted attacks on the world air transportation network : impact on its regional structure*". Complex Networks XIII : Proceedings of the 13th Conference on Complex Networks, CompleNet Live 2022, Exeter, UK, May 30-June 1, 2022. Springer Nature. https ://doi.org/10.1007/978-3-031-17658-6_1.

Preliminary works, Abstracts & Posters

Issa Moussa Diop, Cherif Diallo, Chantal Cherifi, Hocine Cherifi. "*Analyse de la robustesse du réseau de transport aérien mondial : impact sur sa structure en composante*". French Regional Conference on Complex Systems, FRCCS 2022.

Issa Moussa Diop, Chantal Cherifi, Cherif Diallo, Hocine Cherifi. "*On local and global components of the air transportation network*". Conference on Complex Systems (CCS), 2020.

Issa Moussa Diop, Cherif Diallo, Chantal Cherifi, Hocine Cherifi. "*Exploring The Mesoscopic Structure Of The World Air Transportation Network*". Belgian Network Research Meeting (BENet), 2020.

Issa Moussa Diop, Cherif Diallo, Chantal Cherifi, Hocine Cherifi. "*Core Analysis of the components of the air transportation network*". Symposium DSTN, Université Gaston Berger, December 6-10, 2021.

Issa Moussa Diop, Chantal Cherifi, Cherif Diallo, Hocine Cherifi. "*On local and global components of the air transportation network*". Symposium DSTN, CEA-MITIC, Université Gaston Berger, December 6-10, 2021.

Bibliography

[Alves, 2020] Luiz GA Alves, Alberto Aleta, Francisco A Rodrigues, Yamir Moreno et Luis A Nunes Amaral. « Centrality anomalies in complex networks as a result of model over-simplification ». *New Journal of Physics* 22.1 (2020), p. 013043 (cf. p. 32, 85).

[Aparicio, 2022] Joao Tiago Aparicio, Elisabete Arsenio et Rui Henriques. « Assessing robustness in multimodal transportation systems : a case study in Lisbon ». *European Transport Research Review* 14.1 (2022), p. 1-18 (cf. p. 134).

[Bagler, 2008] Ganesh Bagler. « Analysis of the airport network of India as a complex weighted network ». *Physica A : Statistical Mechanics and its Applications* 387.12 (2008), p. 2972-2980 (cf. p. 29, 84, 94, 98).

[Bao, 2021] Xiaoge Bao, Peng Ji, Wei Lin, Matjaž Perc et Jürgen Kurths. « The impact of COVID-19 on the worldwide air transportation network ». *Royal Society open science* 8.11 (2021), p. 210682 (cf. p. 134).

[Barabási, 1999] Albert-László Barabási et Réka Albert. « Emergence of scaling in random networks ». *science* 286.5439 (1999), p. 509-512 (cf. p. 10).

[Barrat, 2004] Alain Barrat, Marc Barthelemy, Romualdo Pastor-Satorras et Alessandro Vespignani. « The architecture of complex weighted networks ». *Proceedings of the national academy of sciences* 101.11 (2004), p. 3747-3752 (cf. p. 85, 94, 101).

[Blöcker, 2022] Christopher Blöcker, Juan Carlos Nieves et Martin Rosvall. « Map equation centrality : community-aware centrality based on the map equation ». *Applied Network Science* 7.1 (2022), p. 56 (cf. p. 103).

[Blondel, 2008] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte et Etienne Lefebvre. « Fast unfolding of communities in large networks ». *Journal of statistical mechanics : theory and experiment* 2008.10 (2008), P10008 (cf. p. 12-14, 33, 86, 140).

[Borgatti, 2000] Stephen P Borgatti et Martin G Everett. « Models of core/periphery structures ». *Social networks* 21.4 (2000), p. 375-395 (cf. p. 14, 26, 83).

[Boyd, 2006] John P Boyd, William J Fitzgerald et Robert J Beck. « Computing core/periphery structures and permutation tests for social relations data ». *Social networks* 28.2 (2006), p. 165-178 (cf. p. 14).

[Brandes, 2001] Ulrik Brandes. « A faster algorithm for betweenness centrality ». *Journal of mathematical sociology* 25.2 (2001), p. 163-177 (cf. p. 16, 137).

[Brandes, 2007] Ulrik Brandes, Daniel Delling, Marco Gaertler, Robert Gorke, Martin Hoefer, Zoran Nikoloski et al. « On modularity clustering ». *IEEE transactions on knowledge and data engineering* 20.2 (2007), p. 172-188 (cf. p. 13).

[Broido, 2019] Anna D Broido et Aaron Clauset. « Scale-free networks are rare ». *Nature communications* 10.1 (2019), p. 1-10 (cf. p. 11, 40, 41).

[Carrington, 2005] Peter J Carrington, John Scott et Stanley Wasserman. *Models and methods in social network analysis*. T. 28. Cambridge university press, 2005 (cf. p. 3).

