Working Papers

ISSN 2219-780X

3

SEPTEMBER 2018

Taxation, Big Data and Network Analytics:

An introductory analysis to the global network of double taxation treaties

Santiago Díaz de Sarralde Miguez

CIAT

Taxation, Big Data and Network Analytics:

An introductory analysis to the global network of double taxation treaties

Santiago Diaz de Sarralde Miguez.
CIAT Director of Tax Studies and Research.

Taxation, Big Data and Network Analytics:
An introductory analysis to the global network of double taxation treaties

ISSN: 2219-780X

Intellectual property

Content

Tables and Graphics

Tables

Graphics

Annex 1

1. Introduction

Among the approximately 200 countries and jurisdictions existing in the world, we can trace back some 40,000 possible ways in which the profits can be transferred between two different countries. There are now more than 3,500 bilateral double taxation agreements (DTA), as well as regional agreements (European Union, CARICOM, West African Economic and Monetary Union, etc.) that alter the conditions of taxation of dividends, interests or royalties, modifying more than 7,000 of those possible paths, generally in a favorable direction. If we also consider that for the transfer of profits between jurisdictions, hundreds of thousands of indirect ways could be used, taking advantage of differences in taxation justifiably or as a form of treaty abuse (treaty shopping), the complexity of the international taxation panorama may seem unmanageable.

In recent years however, the coincidence of the development of storage capacities and data processing, together with their increased availability in many sectors, has caused an exponential development of mathematics, statistics and computer techniques for their analysis.

The paradigm that currently defines the availability of complex databases of massive dimensions is known as Big Data, while among the new techniques for analysis, the Network Analysis is highlighted for the novelty of its approach.

In this paper, we aim to provide an introduction to the use of network analysis in the study of massive databases that can be useful for Tax Administration, using the analysis of the global network of double taxation agreements as an illustrative example.

The following section introduces to Network Analysis and justifies its potential usefulness in the field of tax administration. Next, we will focus on its application to the global network of double taxation agreements, describing their characteristics and analyzing the position of the different countries in the network for the particular case of international transfers of dividends.

Finally, we will collect other experiences and fields of application of the Network Analysis Techniques in the tax administration scope.

2. Introduction to Network Analysis[1]

A simple example: a company X, resident in country C, has a subsidiary in country A. Both in A and in C, the withholding of the dividend distributed to nonresident is 10%. Between A and C there is no DTA, while it is in place between A and B (with a reduced fee of 5% for dividends paid to non-residents) and between B and C (with a reduced rate of 1%). In all countries the domestic distribution of dividends are not taxed.

What options does this company X have to repatriate their dividends from A? Are there possibilities of Treaty Shopping? How might the different countries be interested in negotiating their network of agreements or modifying their withholding policies?

For the analysis, the parameters of the problem can be synthesized by a matrix and represented by a graph[2] in which each country would be symbolized by a node or vertex and the relationships between these by arcs or links (edges):

Table 1. Example of international taxation of dividends

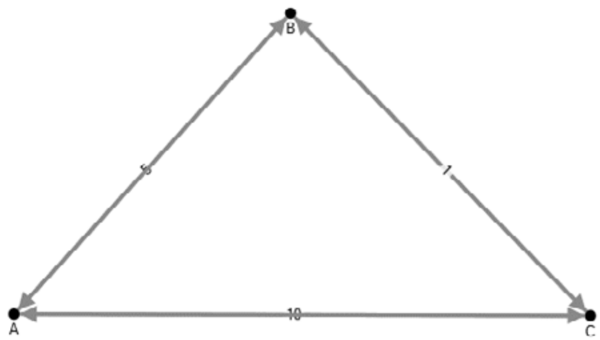
		RESIDENCE		
		A	B	C
SOURCE	A	0	5	10
	B	5	0	1
	C	10	1	0

Source: Own elaboration.

1 The origin of the mathematical analysis of networks is often attributed to Euler and his analysis of the problem of the bridges of Königsberg (now Kaliningrad), a city divided into four parts by the river Pregel and that had 7 bridges. Euler proved that it was impossible to draw a path that would allow crossing the seven bridges by crossing each of them only once.

2 Graphs represented by Excel NodeXL.

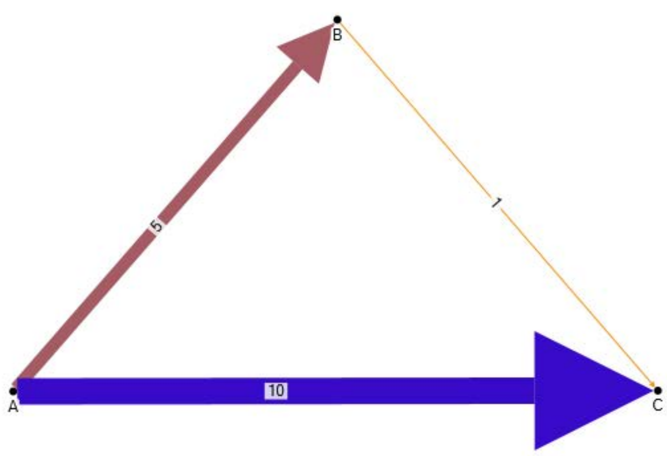
Graph 1. Example of international taxation of dividends



Source: Own elaboration wirth NodeXL.

In this analysis framework the determination, for example, of the shortest path (defined as the lowest taxation) from A to C is obvious: it could be profitable to use an intermediary company (conduit company) located in B to take advantage of the lower taxation deriving from the agreements network (lowering the rates from 10 to 5 + 1).

Graph 2. Treaty Shopping

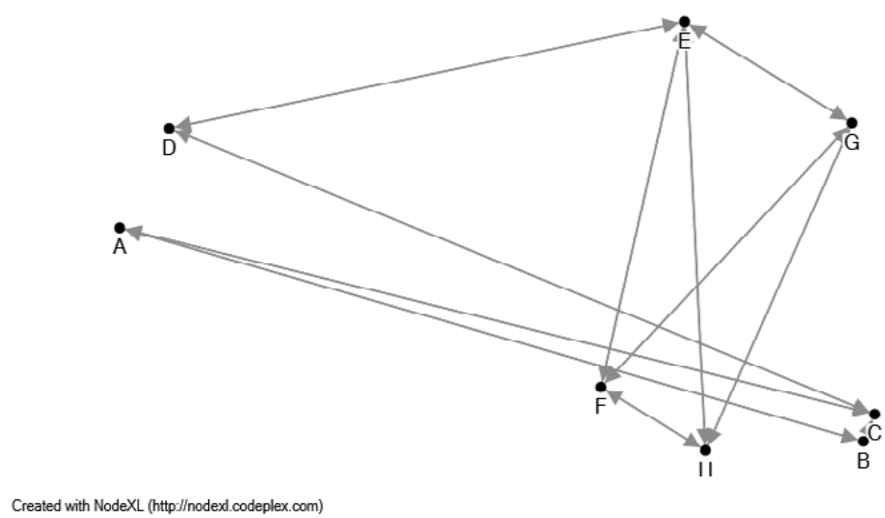


Source: Own elaboration with NodeXL.

From the point of view of the tax administration of country A, identifying possible tax minimizing strategies can be used to assess agreement renegotiation options (e.g. eliminating or increasing the reduced rate between A and B or negotiating a treaty with C that would make the route competitive) or the introduction of anti-abuse provisions.

The usefulness of this systematic approach to the data analysis multiplies exponentially with the number of actors, while available metrical indicators expand. Imagine an expanded network of double taxation[3] as shown in Graph 3.

