Università Commerciale Luigi Bocconi

Department of Decision Sciences
Bachelor of Science in Economics, Management and Computer Science

Construction and Analysis of Transnational Illicit Drug Trafficking Networks

Supervised by:
Prof. Omiros Papaspiliopoulos
Prof. Daniele Durante

Matei Gabriel Coșa
Student ID: 3159774

Academic Year 2023/2024

Acknowledgements

I would like to thank my supervisors, prof. Omiros Papaspiliopoulos and prof. Daniele Durante for their continuous support and guidance, from the beginning of my work in the Bocconi Institute for Data Science and Analytics (BIDSA), to the completion of my Bachelor Thesis. I would also like to express my gratitude towards prof. Luca Giommoni (Cardiff Univeristy) for providing me with access to the International Drug Seizures dataset, the backbone of our empirical analysis, as well as with important advice for improving upon current methodologies. Similarly, prof. Alberto Aziani and the members of the Transcrime research group have provided valuable insights for completing this work. Finally, I am thankful to all my friends, colleagues, and family that supported me throughout the entire degree program.

Contents

Chapter 1

Introduction

Transnational trafficking of illegal substances is an ever-growing phenomenon that poses significant challenges to policymakers, law enforcement agencies, and researchers alike. According to the United Nations Office for Drugs and Crime (UNODC), one in every seventeen people worldwide consumed drugs in 2021, and cocaine alone reached twenty two million consumers globally [19].

While many drugs such cannabis or methamphetamine can be produced in virtually any country, cocaine and heroin are only sourced from specific geographic regions and have a limited number of producers [7]. For this reason, mapping the routes taken by these drugs from their producers to the final consumption markets is essential for gaining insight into the main factors that shape cross-border flows. Consequently, such information can be used for improving targeted law enforcement policies that aim to disrupt the global network [9].

1.1 Objective

Our first goal is to build upon the existing methodology for constructing international drug trafficking networks from empirical data. This requires the use of statistical and mathematical tools that can be of great importance in the field of criminology and quantitative social science. In particular, we will show how spectral analysis and the theory of non-negative matrices relates to the process of estimating actual drug flows between countries from observed seizure data. Our main contribution revolves around connecting the topological structure of the underlying graph to the uniqueness, stability and properties of the estimated flows. Furthermore, we provide a novel methodology that is capable of producing numerically stable solutions with desirable properties without the need of relying on large amounts of sometimes unreliable and unavailable data.

Our second objective is centred around using this approach to construct the international cocaine trafficking network from data covering the period between 2006 and 2017. This allows us to empirically validate our model and analyze the resulting network to identify its structural properties in order to determine the most important countries and trafficking patterns.

1.2 Previous Work

The construction and analysis of transnational drug trafficking networks presents itself as a problem well suited for the techniques of social network science [6]. This branch of statistics and data science provides several tools for understanding the underlying factors behind the formation of complex networks, as well as models for studying their structural features to extract meaningful insights [3].

The current state-of-the-art methodology for constructing transnational drug trafficking networks is based on [17]. This approach has been implemented in several studies [5], [2], [1], [8] as a foundational step for setting up further models relating to aspects such as link formation or interdiction strategies.

A key aspect of the network construction process is related to flow estimation, which essentially provides weights for each edge in the network. Without the weights, important information about the network and its properties cannot be analyzed [12]. This is why we will devote a significant portion of our work to the flow estimation procedure.

Chapter 2

Network Construction

2.1 Methodology & Data

We are interested in building networks of type $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$, where $\mathcal{V} = \{1, 2, \ldots, n\}$ is the set of nodes (countries) and $\mathcal{E}_t$ is the set of edges present in the trafficking network at time t. For us, $t = \{2006, 2007, \ldots 2017\}$, since this is the period for which we have access to all the required data for the construction process.

We follow the general approach of [17] which includes several steps summarized below[1]:

1. **Identifying the connections**: this step relies on data from the IDS dataset [1], as well as the average purity by country [4]. The IDS dataset contains individual seizure cases reported by UNODC member states. For each seizure case, information regarding the type of drug, quantity, hiding place, and several other characteristics are recorded. For some of the entries, information about the place of seizure, as well as the source country and/or future destination countries are provided. This allows us to build a matrix $S_t \in \mathbb{R}^{n \times n}$ such that $(S_t)_{i,j}$ represents the total quantity seized at time t along the directed edge from i to j. As per the standard practices in the literature [13], [17], we use conversion rates [18] in order to transform every quantity in kilograms of pure cocaine, according to the average purity at wholesale level in the country of seizure. Finally, we let $S_{i,j} = \frac{1}{12} \sum_{t=2006}^{2017} (S_t)_{i,j} \ \forall i, j \in \mathcal{V}$.

[1]Whenever we encountered missing data, we performed imputation through linear interpolation and extrapolation across time, combined with sub-regional or regional averages, as [9] did in their study.

We do this in order to overcome potential biases induced by the variability in the number and accuracy of reported seizures across the twelve year period.