[Chakraborty, 2016] Debayan Chakraborty, Anurag Singh et Hocine Cherifi. « Immunization strategies based on the overlapping nodes in networks with community structure ». *International conference on computational social networks*. Springer, Cham. 2016, p. 62-73 (cf. p. 103, 126, 136).

[Cherifi, 2018] Hocine Cherifi. *Non-overlapping community detection*. 2018. arXiv : 1805.11584 [cs.SI] (cf. p. 33).

[Cherifi, 2019a] Hocine Cherifi, Gergely Palla, Boleslaw K Szymanski et Xiaoyan Lu. « On community structure in complex networks : challenges and opportunities ». *Applied Network Science* 4.1 (2019), p. 1-35 (cf. p. 33).

[Cherifi, 2019b] Hocine Cherifi, Gergely Palla, Boleslaw K Szymanski et Xiaoyan Lu. « On community structure in complex networks : challenges and opportunities ». *Applied Network Science* 4.1 (2019), p. 1-35 (cf. p. 83, 135).

[Cheung, 2012] Dorothy P Cheung et Mehmet Hadi Gunes. « A complex network analysis of the United States air transportation ». *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE. 2012, p. 699-701 (cf. p. 30).

[Cheung, 2020] Tommy KY Cheung, Collin WH Wong et Anming Zhang. « The evolution of aviation network : Global airport connectivity index 2006–2016 ». *Transportation Research Part E : Logistics and Transportation Review* 133 (2020), p. 101826 (cf. p. 28, 84, 85).

[Chunaev, 2020] Petr Chunaev. « Community detection in node-attributed social networks : AÂ survey ». *Computer Science Review* 37 (2020), p. 100286 (cf. p. 33).

[Colizza, 2006] Vittoria Colizza, Alain Barrat, Marc Barthélemy et Alessandro Vespignani. « The role of the airline transportation network in the prediction and predictability of global epidemics ». *Proceedings of the National Academy of Sciences* 103.7 (2006), p. 2015-2020 (cf. p. 26).

[Coscia, 2011] Michele Coscia, Fosca Giannotti et Dino Pedreschi. « A classification for community discovery methods in complex networks ». *Statistical Analysis and Data Mining : The ASA Data Science Journal* 4.5 (2011), p. 512-546 (cf. p. 12).

[Costa, 2018] Bernardo Costa, João Victor Bechara, Klaus Wehmuth et Artur Ziviani. *A Multilayer and Time-varying Structural Analysis of the Brazilian Air Transportation Network*. 2018. arXiv : 1709.03360 [cs.SI] (cf. p. 30).

[Costa, 2021] Luciano da F Costa. « Further generalizations of the Jaccard index ». *arXiv preprint arXiv :2110.09619* (2021) (cf. p. 138).

[Couto, 2015] Guilherme S Couto, Ana Paula Couto DA Silva, Linnyer B Ruiz et Fabrıcio Benevenuto. « Structural properties of the Brazilian air transportation network ». *Anais da Academia Brasileira de Ciências* 87.3 (2015), p. 1653-1674 (cf. p. 30, 56).

[Crucitti, 2003] Paolo Crucitti, Vito Latora, Massimo Marchiori et Andrea Rapisarda. « Efficiency of scale-free networks : error and attack tolerance ». *Physica A : Statistical Mechanics and its Applications* 320 (2003), p. 622-642 (cf. p. 17, 137).

[Cucuringu, 2016] Mihai Cucuringu, Puck Rombach, Sang Hoon Lee et Mason A Porter. « Detection of core–periphery structure in networks using spectral methods and geodesic paths ». *European Journal of Applied Mathematics* 27.6 (2016), p. 846-887 (cf. p. 14).

[Da Silva, 2008] Marcio Rosa Da Silva, Hongwu Ma et An-Ping Zeng. « Centrality, network capacity, and modularity as parameters to analyze the core-periphery structure in metabolic networks ». *Proceedings of the IEEE* 96.8 (2008), p. 1411-1420 (cf. p. 14).

[Dai, 2018] Liang Dai, Ben Derudder et Xingjian Liu. « The evolving structure of the Southeast Asian air transport network through the lens of complex networks, 1979–2012 ». *Journal of Transport Geography* 68 (2018), p. 67-77 (cf. p. 26, 29).

[Das, 2018] Kousik Das, Sovan Samanta et Madhumangal Pal. « Study on centrality measures in social networks : a survey ». *Social Network Analysis and Mining* 8.1 (2018), p. 13 (cf. p. 103).

[Ding, 2011] Ying Ding. « Applying weighted PageRank to author citation networks ». *Journal of the American Society for Information Science and Technology* 62.2 (2011), p. 236-245 (cf. p. 16, 137).

[Diop, 2021] Issa Moussa Diop, Chantal Cherifi, Cherif Diallo et Hocine Cherifi. « Revealing the component structure of the world air transportation network ». *Applied Network Science* 6.1 (2021), p. 1-50 (cf. p. 19, 23, 25, 83, 86, 90, 103, 105, 107, 125, 127, 135, 140, 157, 159).