Graph 3. Example of international network of agreements



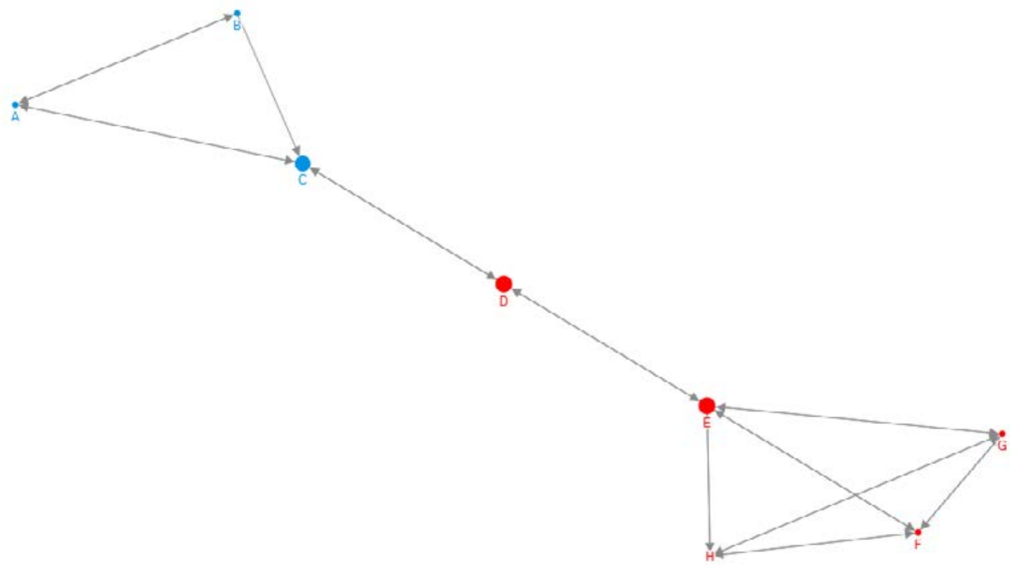
Source: Own elaboration with NodeXL.

In this case, the internal structure of the network and possible strategies are more difficult to identify at a glance. Fortunately, the available software allows reorganizing the graphical representation according to different guidelines through multiple algorithms (layout algorithms) that facilitate the understanding of relationships. In this case, the application of the Harel-Koren Fast Multiscale Layout method (based on what are called attraction and repulsion forces, determined by the connections between the nodes), the formation of groups or clusters (represented by colors) and differentiation of nodes size depending on their "centrality"[4] give us a much clearer picture of the relationships (Graph 4).

3 We can assume that other connections may not even still be profitable in the absence of agreement and the transfer of dividends may be subject to higher withholding rates. Additionally, as a particular case to introduce variability, the connection from B to C is considered unidirectional. The weight of the new edges is 1, except between C and D (5).

4 In this case, the betweenness centrality whose interpretation and calculation are explained below.

Graph 4. Example of presentation of international agreements network



Source: Own elaboration with NodeXL.

Thus, the role that the networks of agreements by groups of countries and, at the same time, the relative importance of each in the set of connections is much clearer (rendering techniques allow improving representation in a highly flexible way, giving colors and dimensions to the nodes and links according to multiple parameters). Countries A, B and C would form a group, with less individual agreements in force, while E, F, G and H form another, more integrated, cluster. Individually, C, D and E play an essential role as a bridge in the communication between the different countries, with E including concentrating more direct and strategic connection, while D, despite its small number of relationships, maximizes its relevance thanks to its intermediary position.

As our object of study becomes more complicated, the simple graphical representation, despite its usefulness as a first approximation, may not be accurate enough to analyze the connections. This requires using different indicators or statistics that summarize various features of the whole network, the groups of nodes and each of the individual vertices.

In the above example the basic overall characteristics of the network agreements are:

- Diameter: 5; the maximum number of connections necessary to link the farthests node.
- Graph density: 0.375, ratio of the number of existing links (21) relative to the total of possibilities (n * (n-1) = 8 * 7 = 56).
- Average path length: 2.321; the average of the shortest paths between all countries.

- The interpretation of these statistics varies depending on the context, being much more useful when taking measurements of similar networks or the same network at different times to compare. In any case, they show how complete and "efficient" the network is.

The main individual statistics that we can handle are summarized in Table 2.

Table 2. Example network analysis of international conventions

Countries	In degree	Out degree	Eccentricity	Betweenness Centrality	Closeness Centrality	Centrality Eigenvector	Page rank	Clustering coefficient
A	2	2	4	0.14	0,37	0.13	0.11	0.50
B	1	2	4	0.00	0,37	0.06	0.06	1.00
C	3	2	3	0,48	0,47	0.24	0.14	0.33
D	2	2	3	0,57	0,54	0,42	0.12	0.00
E	4	4	4	0,57	0,54	1.00	0.18	0.50
F	3	3	5	0.00	0,41	0.89	0.13	1.00
G	3	3	5	0.00	0,41	0.89	0.13	1.00
H	3	3	5	0.00	0,41	0.89	0.13	1.00

Source: Own elaboration with Gephi (https://gephi.org/).

Very briefly, the input and output degrees (in-degree, out-degree) show the number of directional relationships[5] that affect each node, in our case the input or output paths for dividends. Eccentricity measures the maximum number of steps to be performed from each node to reach the farthest point on the network. They are the simplest measures and a first approximation (very nuanced, as we shall see) to the importance (greater degree; lesser eccentricity) and role of each node in the network[6] .

Centrality measures offer different perspectives on the role of each node (country) in the internal network connections (circulation of dividends, potentially minimizing international taxation by leveraging agreements).

Betweenness centrality measures the frequency with which each node is on the shortest path between any two nodes (its value in this case is normalized between zero and one, dividing by the total number of possible paths). Closeness (Closeness centrality) is the inverse of the average of the shortest paths from each node to all others[7]. These two measures are inversely related to the weights of each link, in this case with the rates to be paid, meaning that the lower the rates established in the agreements the higher the index value of centrality. Because of their ease of interpretation, these indicators are the

5 Of course, these techniques can also be applied to undirected networks.
6 Unweighted values are given in the table, that is, regardless of the value of links (in our case, the rates established in the agreements).
7 The treatment of unreachable nodes represents a difficulty. Overall this index is zero when it is an isolated node. The harmonic centrality measure (*harmonic centrality*) is a variant that tries to simplify this problem by calculating the average of the inverse of the shortest paths.

most commonly used, particularly the one of intermediation, in the analysis of double taxation treaties networks.

The Eigenvector Centrality is an extension of the measures of centrality in which the neighbors (neighboring nodes) are weighted, in function of their respective centrality[8]. Using this information, the PageRank value can be obtained iteratively, measuring the frequency or probability of ending up in each of the nodes taking into account their connections[9]. The "page ranking" method of Google is the best-known example of using such algorithms using data links, searches and frequency of visits. In the case of the analysis of the connections between countries via agreements (bilateral in nature), it is not easy to interpret these indicators especially when the value of the rates is included, weighting further the highest values. This makes their value a combination of two contradictory effects (greater when the "neighbors" have more agreements, but also when rates are higher). However, these features can be very useful in other contexts, such as the study of possible fraud networks, relying on information on trade and financial flows.