2. **Estimating the national markets**: this is done using data from several datasets relating to population [2], drug prevalence [3], and global production [18]. For a country $i \in \mathcal{V}$, let $\lambda_{i,t}$ be the national prevalence of cocaine, and $p_{i,t}$ be the population in the age category 15-64. By dividing the total production of cocaine by the estimated global number of users, we obtain the average quantity consumed by a user in year t, denoted by $\bar{c}_t$. Using this quantity, we can estimate the average annual consumption as $C_{i,t} = \lambda_{i,t} p_{i,t} \bar{c}_t$. Then we can write the total market of country i at time t as $b_{i,t} = C_{i,t} + \sum_j (S_t)_{j,i}$, meaning the sum between the total consumption and the seized quantity. An additional complication stems from the fact that some countries also produce cocaine domestically. To address this, we will subtract the estimated annual production from the national market for the largest producers of cocaine: Colombia, the Plurinational State of Bolivia, and Peru. Analogously to the the previous step, we average across the entire time period to obtain $b \in \mathbb{R}^n$, the vector of averaged national markets.

3. **Estimating the flows**: this final stage requires combining all the previous estimates. Since seizures only convey information about drugs that have been captured and reported by law enforcement, we would like to retrieve some credible latent structure of flows that could have potentially generated the observed seizures. To do this, we make the strong assumption that the relative weights of the seized quantities fully represent the true, underlying relative weights of the flows. Let $F \in \mathbb{R}^{n \times n}$ be the matrix of flows, i.e. $F_{i,j}$ is the total quantity flowing from i to j. For a country i, we can thus write the flows as:

$$\sum_{j=1}^{n} F_{j,i} = b_i + \sum_{j=1}^{n} F_{i,j} \tag{2.1}$$

The expression 2.1 has a simple interpretation: the market of a country is the difference between imports and exports. We denote with $I_i = \sum_{j=1}^{n} F_{j,i}$ the total

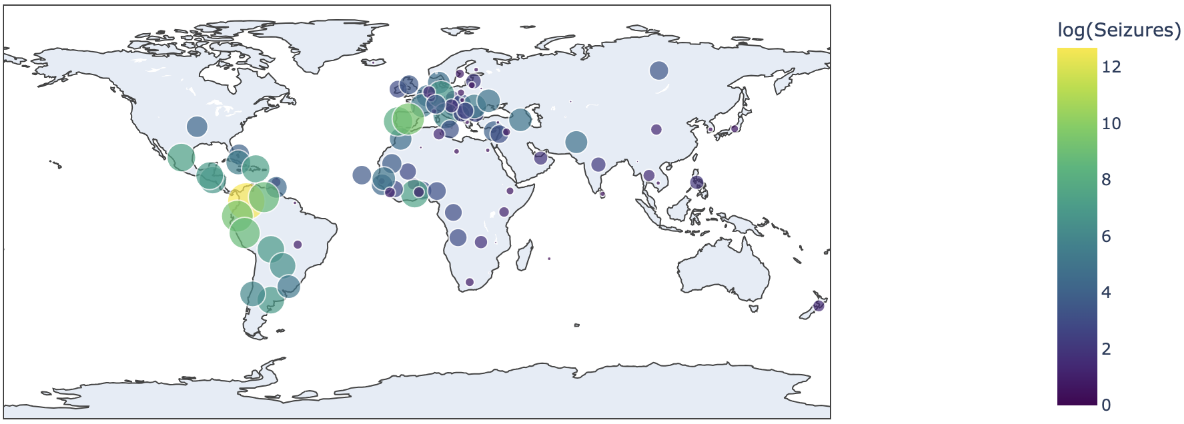

Figure 2.1: Global Seizures of Cocaine (2006-2017)

imports of country i. A key observation is that, under our assumption, the exports can be written in terms of the imports using the relative weights induced by the seizures:

$$F_{i,j} = \frac{S_{i,j}}{\sum_{k=1}^{n} S_{k,j}} I_j \tag{2.2}$$

Clearly, an analogous expression allows us to write the exports in terms of the imports and the corresponding relative weights. These observations lead to a much more practical and elegant formulation of 2.1 that will be explored extensively in Chapter 3.

2.2 Limitations and Possible Improvements

As Giommoni et al.[9] note, this methodology is not without its flaws. Basing the network reconstruction process on seizure data requires one to rely on the competence and transparency of law enforcement agencies and governments with heterogeneous policies, mandates, and resources [14]. Moreover, as we hinted in section 2.1, assuming that seizures accurately reflect the relative weights of the true flows is a very strong leap of faith. Figures 2.1 and 2.2 showcase the difference in magnitude between the seizures and the estimated markets (which include both consumption and seizures). Most notably, there are several countries with small seized quantities, but sizeable internal consump-

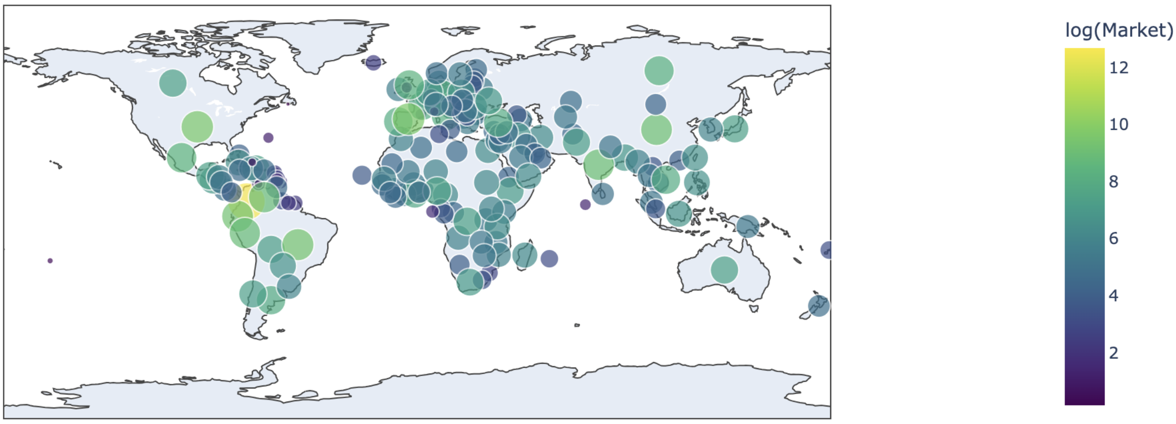

Figure 2.2: Global Markets of Cocaine (2006-2017)

tion. Estimating the real flows based on relative weights from the seizures can thus be problematic for these states.