[Diop, 2022a] Issa Moussa Diop, Chantal Cherifi, Cherif Diallo et Hocine Cherifi. « Robustness of the Weighted World Air Transportation Network Components ». *2022 IEEE Workshop on Complexity in Engineering (COMPENG)*. IEEE. 2022, p. 1-6 (cf. p. 20, 24, 133).

[Diop, 2022b] Issa Moussa Diop, Chantal Cherifi, Cherif Diallo et Hocine Cherifi. « Targeted Attack of the Air Transportation Network Global Component ». *The 11th International Conference on Computational Data and Social Networks, CSoNet 2022, December 5-7, 2022 – Virtual (Tampa, FL), Springer Nature*. 2022 (cf. p. 20, 24, 156).

[Diop, 2022c] Issa Moussa Diop, Chantal Cherifi, Cherif Diallo et Hocine Cherifi. « Targeted attacks on the world air transportation network : impact on its regional structure ». *arXiv preprint arXiv :2209.05617, Complex Networks XIII : Proceedings of the 13th Conference on Complex Networks, CompleNet 2022. Springer Nature* (2022) (cf. p. 20, 24, 123, 159).

[Diop, 2022d] Issa Moussa Diop, Cherif Diallo, Chantal Cherifi et Hocine Cherifi. « Exploring the Regional Structure of the Worldwide Air Traffic and Route Networks ». *Submitted to IEEE Access* (2022) (cf. p. 19, 24, 82).

[Diop, 2022e] Issa Moussa Diop, Cherif Diallo, Chantal Cherifi et Hocine Cherifi. « Robustness Analysis of the Regional and Inter-Regional Components of the WeightedWorld Air Transportation Network ». *Complexity* 2022, Article ID 6595314 (2022) (cf. p. 20, 24, 133).

[Dobruszkes, 2006] Frédéric Dobruszkes. « An analysis of European low-cost airlines and their networks ». *Journal of Transport Geography* 14.4 (2006), p. 249-264 (cf. p. 55).

[Du, 2017a] Wen-Bo Du, Bo-Yuan Liang, Chen Hong et Oriol Lordan. « Analysis of the Chinese provincial air transportation network ». *Physica A : Statistical Mechanics and its Applications* 465 (2017), p. 579-586 (cf. p. 29).

[Du, 2017b] Wenbo Du, Boyuan Liang, Gang Yan, Oriol Lordan et Xianbin Cao. « Identifying vital edges in Chinese air route network via memetic algorithm ». *Chinese Journal of Aeronautics* 30.1 (2017), p. 330-336 (cf. p. 29).

[Eidsaa, 2016] Marius Eidsaa et Eivind Almaas. « Investigating the relationship between k-core and s-core network decompositions ». *Physica A : Statistical Mechanics and its Applications* 449 (2016), p. 111-125 (cf. p. 113).

[Erd6s, 1959] P Erd6s et A Renyi. « On random graphs ». *Publ. Math. Debrecen* 6 (1959), p. 290-297 (cf. p. 9).

[Flightaware,] Flightaware. « https ://flightaware.com/ » () (cf. p. 32, 85, 126, 138, 157).

[Fortunato, 2016a] Santo Fortunato et Darko Hric. « Community detection in networks : A user guide ». *Physics reports* 659 (2016), p. 1-44 (cf. p. 11, 26).

[Fortunato, 2016b] Santo Fortunato et Darko Hric. « Community detection in networks : A user guide ». *Physics Reports* 659 (2016), p. 1-44 (cf. p. 33, 83).

[Ghalmane, 2019a] Zakariya Ghalmane, Chantal Cherifi, Hocine Cherifi et Mohammed El Hassouni. « Centrality in complex networks with overlapping community structure ». *Scientific reports* 9.1 (2019), p. 1-29 (cf. p. 17, 26, 31, 136, 157).

[Ghalmane, 2019b] Zakariya Ghalmane, Mohammed El Hassouni, Chantal Cherifi et Hocine Cherifi. « Centrality in modular networks ». *EPJ Data Science* 8.1 (2019), p. 15 (cf. p. 103).

[Girvan, 2002] Michelle Girvan et Mark EJ Newman. « Community structure in social and biological networks ». *Proceedings of the national academy of sciences* 99.12 (2002), p. 7821-7826 (cf. p. 13).

[Guida, 2007] Michele Guida et Funaro Maria. « Topology of the Italian airport network : A scale-free small-world network with a fractal structure ? » *Chaos, Solitons & Fractals* 31.3 (2007), p. 527-536 (cf. p. 29).

[Guimera, 2004] Roger Guimera et Luis A Nunes Amaral. « Modeling the world-wide airport network ». *The European Physical Journal B* 38.2 (2004), p. 381-385 (cf. p. 27).

[Guimera, 2005] Roger Guimera, Stefano Mossa, Adrian Turtschi et LA Nunes Amaral. « The worldwide air transportation network : Anomalous centrality, community structure, and cities' global roles ». *Proceedings of the National Academy of Sciences* 102.22 (2005), p. 7794-7799 (cf. p. 26, 28, 84).