Among the indicators shown in the table we also find the coefficient of Clustering, which shows for each node the probability for its closest nodes[10] to be in turn connected to each other, which allows us to identify the most interconnected groups (identified by colors on the graph). Globally, the network would have a Clustering coefficient of 0.667 (its values being between 0 and 1, indicating the frequency of triangular connections in relation to all possible ones).

As can be seen, the calculation of indicators reinforces, makes accurate and quantifies the features already observed in the visual analysis of the graph (Graph 4), which will be most useful in complex environments, while expanding the possibilities of analysis to other environments of interest for tax administrations.

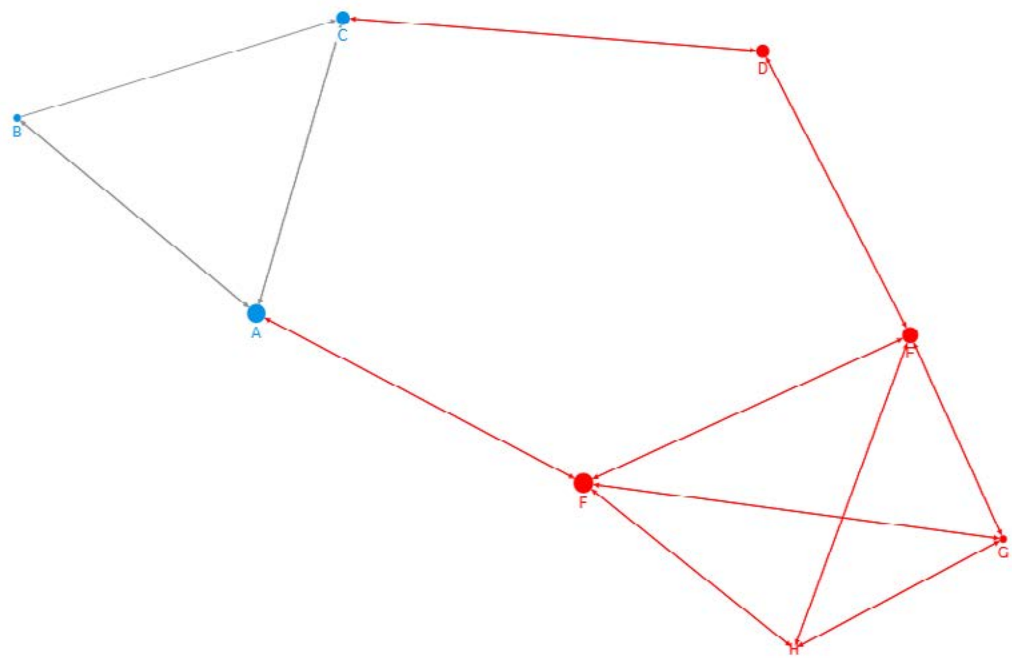
In addition, the analysis can take a dynamic perspective and/or allow comparisons of potential transformations. For example, we could analyze the consequences of signing a new agreement between A and F (with rate 1). The network would become as shown in Graph 5, significantly increasing the role of both countries in the flow of dividends, which would displace D and E in the classifications based on the values of the centrality indicators. Of course, at the same time overall network connectivity would increase, reducing its diameter to 3, increasing the density to 0.411 and lowering the average length of the connections to 1.786.

8 Its value is normalized, being 1 for the most connected node according to these parameters.

9 *HITS* Indicators*, Hub and Authority Scoring*, originally also designed to rank websites, assess centrality by the role of each node: receiving links from other relevant nodes (*Authority*) or as a link to other important nodes (*Hub*).

10 Intuitively, it involves analyzing whether our "friends" are, in turn, friends among them. This effect is called "*small-world* ". Technically, it proceeds to identify all possible "triangles" and measure the actually existing network connections.

Graph 5. Example transformation network of international conventions

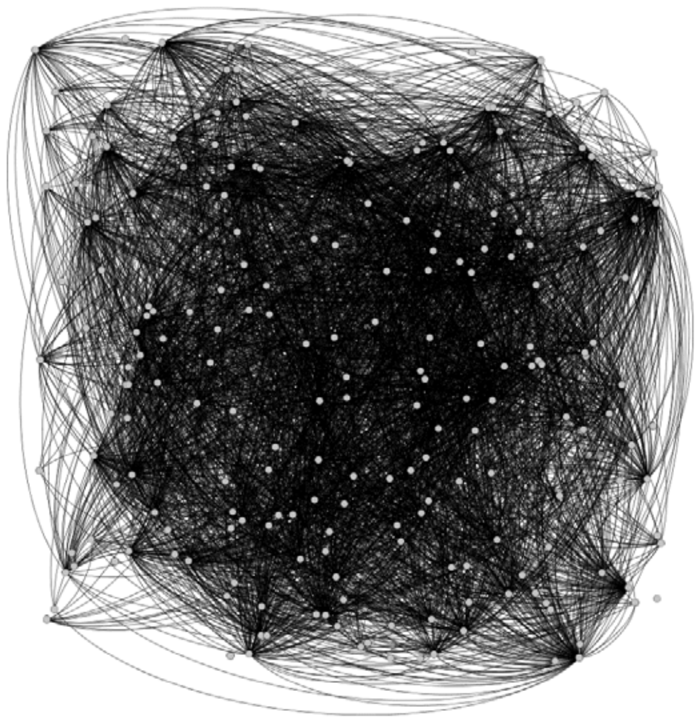


Source: Own elaboration with NodeXL.

3. The international network of Double Taxation Agreements

Based on the database of IBFD[11] tax conventions, we can set up a network of 217 nodes (countries / jurisdictions) and 3,417 agreements (and double the number of bilateral-edges). Its first graphic representation with random arrangement (Graph 6) clearly highlights its complexity.