One area in which we think improvements can be made is step 2, which makes a series of potentially oversimplifying assumptions. The major point of concern is the calculation of the average annual consumption per user, which neglects country-specific behaviours, as well as differences between light and heavy users.

Relating to the combination of supply and demand-side estimates, there is also a potential time lag between production and consumption, which makes such approaches problematic [11]. Even when the values obtained for the national markets appear reasonable in relative terms, their magnitude may not be accurate. Furthermore, errors may have a significant impact on the existence and quality of the solutions obtained from step 3. For this reason, we will introduce an approach that aims to circumvent the necessity of estimating the national markets in section 3.6.

Regarding step 3, while very clear at an intuitive level, the system formulated as above poses several mathematical questions that should be treated with care. Our goal in chapter 3 will be precisely to identify these challenges and propose theoretical solutions that guarantee desirable outcomes in the empirical setting.

2.3 Network Statistics

We will focus on exploring the topological structure of our seizure graph constructed in step 1. The seizure graph contains 164 nodes and 1098 edges. At first glance, we wish to control for possible anomalies stemming from missing or unreliable data. One such problem is related to the existence of non-producing countries that do not have any incoming edges. This clearly results from the lack of data and can pose several problems in the estimation process. The natural course of action is to identify the strongly connected components of our seizure graph.

Interestingly, we observe a large strongly connected component of 111 nodes and 53 components containing only one node. The main connected component highlights the centralization of cocaine production in a few countries in South America, from which the entire network satisfies its demand. If there were several producing areas across the globe, we would have potentially seen multiple large connected components. Out of the 53 single-node components, 40 of them have no outgoing edges, which means they are countries that do not export and serve as final destinations for cocaine. Such countries are referred to as consumers in the literature [10]. The remaining 13 nodes/components (Guyana, Cook Islands, Sri Lanka, Bermuda, Tunisia, Mauritania, Antigua and Barbuda, French Guiana, Grenada, Martinique, Saint Pierre and Miquelon, Papua New Guinea, Bosnia and Herzegovina) lack incoming edges, which means they fall into the problematic category mentioned above, so we decide to remove them.

We are left with a sparse graph of 151 nodes and 1081 edges. The average node degree is close to 7, indicating that with the exception of large transit countries [10] and hubs, most nodes do not import or export to many others. Regarding the local connectivity of the network, the average clustering coefficient is 0.41. This suggests that for a random node, the probability that two of its neighbors are connected is close to 40% [3]. We can thus expect a reasonable number of triangle structures.

To better understand the relevance of each node within the network, we turn to measures of centrality, in particular: in-degree, out-degree, and betweenness centrality. Spain is by far the single most important country within the network in terms of centrality. Not

Betweenness	**In-degree**	**Out-degree**
Spain: 0.35	Spain: 0.46	Ecuador: 0.6
Ecuador: 0.11	Italy: 0.3	Spain: 0.58
Italy: 0.07	Portugal: 0.25	Venezuela, Bolivarian Republic of: 0.51
Portugal: 0.07	Netherlands: 0.19	Germany: 0.33
France: 0.06	Austria: 0.16	France: 0.31
Germany: 0.04	United Kingdom: 0.15	Portugal: 0.27
Belgium: 0.04	Russian Federation: 0.15	Italy: 0.26
Colombia: 0.03	Bulgaria: 0.15	Argentina: 0.22
Turkey: 0.03	Nigeria: 0.15	Nigeria: 0.21
Australia: 0.03	France: 0.15	Belgium: 0.21

Table 2.1: Top 10 Countries by Centrality Scores

only is Spain located along the most paths between two other nodes, but it also has a very large number of incoming and outgoing edges, making it a central hub within the network. Ecuador is another country which plays a significant role in the distribution of cocaine from South America to the rest of the world. European countries also have high centrality scores, which is probably due to the large markets and high movement of people and goods within the EU. Bulgaria, Russia, Nigeria, and Turkey are also located along important trafficking routes, as noted by [9]. Finally, Colombia, the world's largest producer of cocaine, as well as its neighbors, naturally position themselves as large exporters.

Figure 2.3 presents the in-degrees plotted against the out-degrees and colored according to the market sizes estimated in step 2. This plot is interesting since it points out several features of the network. Firstly, the size of the market is not directly related to the number of incident edges, as illustrated by Colombia. This country also has a surprisingly large number of incoming edges in spite of being the largest producer. This highlights another crucial point: total imports and exports of a country are not directly related to the number of incident edges. In fact, after we estimate the actual flows in section 3.7, we will observe that Colombia's imports are negligible. The main takeaway is that although the topology of the network provides several insights, it is not enough to paint a clear picture without the estimated flows.