[Holme, 2005] Petter Holme. « Core-periphery organization of complex networks ». *Physical Review E* 72.4 (2005), p. 046111 (cf. p. 14).

[Hossain, 2017] Md Murad Hossain et Sameer Alam. « A complex network approach towards modeling and analysis of the Australian Airport Network ». *Journal of Air Transport Management* 60 (2017), p. 1-9 (cf. p. 30, 84).

[Huffman, 1952] David A Huffman. « A method for the construction of minimum-redundancy codes ». *Proceedings of the IRE* 40.9 (1952), p. 1098-1101 (cf. p. 12).

[Ibnoulouafi, 2018] Ahmed Ibnoulouafi, Mohamed El Haziti et Hocine Cherifi. « M-centrality : identifying key nodes based on global position and local degree variation ». *Journal of Statistical Mechanics : Theory and Experiment* 2018.7 (2018), p. 073407 (cf. p. 103, 126, 136).

[Javed, 2018] Muhammad Aqib Javed, Muhammad Shahzad Younis, Siddique Latif, Junaid Qadir et Adeel Baig. « Community detection in networks : A multidisciplinary review ». *Journal of Network and Computer Applications* 108 (2018), p. 87-111 (cf. p. 33).

[Jebabli, 2018] Malek Jebabli, Hocine Cherifi, Chantal Cherifi et Atef Hamouda. « Community detection algorithm evaluation with ground-truth data ». *Physica a : Statistical mechanics and its applications* 492 (2018), p. 651-706 (cf. p. 33).

[Jeude, 2019] J van Lidth de Jeude, Guido Caldarelli et Tiziano Squartini. « Detecting core-periphery structures by surprise ». *EPL (Europhysics Letters)* 125.6 (2019), p. 68001 (cf. p. 14).

[Jia, 2014] Tao Jia, Kun Qin et Jie Shan. « An exploratory analysis on the evolution of the US airport network ». *Physica A : Statistical Mechanics and its Applications* 413 (2014), p. 266-279 (cf. p. 30).

[Jin, 2021a] Di Jin, Zhizhi Yu, Pengfei Jiao, Shirui Pan, Dongxiao He, Jia Wu et al. « A Survey of Community Detection Approaches : From Statistical Modeling to Deep Learning ». *IEEE Transactions on Knowledge and Data Engineering* (2021), p. 1-1 (cf. p. 140).

[Jin, 2021b] Di Jin, Zhizhi Yu, Pengfei Jiao, Shirui Pan, Dongxiao He, Jia Wu et al. « A survey of community detection approaches : From statistical modeling to deep learning ». *IEEE Transactions on Knowledge and Data Engineering* (2021) (cf. p. 135).

[Jusup, 2022] Marko Jusup, Petter Holme, Kiyoshi Kanazawa, Misako Takayasu, Ivan Romić, Zhen Wang et al. « Social physics ». *Physics Reports* 948 (2022), p. 1-148 (cf. p. 134).

[Karrer, 2008] Brian Karrer, Elizaveta Levina et Mark EJ Newman. « Robustness of community structure in networks ». *Physical review E* 77.4 (2008), p. 046119 (cf. p. 13).

[Kim, 2019] Seyun Kim et Yoonjin Yoon. « On node criticality of the Northeast Asian air route network ». *Journal of Air Transport Management* 80 (2019), p. 101693 (cf. p. 135).

[Kirkley, 2021] Alec Kirkley et M. E. J. Newman. *Representative community divisions of networks*. 2021. arXiv : 2105.04612 [cs.SI] (cf. p. 33).

[Kojaku, 2017a] Sadamori Kojaku et Naoki Masuda. « Finding multiple core-periphery pairs in networks ». *Physical Review E* 96.5 (2017), p. 052313 (cf. p. 14, 83, 135).

[Kojaku, 2017b] Sadamori Kojaku et Naoki Masuda. « Finding multiple core-periphery pairs in networks ». *Phys. Rev. E* 96 (5 2017), p. 052313 (cf. p. 135).

[Kong, 2019a] Yi-Xiu Kong, Gui-Yuan Shi, Rui-Jie Wu et Yi-Cheng Zhang. « k-core : Theories and applications ». *Physics Reports* 832 (2019), p. 1-32 (cf. p. 113).

[Kong, 2019b] Yi-Xiu Kong, Gui-Yuan Shi, Rui-Jie Wu et Yi-Cheng Zhang. « k-core : Theories and applications ». *Physics Reports* 832 (2019), p. 1-32 (cf. p. 113).

[Kumar, 2018] Manish Kumar, Anurag Singh et Hocine Cherifi. « An efficient immunization strategy using overlapping nodes and its neighborhoods ». *Companion Proceedings of the The Web Conference 2018*. 2018, p. 1269-1275 (cf. p. 103, 136).