Graph 6. "Gross" view of the international network of double taxation agreements

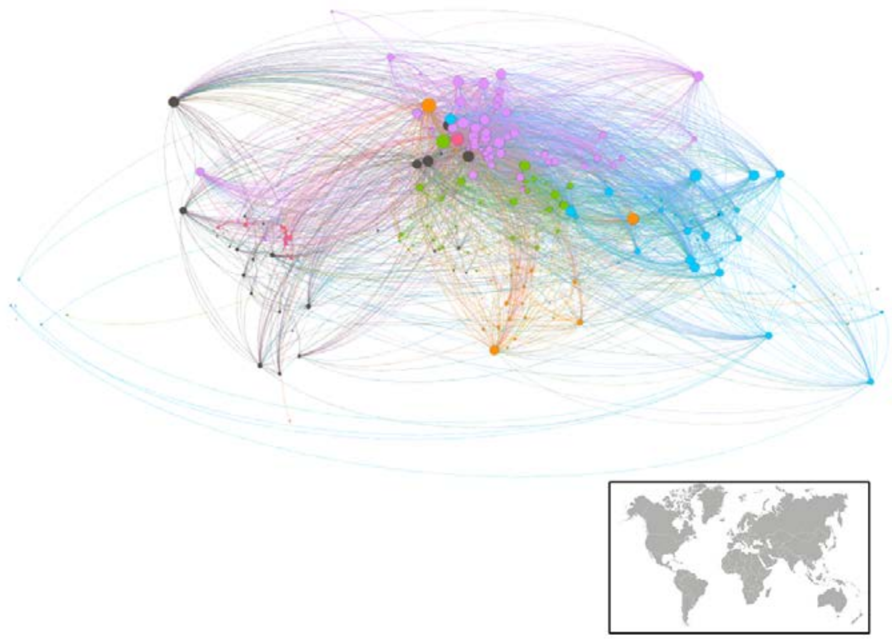


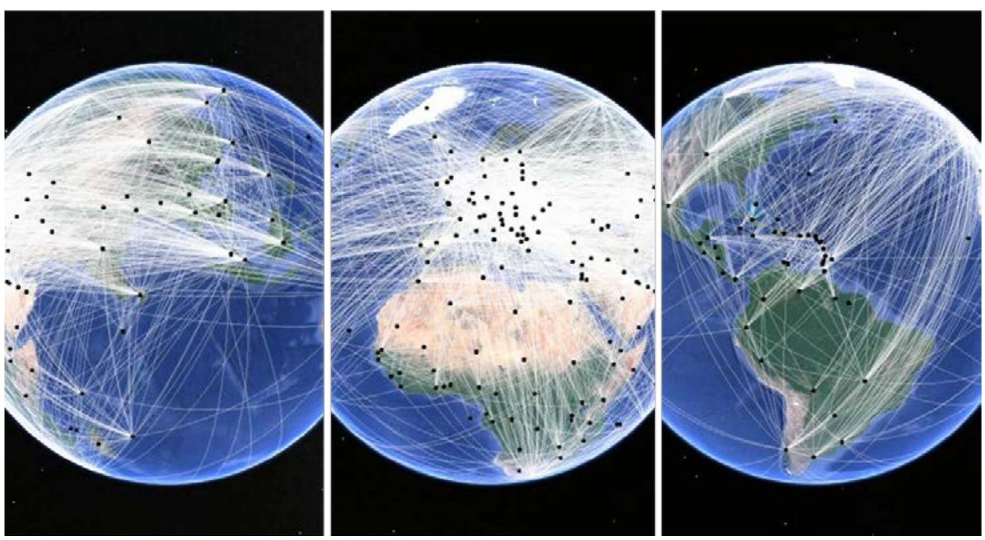
Source: Own elaboration with Gephi.

However, the network analysis tools and graphs allow us to clarify the picture. Graph 7 shows the same information spatially ordered (through longitude and latitude data of each country), dimensioning the nodes based on their Betweenness index and classifying countries by groups (clusters). For purposes of dissemination, the information can even be transposed in a three-dimensional form to Google Earth and even be presented dynamically.

11 https://www.ibfd.org/IBFD-Products/Tax-Treaties-Database

Graph 7. The international network of double taxation agreements with an organized presentation





Source: Own elaboration with Gephi and Google Earth.

Next, the quantitative indicators would allow us a more technical analysis. Based on the information provided by the mere existence of agreements (regardless of the rates for each of the passive income modalities -dividends; interests; royalties-), we get a non-directional graph without weighting factors. Table 3 shows the top twenty countries in terms of number of agreements and according to the Betweenness index value. In both cases, we find a majority of European countries and with the same three countries on top of the list (UK, France and Switzerland). However, the positions are altered depending on the sorting criteria. For example, Portugal excels in terms of mediation or bridge (4th) despite its smaller number of agreements (76). Similarly, Spain, Australia and South Africa improve their position by taking into account their potential role as a link, while others such as China go down significantly in the rankings.

Table 3. Analysis of the international network of double taxation agreements

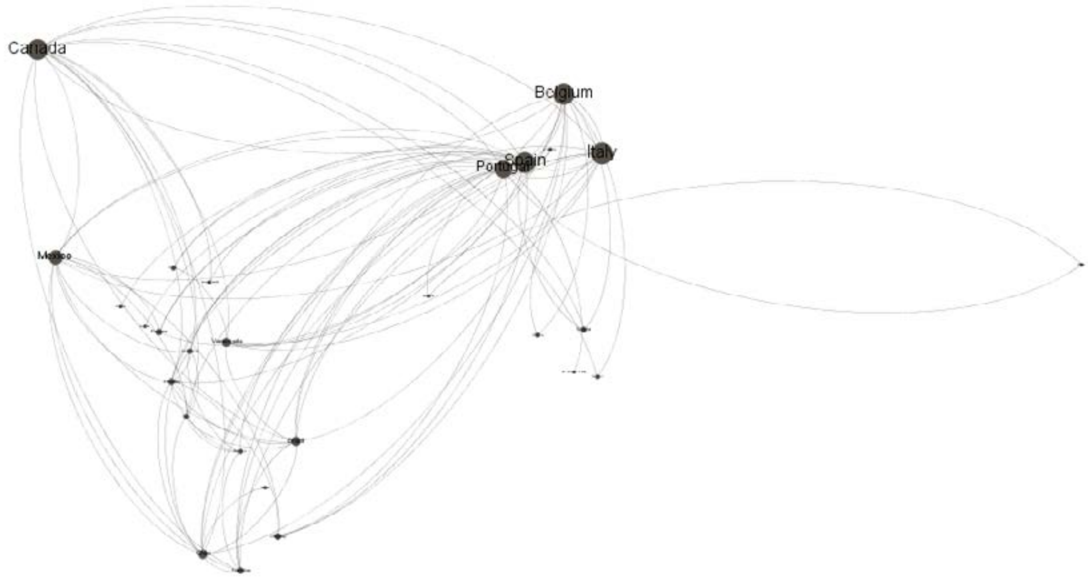
Countries	N. of DTT	Countries	Betweenness	Example Cluster	
United Kingdom	130	France	0,096537	Portugal	Nigeria
France	121	United Kingdom	0,090146	Spain	Paraguay
Switzerland	105	Switzerland	0,046105	Canada	Gabon
China (People's Rep,)	103	Portugal	0,030836	Belgium	Cape Verde
India	100	Spain	0,024188	Italy	Costa Rica
Italy	99	Norway	0,022729	Chile	Dominican Republic
Germany	96	India	0,022503	Mexico	S.Tome and Principe
Norway	95	Australia	0,020902	Brazil	El Salvador
Canada	94	South Africa	0,020666	Venezuela	
Sweden	94	Germany	0,020575	Ecuador	
Spain	93	United Arab Emirates	0,020328	Bolivia	
Netherlands	93	China (People's Rep,)	0,019311	Argentina	
Belgium	93	Canada	0,018979	Macau	
Korea (Rep,)	93	Sweden	0,014817	Colombia	
Austria	89	Netherlands	0,014239	Uruguay	
Romania	89	Belgium	0,01372	Peru	
Czech Republic	89	Singapore	0,012885	Andorra	
United Arab Emirates	87	Italy	0,012734	Panama	
Poland	87	Korea (Rep,)	0,011767	Cuba	
Finland	85	Denmark	0,011685	Ghana	

Source: Own elaboration with Gephi.

The clusters -formed, as mentioned before, according to the similarity of their network agreements and their (relative) distance from networks of other countries and identified by color in Graph 7 -show a remarkable geographical grouping pattern with some exceptions resulting from specific areas of influence of certain countries[12]. The last columns of the table show the countries included in one of the clusters, where 21 of the 28 countries are members of CIAT, with a majority of American countries, but also including, for example, Portugal, Spain, Italy or Nigeria, which are also CIAT members.

12 For example, Switzerland appears drawn in a group with most Caribbean countries. France joins much of the Arab countries and North Africa. United Kingdom is next to South Africa and India. Annex 1 lists the main *clusters* and their graphic representation.

Graph 8. Example of cluster graph



Source: Own elaboration with Gephi.

This first approach would help to present the data in an orderly fashion to identify, for example, possible ways to minimize taxation at international level. However, without taking into account at least the rates stipulated in each of the treaties, the existing withholding rates in the absence of agreement, and regional multilateral agreements of tax reduction, the vision would remain incomplete.