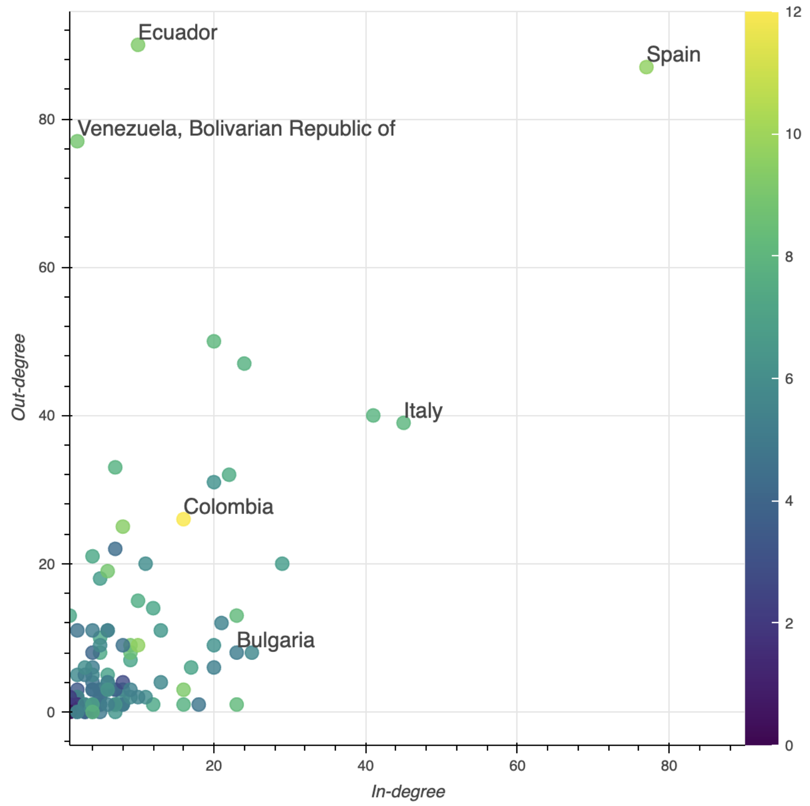

Figure 2.3: In-degree & Out-degree Colored According to log(Market Size)

Chapter 3

Flow Estimation

In this chapter we seek to outline the mathematical framework of the flow estimation problem, propose a novel approach based on theoretical findings, and apply it to the data from section 2.1. We start by presenting some fundamental elements of graph theory, outlining the the problem setting and highlighting some of its properties.

3.1 Preliminaries

The underlying mathematical object that represents a network is a graph, i.e. a pair $\mathcal{G} := (\mathcal{V}, \mathcal{E})$. $\mathcal{V}$ is a set of nodes and $\mathcal{E}$ is a set of edges, meaning $(u, v) \in \mathcal{E}$ if there exists an edge between nodes u and v, where $u, v \in \mathcal{V}$. When $(u, v) \in \mathcal{E}$ does not imply $(v, u) \in \mathcal{E}$, the graph is said to be directed, otherwise it is undirected.

A path in $\mathcal{G}$ is a sequence of nodes $u_1, u_2, \ldots, u_k \in \mathcal{V}$ such that $(u_i, u_{i+1}) \in \mathcal{E}$ for all $i < k$. A connected component is a set C such that for every $u, v \in C$, there exists a path from u to v. In the case of directed graphs, the paths are composed of directed edges and C is called a strongly connected component. If $C = \mathcal{V}$, then we say $\mathcal{G}$ is connected or strongly connected, depending of the type of the graph.

To represent the graph $\mathcal{G}$, we define a matrix $\mathbf{A}_{\mathcal{G}} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$, where $A_{i,j} = 1$ if $(i, j) \in \mathcal{E}$ and $A_{i,j} = 0$ otherwise. An edge from a node to itself is called a self-loop and we will assume $\mathcal{G}$ does not contain such edges, meaning $A_{i,i} = 0$ for all $i \in \mathcal{V}$. $\mathbf{A}_{\mathcal{G}}$ is called the adjacency matrix of $\mathcal{G}$ and it is symmetric whenever $\mathcal{G}$ is directed.

3.2 Problem Setting

We consider a set of nodes $\mathcal{V} := \{1, 2, \ldots, n\}$ and a matrix of seizures $\mathbf{S} \in \mathbb{R}^{n \times n}$, such that $S_{i,j}$ represents the quantity of drugs seized while traveling from country i to country j, with $S_{i,i} = 0 \; \forall i \in \mathcal{V}$. We further define $\mathcal{E} := \{(i, j) \in \mathcal{V}^2 : S_{i,j} > 0\}$ to be the set of pairs of nodes such that i exports to j. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the graph induced by the matrix of seizures $\mathbf{S}$. Since $\mathbf{S}$ is not necessarily symmetric, $\mathcal{G}$ is a directed graph with no self-loops. In addition to the topological structure defined above, we want to endow $\mathcal{G}$ with weights representing the relative proportion of inflows/outflows with respect to the total flows of a given node. To this end, let $\mathbf{W} \in \mathbb{R}^{n \times n}$ be defined as:

$$W_{i,j} := \frac{S_{i,j}}{\sum_{k=1}^{n} S_{i,k}} \quad \forall i, j \in \mathcal{V} \tag{3.1}$$

Furthermore, let $\widetilde{\mathbf{W}} \in \mathbb{R}^{n \times n}$ be defined as:

$$\widetilde{W}_{i,j} := \frac{S_{i,j}}{\sum_{k=1}^{n} S_{k,j}} \quad \forall i, j \in \mathcal{V} \tag{3.2}$$

$\mathbf{W}$ can be interpreted as containing the relative weight of each outflow from a node with respect to the total quantity exported by that node. Similarly, $\widetilde{\mathbf{W}}$ contains the relative weight of each inflow into a node with respect to the total quantity imported by that node. A point of care has to do with the denominators in expressions 3.1 and 3.2. If a node has zero imports or exports, then the fraction is not well-defined. To avoid this issue, by definition, we set the entries were this occurs equal to 0.