[Lancichinetti, 2010] Andrea Lancichinetti, Mikko Kivelä, Jari Saramäki et Santo Fortunato. « Characterizing the community structure of complex networks ». *PloS one* 5.8 (2010), e11976 (cf. p. 39).

[LÃŒ, 2016] Linyuan LÃŒ, Duanbing Chen, Xiao-Long Ren, Qian-Ming Zhang, Yi-Cheng Zhang et Tao Zhou. « Vital nodes identification in complex networks ». *Physics Reports* 650 (2016), p. 1-63 (cf. p. 103).

[Lee, 2014] Sang Hoon Lee, Mihai Cucuringu et Mason A Porter. « Density-based and transport-based core-periphery structures in networks ». *Physical Review E* 89.3 (2014), p. 032810 (cf. p. 14, 83).

[Lordan, 2017] Oriol Lordan et Jose M Sallan. « Analyzing the multilevel structure of the European airport network ». *Chinese Journal of Aeronautics* 30.2 (2017), p. 554-560 (cf. p. 26, 28, 62).

[Lordan, 2019] Oriol Lordan et Jose M Sallan. « Core and critical cities of global region airport networks ». *Physica A : Statistical Mechanics and its Applications* 513 (2019), p. 724-733 (cf. p. 26, 29, 124, 127).

[Lordan, 2016] Oriol Lordan, Jose M Sallan, Nuria Escorihuela et David Gonzalez-Prieto. « Robustness of airline route networks ». *Physica A : Statistical Mechanics and its Applications* 445 (2016), p. 18-26 (cf. p. 124).

[Lordan, 2014a] Oriol Lordan, Jose M Sallan et Pep Simo. « Study of the topology and robustness of airline route networks from the complex network approach : a survey and research agenda ». *Journal of Transport Geography* 37 (2014), p. 112-120 (cf. p. 124, 134).

[Lordan, 2014b] Oriol Lordan, Jose M Sallan, Pep Simo et David Gonzalez-Prieto. « Robustness of the air transport network ». *Transportation Research Part E : Logistics and Transportation Review* 68 (2014), p. 155-163 (cf. p. 134).

[Lordan, 2014c] Oriol Lordan, Jose M. Sallan et Pep Simo. « Study of the topology and robustness of airline route networks from the complex network approach : a survey and research agenda ». *Journal of Transport Geography* 37 (2014), p. 112-120 (cf. p. 27).

[Lü, 2016] Linyuan Lü, Duanbing Chen, Xiao-Long Ren, Qian-Ming Zhang, Yi-Cheng Zhang et Tao Zhou. « Vital nodes identification in complex networks ». *Physics Reports* 650 (2016), p. 1-63 (cf. p. 15, 17, 136, 137).

[Menczer, 2020] Filippo Menczer, Santo Fortunato et Clayton A Davis. *A first course in network science*. Cambridge University Press, 2020 (cf. p. 2).

[Milgram, 1967] Stanley Milgram. « The small world problem ». *Psychology today* 2.1 (1967), p. 60-67 (cf. p. 3, 10).

[Montgolfier, 2011] Fabien de Montgolfier, Mauricio Soto et Laurent Viennot. « Asymptotic modularity of some graph classes ». *International Symposium on Algorithms and Computation*. Springer. 2011, p. 435-444 (cf. p. 13).

[Newman, 2018] Mark Newman. *Networks*. Oxford university press, 2018 (cf. p. 2).

[Newman, 2004] Mark EJ Newman. « Detecting community structure in networks ». *The European physical journal B* 38.2 (2004), p. 321-330 (cf. p. 83).

[Oehlers, 2021] Milena Oehlers et Benjamin Fabian. « Graph Metrics for Network RobustnessâA Survey ». *Mathematics* 9.8 (2021) (cf. p. 137).

[Oliveira, 2020] Izabela M Oliveira, Laura C Carpi et A P F Atman. « The Multiplex Efficiency Index : unveiling the Brazilian air transportation multiplex network—BATMN ». *Scientific Reports* 10.1 (2020), p. 13339 (cf. p. 30).

[Orman, 2011a] Günce Keziban Orman, Vincent Labatut et Hocine Cherifi. « On accuracy of community structure discovery algorithms ». *arXiv preprint arXiv :1112.4134* (2011) (cf. p. 86).

[Orman, 2011b] Günce Keziban Orman, Vincent Labatut et Hocine Cherifi. « Qualitative comparison of community detection algorithms ». *International conference on digital information and communication technology and its applications*. Springer, Berlin, Heidelberg. 2011, p. 265-279 (cf. p. 86).

[Orman, 2012] Günce Keziban Orman, Vincent Labatut et Hocine Cherifi. « Comparative evaluation of community detection algorithms : a topological approach ». *Journal of Statistical Mechanics : Theory and Experiment* 2012.08 (2012), P08001 (cf. p. 33, 34).

[Orman, 2013] Keziban Orman, Vincent Labatut et Hocine Cherifi. « An empirical study of the relation between community structure and transitivity ». *Complex Networks*. Springer, Berlin, Heidelberg, 2013, p. 99-110 (cf. p. 135).