The incorporation of this information is not easy because the laws and treaties stipulate in most cases various possible rates depending on the circumstances (degree of business participation, the existence of prior levy, the source of royalties, type of debt that generates interests, etc.). For our simplified example, we stick to the taxation of dividends, assuming it applies the most favorable rate possible (including regional agreements of the European Union, CARICOM and the West African Economic and Monetary Union). We end up with a matrix of source and residence countries, identifying the most favorable routes in each case by comparison with overall withholding rates to eliminate non-relevant agreements (which prevail when being the lowest possible rates). Additionally, in all cases we add an additional minimum cost (0.001%), so the increase in steps for the dividend repatriation is at least slightly penalized, reflecting the additional costs of a presence in various jurisdictions[13].

Table 4 shows the classification obtained based on the Betweenness Index, which, as we have seen, reflects the possibility for each jurisdiction to be in the shortest path -in terms of taxation- between any two other countries.

13 This adjustment would be necessary, in any case, so the links where the levy is zero could be taken into account, because otherwise most standard programs for network analysis do not recognize the existence of these routes.

Table 4. Analysis of the international network of double taxation agreements with consideration of rates

(1-20)	(21-40)	(41-60)	(61-80)	(81-100)	(101-120)	(121-140)	(141-160)	(161-180)	(181-200)
United Kingdom	South Africa	Monaco	Mauritania	Ukraine	Macedonia (FYR)	Montenegro	Ethiopia	Syria	Grenada
France	Moldova	Romania	Ghana	Senegal	Azerbaijan	Montserrat	Kenya	Georgia	Guinea
India	Portugal	Czech Republic	Liechtenstein	Croatia	Zambia	Turkmenistan	Panama	Andorra	Iraq
Niger	Iran	San Marino	Yemen	Fiji	Oman	Argentina	Antigua and Barbuda	Anguilla	Kiribati
Singapore	Belgium	Chile	Mexico	Israel	Kazakhstan	Ecuador	Mozambique	Bermuda	Liberia
Vietnam	Netherlands	Luxembourg	Pakistan	Trinidad and Tobago	Peru	Albania	Togo	Bhutan	Malawi
Mauritius	Norway	Turkey	Macau	Saudi Arabia	Uzbekistan	Bosnia and Herzegovina	Laos	British Virgin Islands	New Caledonia
Malaysia	Italy	Denmark	Ireland	Guernsey	Algeria	St, Kitts and Nevis	Nigeria	Cambodia	Samoa
Hungary	Jordan	Morocco	Nepal	Isle of Man	Lesotho	Kyrgyzstan	Rwanda	Cameroon	Sierra Leone
Canada	New Zealand	Austria	Egypt	Jamaica	Barbados	Tajikistan	Gambia	Chad	Solomon Islands
Germany	Belize	Kosovo	Myanmar	Ivory Coast	Seychelles	Mongolia	Paraguay	Congo (Rep,)	St, Martin
Spain	St, Lucia	Tunisia	Slovak Republic	Slovenia	Philippines	Uruguay	Congo (Dem, Rep,)	Costa Rica	St, Pierre and Miquelon
Kuwait	St, Vincent and the Grenadines	Russia	Sudan	Serbia	Iceland	Bangladesh	Namibia	Dominican Republic	Suriname
Qatar	Korea (Rep,)	Korea (Dem, People's Rep,)	Belarus	Central African Republic	Lebanon	Bolivia	Swaziland	El Salvador	São Tomé and Príncipe
Brazil	China (People's Rep,)	Brunei	Thailand	Madagascar	Aruba	Guyana	Uganda	Equatorial Guinea	Tanzania
Malta	Sweden	Poland	Papua New Guinea	Greece	Curaçao	Benin	Dominica	Falkland Islands	Turkish Cyprus
Latvia	Hong Kong	Finland	Bulgaria	Netherlands Antilles	St, Maarten	GuineaBissau	Palestine	Faroe Islands	Tuvalu
Cyprus	Taiwan	Jersey	United States	Sri Lanka	Armenia	Colombia	TimorLeste	French Polynesia	French Guiana
Switzerland	Botswana	Libya	Japan	Cape Verde	Venezuela	Burkina Faso	United Arab Emirates	Gabon	Guadeloupe
Estonia	Indonesia	Cuba	Australia	Lithuania	Mali	Zimbabwe	Bahrain	Greenland	Martinique

Source: Own elaboration with Gephi

For example, the UK would have a value of 0.045 in its Betweenness Index, which would imply being present in 4.5% of the shortest paths -while without taking into account the conventions and general withholding rates, Table 3, appears in 9.6% of them- followed by France (3.4%) and India (2.2%).

Among the top twenty, seven of them (United Kingdom, France, India, Germany, Canada and Spain) already appeared in Table 3, for their large number of agreements. The other important factor to achieve this bridging role is to have a zero overall withholding rate -without DTA- for the repatriation of dividends (among the top twenty, the only exceptions are France, Canada, Germany, Spain and Switzerland).

The analysis could still be refined taking into account the regime of taxation of dividends received in each country and their rules for correction of double taxation[14], the lists of tax havens[15] , the taxation of other passive income -interests, royalties-[16] or the security and stability of each country, among other factors. In any case, through this first approach we hope to provide an example of the potential of the Network Analysis technique in the field of tax administration.

14 Barrios al (2009) provides a theoretical framework for its treatment. Petkova et al (2018), Van ´t Riet and Lejour (2014) apply, with variants, this approach.

15 Van ´t Riet and Lejour (2014).

16 Nakamoto and Ikeida (2018).

4. Other areas of application of Network Analysis in Tax Administrations

In its publication Advanced Analytics for Better Tax Administration: Putting Data to Work, OECD (2016) identifies the application of network analysis in several countries (Ireland, Malaysia, Netherlands, Singapore[17], New Zealand) for the prevention of the VAT carousel fraud and for detecting other types of fraud involving networks of economic operators. It warns, however, that it is not yet the most common practice. More specifically, the article selected for the number 44 (2018) of the CIAT / IEF / AEAT Tax Administration Review by Ignacio González García - Spain Tax Agency, AEAT- (Analytics and Big Data. The New Frontier) analyzes the basic technical characteristics of this approach compared to more traditional statistical techniques. This highly recommended reading provides concrete examples developed in the AEAT for the detection of extended family networks, calculation of corporate wealth from taxpayers taking into account indirect and cross-shareholdings, detection of corruption and laundering schemes and structures of fraud. Likewise, the DIAN in Colombia has been using these techniques for detecting patterns of fraud and money laundering in the field of gold trading.