We now consider a vector $\mathbf{b} \in \mathbb{R}^n$ representing the market volume of each country present in the network. The elements of $\mathbf{b}$ may be negative if internal production outweighs consumption and seizures. The literature proposes several techniques for estimating $\mathbf{b}$ by combining supply, consumption, and seizure data, as outlined in step 2 of Section 2.1. Due to the lack of granular, trustworthy and publicly available data, these estimates may not be reliable, as explained in Section 2.2. In light of this problem, we will first treat $\mathbf{b}$ as known and then propose a method for retrieving the flows without going through the process of market estimation referenced above.

Using this general framework, we introduce two problems: determining the total exports and imports of every node in the network based on the seizures and market data. Let us present the following systems of linear equations:

$$(\mathbf{W}^{\mathbf{T}} - \mathbf{I})\mathbf{x} = \mathbf{b} \tag{3.3}$$

$$(\mathbf{I} - \widetilde{\mathbf{W}})\widetilde{\mathbf{x}} = \mathbf{b} \tag{3.4}$$

where $\mathbf{x}, \widetilde{\mathbf{x}}$ represent the estimated total exports and imports, respectively. Intuitively, the two problems should be directly related, since every outflow can be viewed as an inflow in the opposite direction. In fact, we can formulate the explicit relation between $\mathbf{x}$ and $\widetilde{\mathbf{x}}$, assuming both quantities are well-defined:

$$\widetilde{\mathbf{x}} = \mathbf{W}^{\mathbf{T}}\mathbf{x} \tag{3.5}$$

Equation 3.5 can be interpreted as saying that the imports of node i are equal to the sum of exports from all countries to node i. Using 2.2, it is easy to see that the two systems are in fact equivalent to our initial system of flows outlined in Section 2.1, but it is important to stress that, a priori, there exists no mathematical guarantee for the two systems to simultaneously have unique solutions, and for the solutions to be related in the way described above. In fact, one can observe that 3.3 and 3.4 may not admit unique solutions. We will prove this, as well as other interesting properties of the setting described above in the following section. Moreover, from the point of view of our application, we would like to obtain solutions with non-negative components, i.e. $\mathbf{x}, \widetilde{\mathbf{x}} \geq \mathbf{0}$. This is because negative imports or exports have no clear interpretation and may guide us to misleading results.

3.3 Properties

We begin by stating and proving a series of claims about the properties of 3.3 and 3.4, as well the implications of imposing restriction 3.5. In particular, we want to study the existence and/or multiplicity of the solutions, as well as their properties when they exist. For the purposes of what follows, we will briefly introduce the notion of a stochastic matrix, also known as a Markov matrix.

Definition 3.3.1. *Let $\mathbf{A} \in \mathbb{R}^{n \times n}$. If $A_{i,j} \in [0,1]$ and $\sum_{i=1}^{n} A_{i,j} = 1 \ \forall j \in \{1, 2, \ldots, n\}$, then $\mathbf{A}$ is said to be (column) stochastic.*

Theorem 3.3.1. *If the graph $\mathcal{G}$ is strongly connected, then the matrices $\mathbf{W^T}$ and $\widetilde{\mathbf{W}}$ are stochastic.*

Proof. Firstly, all elements in $\mathbf{W^T}$ and $\widetilde{\mathbf{W}}$ are between 0 and 1 by construction. Since $\mathcal{G}$ is strongly connected, for every i there exists at least one $j \in \mathcal{V}$ and at least one $k \in \mathcal{V}$ such that $S_{i,j}, S_{j,k} > 0$. Then, for any $j \in \mathcal{V}$:

$$\sum_{i=1}^{n} (W^T)_{i,j} = \sum_{i=1}^{n} W_{j,i} = \sum_{i=1}^{n} \frac{S_{j,i}}{\sum_{k=1}^{n} S_{j,k}} = 1$$

$$\sum_{i=1}^{n} \widetilde{W}_{i,j} = \sum_{i=1}^{n} \frac{S_{i,j}}{\sum_{k=1}^{n} S_{k,j}} = 1$$

□

We will now introduce two lemmas regarding stochastic matrices which will be useful to prove future results.

Proposition 3.3.1. *If $\mathbf{A} \in \mathbb{R}^{n \times n}$ is stochastic, then $\lambda = 1$ is an eigenvalue of $\mathbf{A}$.*

Proof. Since $\mathbf{A}$ is stochastic, $\mathbf{A^T 1} = \mathbf{1}$, where $\mathbf{1} \in \mathbb{R}^n$ is the identity vector. Then, 1 is an eigenvalue of $\mathbf{A^T}$ and since the spectrum of $\mathbf{A}$ coincides with the spectrum of $\mathbf{A^T}$, 1 is an eigenvalue of $\mathbf{A}$. □

Proposition 3.3.2. *If* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *is stochastic, then* $\det(\mathbf{A} - \mathbf{I}) = 0$.