[Papadopoulos, 2012] Symeon Papadopoulos, Yiannis Kompatsiaris, Athena Vakali et Ploutarchos Spyridonos. « Community detection in Social Media ». *Data Mining and Knowledge Discovery* 24.3 (2012), p. 515-554 (cf. p. 33).

[Pons, 2005] Pascal Pons et Matthieu Latapy. « Computing communities in large networks using random walks ». *International symposium on computer and information sciences*. Springer. 2005, p. 284-293 (cf. p. 12).

[Raghavan, 2007] Usha Nandini Raghavan, Réka Albert et Soundar Kumara. « Near linear time algorithm to detect community structures in large-scale networks ». *Physical review E* 76.3 (2007), p. 036106 (cf. p. 12, 33).

[Rajeh, 2020] Stephany Rajeh, Marinette Savonnet, Eric Leclercq et Hocine Cherifi. « Interplay between hierarchy and centrality in complex networks ». *IEEE Access* 8 (2020), p. 129717-129742 (cf. p. 17, 136).

[Rajeh, 2021] Stephany Rajeh, Marinette Savonnet, Eric Leclercq et Hocine Cherifi. « Characterizing the interactions between classical and community-aware centrality measures in complex networks ». *Scientific reports* 11.1 (2021), p. 1-15 (cf. p. 83).

[Ravasz, 2003] Erzsébet Ravasz et Albert-László Barabási. « Hierarchical organization in complex networks ». *Physical review E* 67.2 (2003), p. 026112 (cf. p. 14, 94).

[Requião da Cunha, 2015] Bruno Requião da Cunha, Juan Carlos González-Avella et Sebastián Gonçalves. « Fast Fragmentation of Networks Using Module-Based Attacks ». *PLOS ONE* 10.11 (2015), p. 1-15 (cf. p. 157).

[Riolo, 2020] Maria A. Riolo et M. E. J. Newman. « Consistency of community structure in complex networks ». *Phys. Rev. E* 101 (5 2020), p. 052306 (cf. p. 34).

[Rocha, 2009] Luis E C da Rocha. « Structural evolution of the Brazilian airport network ». *Journal of Statistical Mechanics : Theory and Experiment* 2009.04 (2009), P04020 (cf. p. 30).

[Rocha, 2017] Luis E.C. Rocha. « Dynamics of air transport networks : A review from a complex systems perspective ». *Chinese Journal of Aeronautics* 30.2 (2017), p. 469-478 (cf. p. 27).

[Rombach, 2017] Puck Rombach, Mason A Porter, James H Fowler et Peter J Mucha. « Core-periphery structure in networks (revisited) ». *SIAM review* 59.3 (2017), p. 619-646 (cf. p. 14).

[Rosvall, 2008] Martin Rosvall et Carl T Bergstrom. « Maps of random walks on complex networks reveal community structure ». *Proceedings of the national academy of sciences* 105.4 (2008), p. 1118-1123 (cf. p. 12).

[Sales-Pardo, 2007] Marta Sales-Pardo, Roger Guimera, André A Moreira et Luis A Nunes Amaral. « Extracting the hierarchical organization of complex systems ». *Proceedings of the National Academy of Sciences* 104.39 (2007), p. 15224-15229 (cf. p. 14).

[Serrat, 2017] Olivier Serrat. « Social network analysis ». *Knowledge solutions*. Springer, 2017, p. 39-43 (cf. p. 3).

[Siozos-Rousoulis, 2021] Leonidas Siozos-Rousoulis, Dimitri Robert et Wouter Verbeke. « A study of the U.S. domestic air transportation network : temporal evolution of network topology and robustness from 2001 to 2016 ». *Journal of Transportation Security* 14.1 (2021), p. 55-78 (cf. p. 30).

[Sobolevsky, 2014] Stanislav Sobolevsky, Riccardo Campari, Alexander Belyi et Carlo Ratti. « General optimization technique for high-quality community detection in complex networks ». *Physical Review E* 90.1 (2014), p. 012811 (cf. p. 86).

[Sun, 2017a] Xiaoqian Sun, Volker Gollnick et Sebastian Wandelt. « Robustness analysis metrics for worldwide airport network : A comprehensive study ». *Chinese Journal of Aeronautics* 30.2 (2017), p. 500-512 (cf. p. 124, 134).

[Sun, 2021] Xiaoqian Sun et Sebastian Wandelt. « Robustness of air transportation as complex networks : Systematic review of 15 years of research and outlook into the future ». *Sustainability* 13.11 (2021), p. 6446 (cf. p. 124, 134, 157).

[Sun, 2017b] Xiaoqian Sun, Sebastian Wandelt et Massimiliano Zanin. « Worldwide air transportation networks : a matter of scale and fractality ? » *Transportmetrica A : Transport Science* 13.7 (2017), p. 607-630 (cf. p. 28).