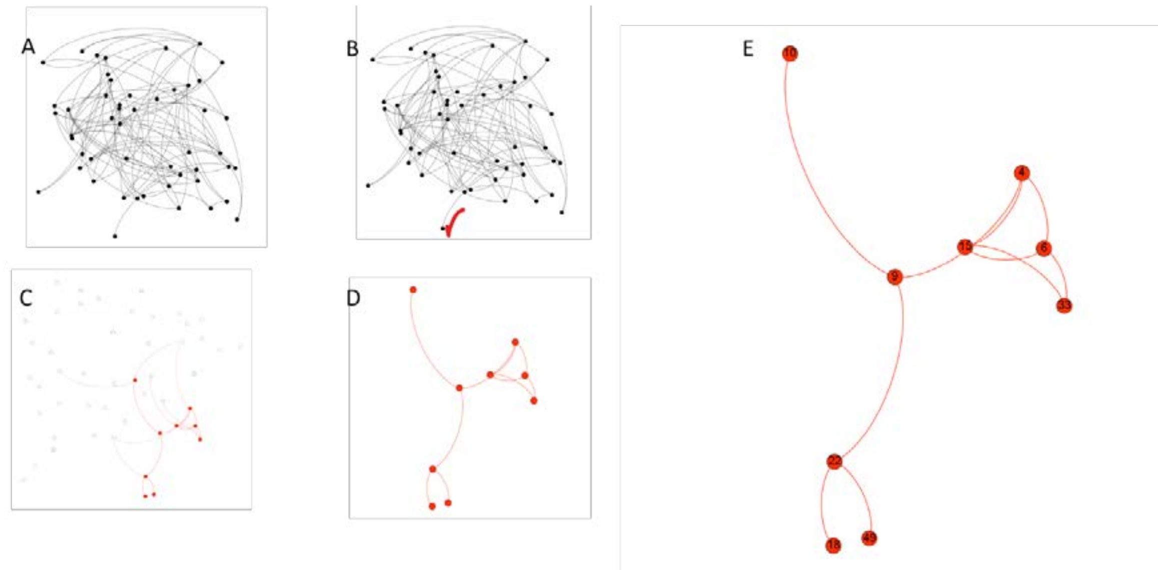
Given the magnitude of the information that Tax Administration[18] handle, the scope of network analysis is broad. The flexibility offered when organizing, visualizing and categorizing the information turns it into an optimal working tool even to address traditional problems. For example, if we have information about economic transactions on a network (purchases and sales) and we suspect that one of the companies has the characteristics of a possible issuer of false invoices or being involved in a carousel fraud, Network Analysis enables us to identify simply and quickly the leading companies connected for later tracking and analysis. Graph 8 shows a system based on a network of 50 nodes and 80 edges generated randomly (A), in which the suspect company is identified (B), the companies most connected to A (C and D) selected -using identifying clusters by modularity, i.e. communities within the network which are characterized by high internal density (many internal connections, and little connection with the rest[19])- and isolated from the other for their analysis by applying filters (9 nodes, 18% of the total, and 10 edges, 12.35%). In addition, Page Rank or HITS indices could be used to determine the importance and role of each node in the network.

17 OECD (2017), *The Changing Tax Compliance Environment and the Role of Audit*, highlights the case of Singapore for the use of network analysis for risk modeling and selection of audits.

18 The introduction of electronic invoicing and other information collection systems exponentially increases the possibilities of analysis.

19 Technically, the method proposed by Blondel et al (2008) is highly efficient in managing large databases and adjustable in its demand for internal cohesion of each community.

Graph 9. Example detection of potentially fraudulent activities



Source: Own elaboration with Gephi.

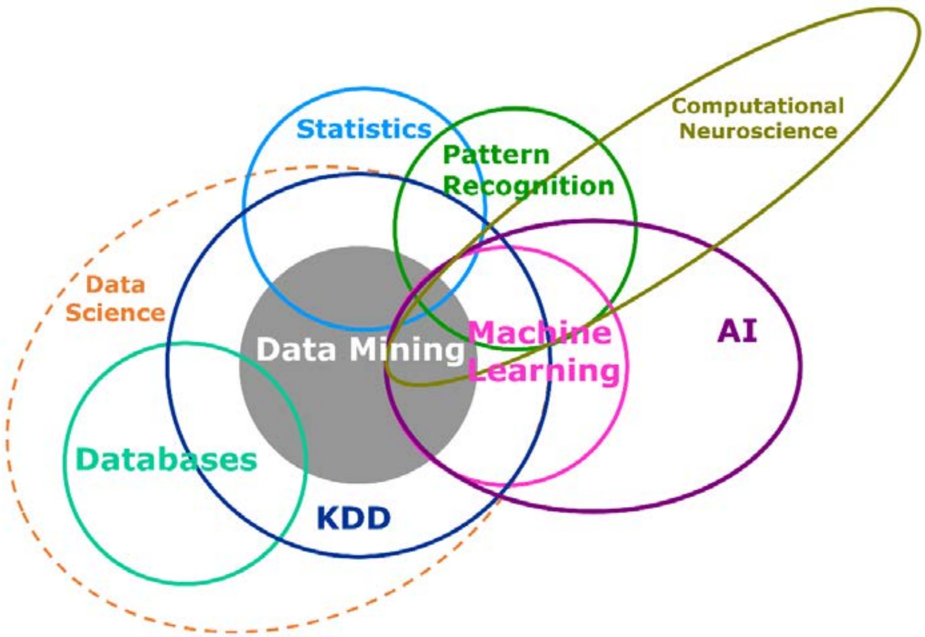
The possibilities for deeper analysis are numerous: attributing additional features through labels for filtering and selection by seniority, previous inspections, geographical location, etc., differentiating between different types of nodes (multipartite graphs) such as, for example, individuals, corporations, holding companies, etc., or by analyzing changes in the time relations (dynamic graph through time-stamps). Whenever the frauds involve building relationships and networks[20], these tools will be useful to organize and analyze information[21].

Finally, it is worth noting that this is an aseptic technique, we could say, in the sense that it does not presuppose a theoretical modeling of fraud, even if it depends on the expert knowledge to guide their use and interpretation. The systems analyst needs to work closely with the tax auditor to organize the information, set search criteria and select the most appropriate indicators. This monitoring can be supplemented by other data analysis techniques, grouped in various fields and denominations that overlap each other: machine learning; artificial intelligence; data mining; Knowledge Discovery in Databases (KDD); etc.

20 Something that seems to be common in many real situations.

21 The book Baesens et al (2015) "Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques" introduces in detail the use of these techniques from the statistical standpoint, including references to other specific applications of fraud detection, for example in the field of Social Security - "GOTCHA! Network-Based Fraud Detection for Social Security Fraud ", published in Van Vlasselaer et al (2017) -. SAS has developed an application called Social Network Analysis for analysis and detection of insurance fraud by applying these techniques (https://www.sas.com/content/dam/SAS/en_us/doc/productbrief/sas- social-network-analysis-103857.pdf).

Graph 10. The multidisciplinary nature and the fuzzy boundaries between data analysis techniques



Source: Hall et al (2014)

These techniques will allow amplifying the expert guidance of the auditor, tracking down existing signals in complex databases regarding fraudulent behavior through statistical techniques and recursive methods for optimization, driven by the enormous capacity and computational speed of current computer equipment (a classic example is the use of so -called neural networks to detect fraud patterns, just as they are used for image or voice recognition)[22] .

22 Ultimately and due to the rise of artificial intelligence, some suggest that it could even lead to the detection of fraudulent behavior without expert supervision ("unsupervised *machine learning* "), thus avoiding bias and the limitations of the models that are based on patterns of fraud previously known. In my opinion, however, it is not useful to be bogged down in arguments about radical approaches such as confrontation between the new "artificial intelligence" and "classical tax intelligence" of experts and auditors. We will always have to face externally determined objectives and search parameters, to validate the results in terms of actual legislation and present evidence before the courts. Objectives, legislation and judges, yet very human.

Bibliographic references

Baesens, B., Van Vlasselaer, V., (2015). Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques .

Barrios, S., Huizinga, H., Laeven, L. and NICODEM, G (2009) "International Taxation and Multinational Firm Location Decisions". Economic Papers 356 | January 2009 European Commission

Blondel, V., Guillaume JL, Lambiotte, R. and Lefebvre, E. (2008) "Fast unfolding of communities in large networks", Journal of Statistical Mechanics: Theory and Experiment. IOP and SISSA journal

González García, I., (2018) "Analytics and Big Data. New Frontier "Tax Administration Review CIAT / IEF / AEAT, number 44, 2018.

Hall, P., Dean, J., Kaynar Kabul, I. Silva. J (2014), "An Overview of Machine Learning with SAS® Enterprise Miner ™". Paper SAS313-2014 . SAS Institute Inc.

Nakamoto, T., Ikeda, Y. (2018) "Identification of Conduit Countries and Community Structures in the Withholding Tax Networks" https://arxiv.org/abs/1806.00799

OECD (2016), Advanced Analytics for Better Tax Administration: Putting Data to Work , OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264256453-en

OECD (2017), The Changing Tax Compliance Environment and the Role of Audit , OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264282186-en

Petkova, K., Stasio, A., Zagler, M. (2018) "On the relevance of double tax Treaties", WU International Taxation Research Paper Series, No. 2018-05

Van't Riet, M., Lejour, A (2014) "Ranking the Stars. Network Analysis of Bilateral Tax Treaties "CPB Discussion Paper | 290.

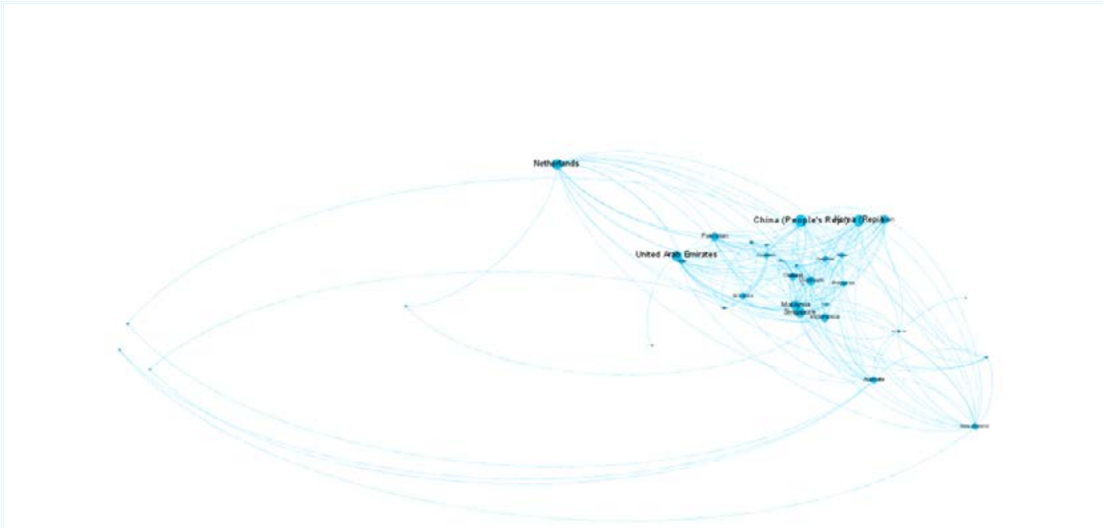
Van Vlasselaer, V., Eliassi-Rad, T., Akoglu, L., Snoeck, M., Baesens, B., (2017), "GOTCHA! Network-Based Fraud Detection for Social Security Fraud ", Management Science Volume 63, Issue 9, September 2017

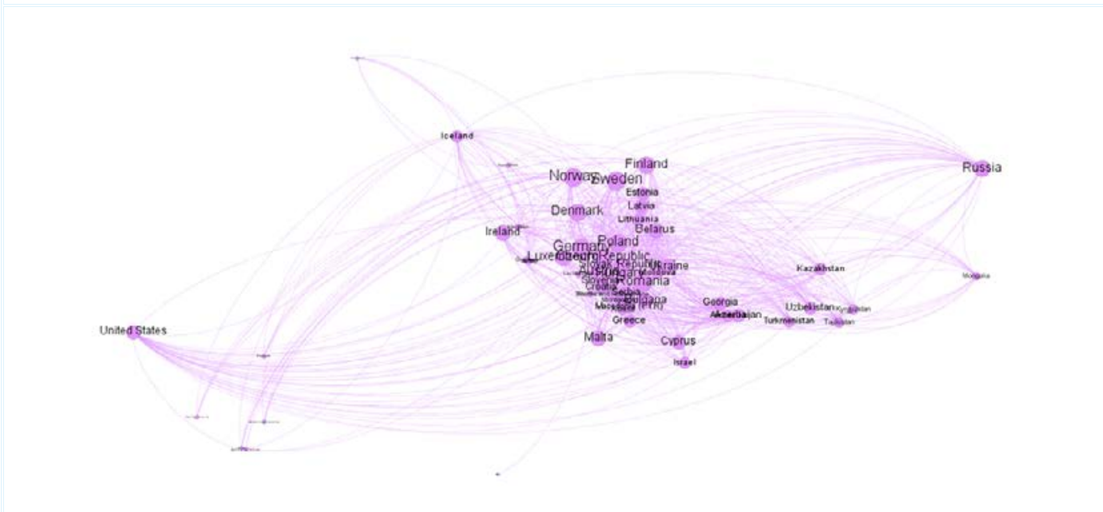
Wujek, B., Hall, P. and Güneş, F. (2016), "Best Practices for Machine Learning Applications" Paper SAS2360-2016. SAS Institute Inc.

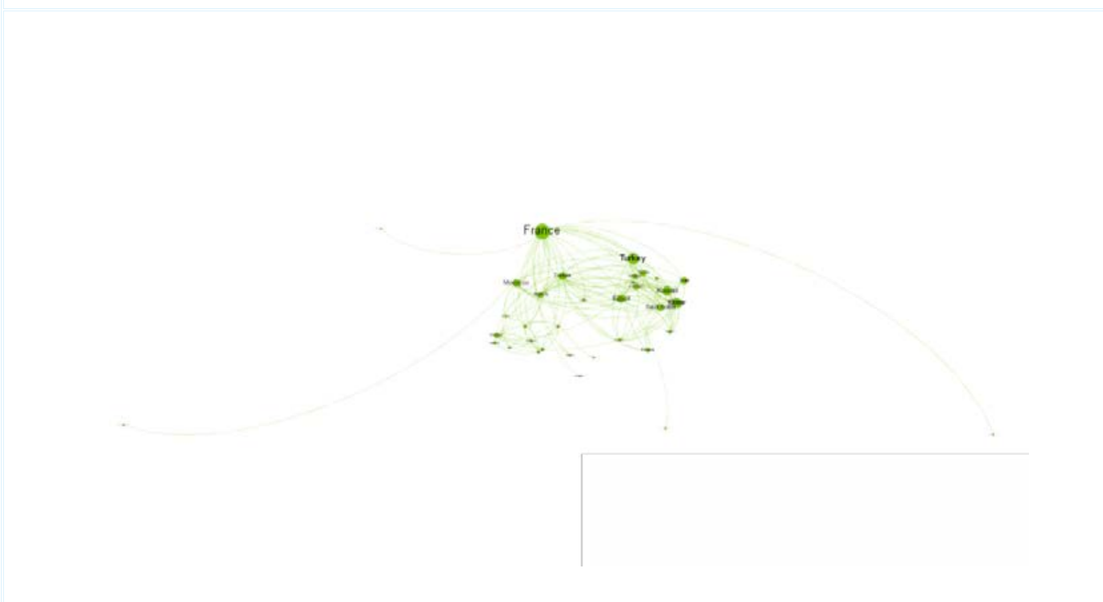
Annex 1. Non-directional Clusters of Double Taxation Conventions graphs