The previous result follows directly from Proposition 3.3.1 and helps us study the existence and multiplicity of the solutions of 3.3 and 3.4.

Theorem 3.3.2. *If* $\mathcal{G}$ *is strongly connected, the systems 3.3 and 3.4 cannot have a unique solution. Moreover, there exists* $\mathbf{b} \in \mathbb{R}^n$ *such that both systems admit solutions.*

Proof. By Proposition 3.3.1, $\mathbf{W^T} - \mathbf{I}$ and $\mathbf{I} - \widetilde{\mathbf{W}}$ are singular. Therefore, depending on $\mathbf{b}$, the systems may have either 0 or infinitely many solutions.
Suppose $\mathbf{b} = \mathbf{0}$. We note that $(\mathbf{W^T} - \mathbf{I})\mathbf{x} = \mathbf{0} \Rightarrow x \in \mathcal{N}(\mathbf{W^T} - \mathbf{I})$, where $\mathcal{N}$ represents the kernel of the linear operator. Being non-singular, $\operatorname{rank}(\mathbf{W^T} - \mathbf{I}) < n$, so, by the rank-nullity theorem, $\dim(\mathbf{W^T} - \mathbf{I}) > 0$. The proof is analogous for $(\mathbf{I} - \widetilde{\mathbf{W}})$. □

The previous theorem indicates that there exists at least one $\mathbf{b}$ such that both systems admit solutions. From an intuitive perspective, the case described above corresponds to a network in which all the mass gets redistributed without any internal consumption or production. For $\mathbf{b} \neq \mathbf{0}$, solutions to 3.3 exist if and only if $\mathbf{b} \in \mathcal{C}(\mathbf{W^T} - \mathbf{I})$, where $\mathcal{C}$ represents the column space. A similar claim is true for the system 3.4. Based on these observations, from now on we will assume $\mathbf{b}$ is such that the systems admit solutions.

Theorem 3.3.3. *The vector* $\mathbf{b} \in \mathbb{R}^n$ *must satisfy* $\sum_{i=1}^{n} b_i = 0$.

Proof. Without loss of generality, we assume there exists $\mathbf{x} \in \mathbf{R}^n$ such that $(\mathbf{W^T} - \mathbf{I}) = \mathbf{b}$. Then let $\mathbf{A} \in \mathbf{R}^{n\times n}$ be a matrix such that for every i, $A_{i,j} = a_i > 0\ \forall j \in \mathcal{V}$. Then $\mathbf{A}(\mathbf{W^T} - \mathbf{I})\mathbf{x} = \mathbf{Ab}$ and by Theorem 3.3.1, $\sum_{j=1}^{n} a_i b_j = 0$ for every $i \in \mathcal{V}$, which yields the conclusion. □

We have thus obtained that the vector of markets, which incorporates both production and consumption, should balance supply and demand. The total quantity from the markets with negative values corresponding to the producers should be equal in absolute value to the total quantity in the positive markets (i.e., consumers).

Theorem 3.3.4. *If* $\widetilde{\mathbf{x}} = \mathbf{W^T x}$ *and* $\mathbf{x}, \widetilde{\mathbf{x}} \geq \mathbf{0}$*, then* $||\widetilde{\mathbf{x}}||_1 = ||\mathbf{x}||_1$.

Proof.

$$
\begin{aligned}
||\widetilde{\mathbf{x}}||_1 = ||\mathbf{W^T x}||_1 &= \sum_i |(\mathbf{W^T x})_i| = \sum_i (\mathbf{W^T x})_i = \sum_i \sum_j \frac{S_{j,i}}{\sum_k S_{j,k}} x_j \\
&= \sum_j x_j \sum_i \frac{S_{j,i}}{\sum_k S_{j,k}} = \sum_j x_j = ||\mathbf{x}||_1
\end{aligned}
$$

□

We observe that any non-negative solutions satisfying 3.5 must have the same modulus, meaning the total sum of exports must be equal to the total sum of imports.

Theorem 3.3.5. *Let* $\mathbf{x}, \widetilde{\mathbf{x}}$ *be solutions to 3.3 and 3.4, respectively, such that* $\widetilde{\mathbf{x}} = \mathbf{W^T x}$*. Then* $\mathbf{x}$ *is an eigenvector corresponding to the eigenvalue 1 for the matrix* $\widetilde{\mathbf{W}}\mathbf{W^T}$.

Proof. From 3.4 and 3.5, we obtain that $(\mathbf{I} - \widetilde{\mathbf{W}})\mathbf{W^T x} = \mathbf{b}$. From 3.3, we obtain that $\mathbf{W^T x} = \mathbf{b} + \mathbf{x}$. The two results yield the conclusion:

$$
\widetilde{\mathbf{W}}\mathbf{W^T x} = \mathbf{x}
$$

□

The previous result is remarkable because it showcases a practical way of obtaining $\mathbf{x}$ with the desired properties. In particular, by solving the eigenvalue problem above, one can determine, up to scaling, solutions for both systems that satisfy 3.5. At this point it is unclear, however, if 1 is an eigenvalue of $\widetilde{\mathbf{W}}\mathbf{W^T}$ and whether it is simple (i.e. its multiplicity is equal to one). We would also like the corresponding eigenvector to be non-negative and unique. Furthermore, we would like to understand under what assumptions we can expect such behaviour and whether these assumptions are consistent with the empirical setting.