[Tang, 2019] Wenli Tang, Liutao Zhao, Wei Liu, Yiping Liu et Bo Yan. « Recent advance on detecting core-periphery structure : a survey ». *CCF Transactions on Pervasive Computing and Interaction* 1.3 (2019), p. 175-189 (cf. p. 135).

[Tarkhov, 2017] SA Tarkhov. « Changes in air transport connectivity of Russian cities in 1990–2015 ». *Regional Research of Russia* 7.2 (2017), p. 127-145 (cf. p. 56).

[Verma, 2014] Trivik Verma, Nuno AM Araújo et Hans J Herrmann. « Revealing the structure of the world airline network ». *Scientific reports* 4.1 (2014), p. 1-6 (cf. p. 134).

[Voitalov, 2019] Ivan Voitalov, Pim van der Hoorn, Remco van der Hofstad et Dmitri Krioukov. « Scale-free networks well done ». *Physical Review Research* 1.3 (2019), p. 033034 (cf. p. 41).

[Wan, 2021] Zelin Wan, Yash Mahajan, Beom Woo Kang, Terrence J Moore et Jin-Hee Cho. « A survey on centrality metrics and their network resilience analysis ». *IEEE Access* 9 (2021), p. 104773-104819 (cf. p. 17, 136).

[Wandelt, 2021] Sebastian Wandelt, Xing Shi et Xiaoqian Sun. « Estimation and improvement of transportation network robustness by exploiting communities ». *Reliability Engineering & System Safety* 206 (2021), p. 107307 (cf. p. 134, 157).

[Wandelt, 2018a] Sebastian Wandelt, Xiaoqian Sun, Daozhong Feng, Massimiliano Zanin et Shlomo Havlin. « A comparative analysis of approaches to network-dismantling ». *Scientific reports* 8.1 (2018), p. 1-15 (cf. p. 17, 136, 157).

[Wandelt, 2018b] Sebastian Wandelt, Xiaoqian Sun, Daozhong Feng, Massimiliano Zanin et Shlomo Havlin. « A comparative analysis of approaches to network-dismantling ». *Scientific Reports* 8.1 (2018), p. 13513 (cf. p. 137).

[Wandelt, 2019] Sebastian Wandelt, Xiaoqian Sun et Jun Zhang. « Evolution of domestic airport networks : a review and comparative analysis ». *Transportmetrica B : Transport Dynamics* 7.1 (2019), p. 1-17. eprint : https://doi.org/10.1080/21680566.2017.1301274 (cf. p. 29).

[Wang, 2011] Jiaoe Wang, Huihui Mo, Fahui Wang et Fengjun Jin. « Exploring the network structure and nodal centrality of China's air transport network : A complex network approach ». *Journal of Transport Geography* 19.4 (2011), p. 712-721 (cf. p. 29).

[Webber, 2010] William Webber, Alistair Moffat et Justin Zobel. « A similarity measure for indefinite rankings ». *ACM Transactions on Information Systems (TOIS)* 28.4 (2010), p. 1-38 (cf. p. 107).

[Wilkinson, 2012] Sean M Wilkinson, Sarah Dunn et Shu Ma. « The vulnerability of the European air traffic network to spatial hazards ». *Natural hazards* 60.3 (2012), p. 1027-1036 (cf. p. 3).

[Xiang, 2018] Bing-Bing Xiang, Zhong-Kui Bao, Chuang Ma, Xingyi Zhang, Han-Shuang Chen et Hai-Feng Zhang. « A unified method of detecting core-periphery structure and community structure in networks ». *Chaos : An Interdisciplinary Journal of Nonlinear Science* 28.1 (2018), p. 013122 (cf. p. 14).

[Xu, 2008] Zengwang Xu et Robert Harriss. « Exploring the structure of the U.S. intercity passenger air transportation network : a weighted complex network approach ». *GeoJournal* 73.2 (2008), p. 87 (cf. p. 30, 84).

[Yang, 2021] Yong Yang, Kai-Jun Xu et Chen Hong. « Network dynamics on the Chinese air transportation multilayer network ». *International Journal of Modern Physics C* 32.05 (2021), p. 2150070. eprint : https://doi.org/10.1142/S0129183121500704 (cf. p. 29).

[Yook, 2002] Soon-Hyung Yook, Hawoong Jeong et Albert-László Barabási. « Modeling the Internet's large-scale topology ». *Proceedings of the National Academy of Sciences* 99.21 (2002), p. 13382-13386. eprint : https://www.pnas.org/content/99/21/13382.full.pdf (cf. p. 28).

[Zanin, 2013] Massimiliano Zanin et Fabrizio Lillo. « Modelling the air transport with complex networks : A short review ». *The European Physical Journal Special Topics* 215.1 (2013), p. 5-21 (cf. p. 26, 84).

[Zhang, 2019] Lu Zhang, Hongru Du, Xiaolei Zhang, Philippe De Maeyer, Bart Dessein et Yannan Zhao. « Robustness and edge addition strategy of air transport networks : a case study of "the belt and road" ». *IEEE Access* 7 (2019), p. 96470-96477 (cf. p. 134, 135).