Cluster 1: 26.27% of the countries			Cluster 2: 16.59%		Cluster 3: 15.67%	
Albania	Greece	Mongolia	Algeria	Mali	Australia	Netherlands
Armenia	Greenland	Montenegro	Bahrain	Mauritania	Bangladesh	New Zealand
Aruba	Guernsey	Netherlands Antilles	Benin	Morocco	Bhutan	Oman
Austria	Hungary	Norway	Burkina Faso	New Caledonia	Brunei	Pakistan
Azerbaijan	Iceland	Poland	Cameroon	Niger	Cambodia	Papua New Guinea
Belarus	Ireland	Romania	Central African Republic	Palestine	China (People's Rep,)	Philippines
Bermuda	Isle of Man	Russia	Congo (Rep,)	Qatar	Comoros Islands	Samoa
Bosnia and Herzegovina	Israel	San Marino	Egypt	Saudi Arabia	Cook Islands	Singapore
British Virgin Islands	Jersey	Serbia	Ethiopia	Senegal	Fiji	Sri Lanka
Bulgaria	Kazakhstan	Slovak Republic	France	St, Pierre and Miquelon	Hong Kong	Suriname
Cayman Islands	Kosovo	Slovenia	French Polynesia	Sudan	Indonesia	Taiwan
Croatia	Kyrgyzstan	Sweden	Guinea	Syria	Japan	Thailand
Cyprus	Latvia	Tajikistan	Guinea-Bissau	Togo	Kiribati	United Arab Emirates
Czech Republic	Liberia	Turkmenistan	Iran	Tunisia	Korea (Rep,)	Vietnam
Denmark	Liechtenstein	Ukraine	Iraq	Turkey	Laos	
Estonia	Lithuania	United States	Jordan	Yemen	Malaysia	
Faroe Islands	Luxembourg	Uzbekistan	Kuwait		Maldives	
Finland	Macedonia (FYR)		Lebanon		Marshall Islands	
Georgia	Malta		Libya		Myanmar	
Germany	Moldova		Madagascar		Nepal	

Clusters 1, 2 and 3

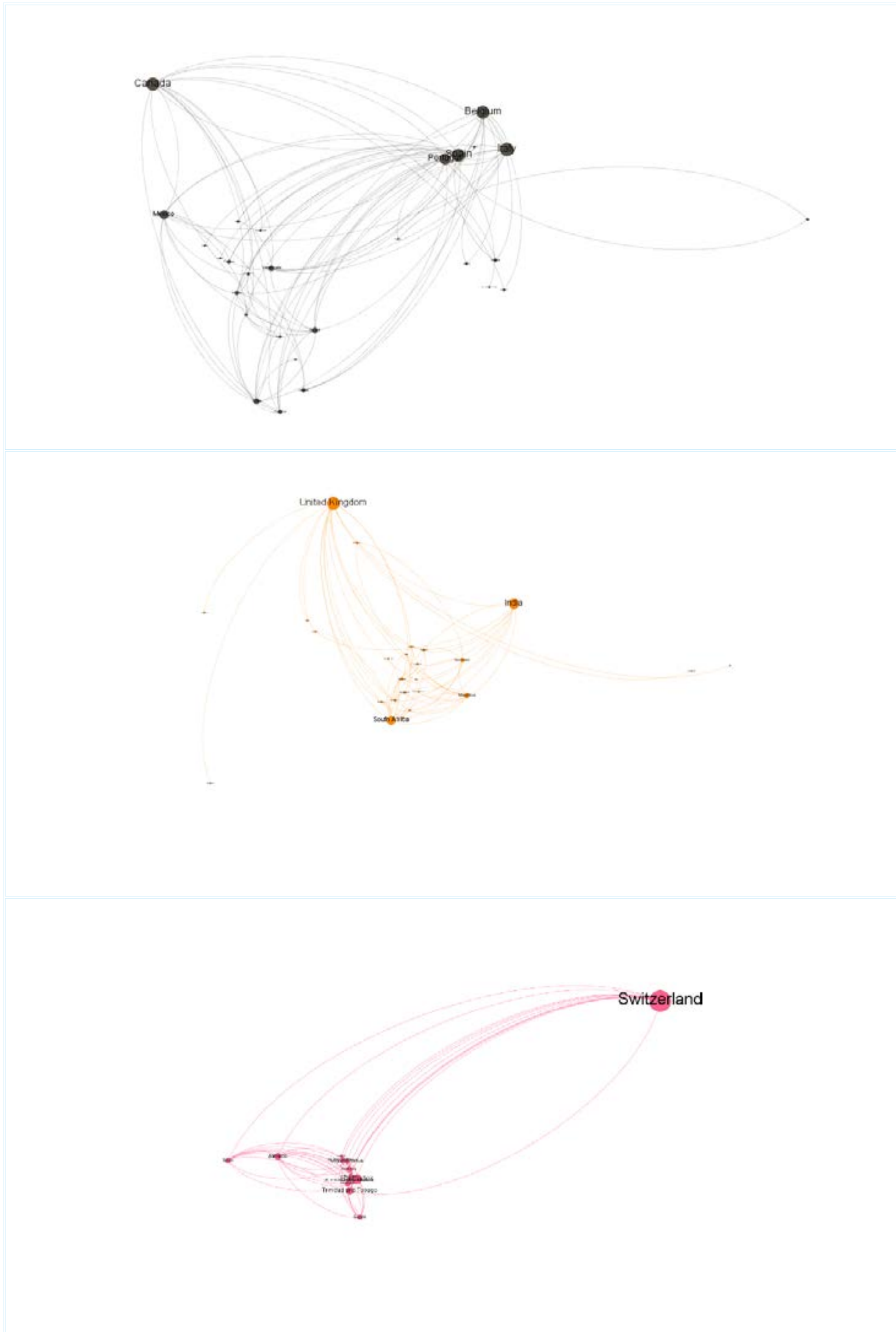






Cluster 4: 12,9%		Cluster 5: 11,52%		Cluster 6: 6%
Andorra	Gabon	Botswana	Seychelles	Anguilla
Argentina	Ghana	Congo (Dem, Rep,)	Sierra Leone	Antigua and Barbuda
Belgium	Italy	Falkland Islands	Solomon Islands	Barbados
Bolivia	Macau	Gambia	South Africa	Belize
Brazil	Mexico	India	Swaziland	Dominica
Canada	Nigeria	Kenya	Tanzania	Grenada
Cape Verde	Panama	Lesotho	Tuvalu	Guyana
Chile	Paraguay	Malawi	Uganda	Jamaica
Colombia	Peru	Mauritius	United Kingdom	St, Kitts and Nevis
Costa Rica	Portugal	Monaco	Zambia	St, Lucia
Cuba	Sao Tomao and Principe	Montserrat	Zimbabwe	St, Vincent and the Grenadines
Dominican Republic	Spain	Mozambique		Switzerland
Ecuador	Uruguay	Namibia		Trinidad and Tobago
El Salvador	Venezuela	Rwanda		

Clusters 4, 5 and 6



Working Papers *Series*

CIAT Executive Secretariat
P.O Box: 0834-02129 , Panama City, Republic of Panama
Phone: (+507) 307 CIAT (2428)
Fax: (507) 2644926
e-mail: ciat@ciat.org
Website: www.ciat.org