3.4 The Perron-Frobenius Theorem for Non-negative Matrices

To provide answers to the previous questions, we turn to the Perron-Frobenius theorem and theory of non-negative matrices [16]. We first introduce some terminology and results and then demonstrate how they relate to our particular application.

Definition 3.4.1. *Let* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *be a matrix such that* $A_{i,j} \geq 0\ \forall 1 \leq i, j \leq n$. *We say* $\mathbf{A}$ *is non-negative and write* $\mathbf{A} \geq \mathbf{O}$.

Definition 3.4.2. *Let* $\mathbf{A} \in \mathbb{R}^{n\times n} \geq \mathbf{O}$ *and* $\mathcal{G}_A$ *be the directed graph represented by* $(\mathcal{V}, \mathcal{E})$, *where* $\mathcal{V} = \{1, 2, \dots, n\}$ *and* $\mathcal{E} = \{(i, j) : A_{i,j} > 0\}$. *We say* $\mathbf{A}$ *is irreducible if* $\mathcal{G}_A$ *is strongly connected.*

Theorem 3.4.1. *If the graph* $\mathcal{G}$ *is strongly connected, then the matrix* $\widetilde{\mathbf{W}}\mathbf{W}^{\mathbf{T}}$ *is irreducible.*

Proof. Let $\mathbf{A}_{\mathcal{G}} \in \mathbb{R}^{n\times n}$ be the adjacency matrix induced by $\widetilde{\mathbf{W}}$, i.e. $(A_{\mathcal{G}})_{i,j} = 1$ if $\widetilde{W}_{i,j} > 0$ and 0 otherwise. Similarly, let $\mathbf{A}_{\mathcal{G}'} \in \mathbb{R}^{n\times n}$ be the adjacency matrix induced by $\mathbf{W}^{\mathbf{T}}$. One can observe that $\mathbf{A}_{\mathcal{G}}$ corresponds to the adjacency matrix of our original graph $\mathcal{G}$, whereas $\mathbf{A}_{\mathcal{G}'}$ corresponds to the adjacency matrix of $\mathcal{G}'$, which is the graph in which the direction of each edge of $\mathcal{G}$ is inverted. Therefore, we have that for any $i, j \in \mathcal{V}$:

$$(A_{\mathcal{G}})_{i,j} = (A_{\mathcal{G}'})_{j,i}$$

Let $\mathbf{V} := \widetilde{\mathbf{W}}\mathbf{W}^{\mathbf{T}}$, $\mathbf{A}_{\mathbf{V}}$ be the adjacency matrix of $\mathbf{V}$ and $\mathcal{G}_V$ be the associated graph. Then $\mathbf{A}_{\mathbf{V}} = \chi\{\mathbf{A}_{\mathcal{G}}\mathbf{A}_{\mathcal{G}'} > 0\}$, where χ is the characteristic function and it is applied element-wise. Using the result from above, we have that:

$$(A_V)_{i,j} = \chi\left\{\sum_k (A_{\mathcal{G}})_{i,k}(A_{\mathcal{G}'})_{k,j} > 0\right\} = \chi\left\{\sum_k (A_{\mathcal{G}})_{j,k}(A_{\mathcal{G}'})_{k,i} > 0\right\} = (A_V)_{j,i}$$

It follows that $\mathbf{A}_{\mathbf{V}}$ is symmetric, so $\mathcal{G}_V$ is an undirected graph, meaning that the notion of strong connectivity coincides with the connectivity of $\mathcal{G}_V$.

Suppose that $\mathcal{G}_V$ contains an isolated vertex v. Then $(A_V)_{i,v} = 0\ \forall i \in \{1, 2, \dots, n\}$. This implies, by the construction of $\mathbf{A_V}$, that $V_{i,v} = 0 \ \forall i \in \{1, 2, \ldots, n\}$, but $\mathbf{V}$ is stochastic, meaning that $\sum_{i=1}^{n} V_{i,v} = 1$, which yields a contradiction. $\square$

The concepts described above allow us to introduce the formulation of the Perron-Frobenius theorem for non-negative irreducible matrices. For a proof of this result, one may look at [15].

Theorem 3.4.2 (Perron-Frobenius). *Let* $\mathbf{A} \in \mathbb{R}^{n \times n} \geq \mathbf{O}$ *be irreducible. We denote with* $\rho(\mathbf{A}) > 0$ *the spectral radius of* $\mathbf{A}$*, i.e. the modulus of the largest eigenvalue. Then the following statements hold:*

1. *There exists an eigenvalue* $r \in \mathbb{R}$ *such that* $r = \rho(\mathbf{A})$*.* r *is called the Perron-Frobenius eigenvalue of* $\mathbf{A}$*;*

2. r *is simple, meaning the left and right eigenspaces associated to* r *are one-dimensional and its algebraic multiplicity is also equal to one;*

3. *The left and right eigenvectors associated to* r *have real non-negative components;*

4. $r \leq \max_i \sum_j A_{i,j}$.