[Zhang, 2021] Lu Zhang, Yannan Zhao, Dongli Chen et Xinhuan Zhang. « Analysis of network robustness in weighted and unweighted approaches : a case study of the air transport network in the belt and road region ». *Journal of Advanced Transportation* 2021 (2021) (cf. p. 124).

[Zhang, 2015] Xiao Zhang, Travis Martin et Mark EJ Newman. « Identification of core-periphery structure in networks ». *Physical Review E* 91.3 (2015), p. 032803 (cf. p. 14, 83).

[Zhou, 2020] Bin Zhou, Xiangyi Meng et H Eugene Stanley. « Power-law distribution of degree–degree distance : A better representation of the scale-free property of complex networks ». *Proceedings of the national academy of sciences* 117.26 (2020), p. 14812-14818 (cf. p. 11, 41).

[Zhou, 2021] Wei Zhou, Hong Huang, Qiang-Sheng Hua, Dongxiao Yu, Hai Jin et Xiaoming Fu. « Core decomposition and maintenance in weighted graph ». *World Wide Web* 24.2 (2021), p. 541-561 (cf. p. 113).

[Zhou, 2019] Yaoming Zhou, Junwei Wang et George Q Huang. « Efficiency and robustness of weighted air transport networks ». *Transportation Research Part E : Logistics and Transportation Review* 122 (2019), p. 14-26 (cf. p. 17, 134, 137).

RÉSUMÉ

Les réseaux complexes constituent un domaine scientifique qui permet de modéliser des systèmes complexes en se basant sur les graphes qui représentent le mieux les interactions de différents objets, personnes, sites, entités, etc. L'étude des réseaux complexes permet ainsi de mieux comprendre plusieurs phénomènes dans différents systèmes complexes tels que les systèmes financiers, la propagation d'épidémie, les réseaux informatiques, les réseaux de transports, etc. L'analyse d'un réseau complexe peut se faire sur trois niveaux: le niveau macroscopique, le niveau microscopique et le niveau mésoscopique. Au niveau macroscopique, différentes mesures globales sont utilisées pour caractériser le réseau. Au niveau microscopique, on s'intéresse à la propriété d'un nœud ou d'un lien. Au niveau mésoscopique, on met l'accent sur l'émergence de groupes de nœuds qui ont des similarités. Les deux structures mésoscopiques les plus étudiées sont la structure communautaire et la structure noyau-périphérie. En effet, elles sont présentes dans beaucoup de réseaux réels et trouvent plusieurs applications. Cependant, aucun des deux ne permet de dissocier les interactions dans les groupes et les interactions entre les groupes. Ainsi, dans cette thèse, nous proposons une nouvelle structure mésoscopique appelée la structure en composante qui permet de démêler les interactions locales et globales des structures communautaires et noyau-périphérie. Ainsi, un réseau peut être vu comme un agrégat de sous-réseaux qu'on peut analyser séparément. Extensible également aux réseaux pondérés, cette nouvelle structure mésoscopique offre de nombreux champs d'application telles que la recherche de noeuds critiques, l'extraction de k-core, la robustesse, etc. Ainsi, nous l'avons appliqué au réseau mondial de transport aérien dont on a aussi étudié la résilience en proposant de nouvelles stratégies d'attaques plus efficaces comparées aux attaques classiques.

MOTS CLÉS

Réseaux complexes, Structure mésoscopique, Structure communautaire, Structure noyau-périphéreie, Structure en composante, Mesure de centralité, k-core, s-core, Robustesse

ABSTRACT

Complex networks are a scientific field that makes it possible to model complex systems based on the graphs that best represent the interactions of different objects, people, sites, entities, etc. the study of complex networks thus makes it possible to better understand several phenomena in different complex systems such as financial systems, the spread of epidemics, computer networks, transport networks, etc. The analysis of a complex network can be done on three levels: the macroscopic level, the microscopic level and the mesoscopic level. At the macroscopic level, different global measurements are used to characterize the network. At the microscopic level, we are interested in the property of a node or a link. At the mesoscopic level, emphasis is placed on the emergence of groups of nodes that have similarities. The two most studied mesoscopic structures are the community structure and the core-periphery structure. Indeed, they are present in many real networks and find several applications. However, neither of the two makes it possible to dissociate the interactions in the groups and the interactions between the groups. Thus, in this thesis, we propose a new mesoscopic structure called the component structure which allows to disentangle the local and global interactions of community and core-periphery structures. Thus, a network can be seen as an aggregate of sub-networks that can be analyzed separately. Extendable also to weighted networks, this new mesoscopic structure offers many fields of application such as the search for critical nodes, k-core extraction, robustness, etc. Thus, we have applied it to the global air transport network, whose resilience has also been studied by proposing new attack strategies that are more effective compared to conventional attacks.

KEYWORDS

Complex networks, Mesoscopic structure, Community structure, Core-periphery structure, Component structure, Centrality measure, k-core, s-core, Robustness