Theorem 3.4.3. *If the graph* $\mathcal{G}$ *is strongly connected, then the matrix* $\widetilde{\mathbf{W}\mathbf{W}^\mathbf{T}}$ *has a unique eigenvalue equal to 1. Moreover, the corresponding eigenspace has dimension* 1 *and contains the unique (up to scaling) eigenvector that has real, non-negative components.*

Proof. By Theorem 3.4.1, $\mathbf{V} := \widetilde{\mathbf{W}\mathbf{W}^\mathbf{T}}$ is irreducible and $\mathbf{V} \geq \mathbf{O}$. We can thus apply the Perron-Frobenius theorem to $\mathbf{V}$. By calling upon the stochastic property of $\mathbf{V}$, we know that $r \leq \max_i \sum_j V_{i,j} = 1$. Furthermore, due to Proposition 3.3.1, $\lambda = 1$ is an eigenvalue of $\mathbf{V}$ so it must be Perron-Frobenius eigenvalue. $\square$

3.5 Beyond Strong Connectivity

The previous results rely on the assumption that $\mathcal{G}$ is strongly connected. This turns out be false in our case. Fortunately, the conclusions stated above still hold true under a weaker assumption. Crucially, we provide a version of Theorem 3.4.3 that does not require $\mathcal{G}$ to be strongly connected and does not rely on the irreducibility of $\widetilde{\mathbf{W}}\mathbf{W}^{\mathbf{T}}$. The sufficient condition is that $\mathcal{G}$ contains at most one strongly connected component of size greater than one, which is satisfied by our seizure graph as observed in Section 2.3.

Theorem 3.5.1. *If the graph $\mathcal{G}$ contains at most one strongly connected component of size greater than one, then the matrix $\widetilde{\mathbf{W}}\mathbf{W}^{\mathbf{T}}$ has a unique eigenvalue equal to 1. Moreover, the corresponding eigenspace has dimension* 1 *and contains the unique (up to scaling) eigenvector that has real, non-negative components.*

Proof. Analogously to the proof of Theorem 3.4.1, the adjacency matrix of $\mathcal{G}_V$ is symmetric. The main difference is that $\mathbf{V}$ is no longer stochastic, since $\widetilde{\mathbf{W}}$ and $\mathbf{W}^{\mathbf{T}}$ may contain columns full of zeros.

Let $C_1, C_2, \ldots, C_m$ be the strongly connected components of $\mathcal{G}$, such that $|C_1| = k = n - m + 1 > 1$ and $|C_i| = 1 \ \forall i \in \{2, 3, \ldots, m\}$. Without loss of generality, we assume $C_1 = \{1, 2, \ldots, k\}$, otherwise we can transform the graph through an isomorphism that guarantees this property. Then we can write $\mathbf{W}^{\mathbf{T}}$ as follows:

$$\mathbf{W}^{\mathbf{T}} = \begin{pmatrix} \mathbf{B} & \mathbf{O}_{n\times(m-1)} \end{pmatrix},$$

where $\mathbf{B} \in \mathbb{R}^{n\times k}$ contains the columns corresponding to the vertices in C_1, i.e. those columns that sum up to one. Although the theorem holds under the assumption on $\mathcal{G}$'s connected components, to simplify the proof we will consider $\widetilde{\mathbf{W}}$ stochastic. This is in fact the case for our application, as our network does not contain nodes with in-degree zero. Then:

$$\mathbf{V} = \widetilde{\mathbf{W}} \begin{pmatrix} \mathbf{B} & \mathbf{O}_{k\times(m-1)} \end{pmatrix} = \begin{pmatrix} \mathbf{C} & \mathbf{O}_{k\times(m-1)} \\ \mathbf{O}_{(m-1)\times k} & \mathbf{O}_{(m-1)\times(m-1)} \end{pmatrix}, \tag{3.6}$$

where $C \in \mathbb{R}^{k\times k}$ is a stochastic matrix. To see why this is true, one should consider that the first k columns of $\mathbf{V}$ sum up to 1, since they are the product of a stochastic matrix with the first k columns of a stochastic matrix. Since $\mathbf{A_V}$ is symmetric, because its last $m-1$ columns are zero vectors, so must be its last $m-1$ rows. By construction, this can only happen if the block bellow $\mathbf{C}$ is a zero matrix.

Let $\mathbf{A_C}$ be the adjacency matrix of $\mathcal{G}_C$, the graph induced by C. Then by an argument similar to the proof of Theorem 3.4.1, because of the strong connectivity of C_1, $\mathcal{G}_C$ is equivalent to a connected undirected graph, meaning it is strongly connected. Consequently, C is irreducible and non-negative, allowing us to apply the Theorem 3.4.3 to the graph $\mathcal{G}_C$ with the irreducible matrix $\mathbf{C}$. It follows that $r = 1$ is the Perron-Frobenius eigenvalue of $\mathbf{C}$, with the unique (up to scaling) eigenvector $\mathbf{v}_C$.

Furthermore, the spectrum of $\mathbf{V}$ contains the spectrum of $\mathbf{C}$ and $m-1$ zeros, due to the block-diagonal shape of $\mathbf{V}$. It follows that $r = 1$ is the unique largest eigenvalue of $\mathcal{V}$. Moreover,

$$\mathbf{v} = \begin{pmatrix} \mathbf{v}_C \\ \mathbf{0}_{m-1} \end{pmatrix}$$

is the unique (up to scaling) eigenvector corresponding to r and its components are non-negative, real numbers. $\square$

3.6 Proposed Method

In light of theorems 3.3.5 and 3.5.1, under the condition regarding the graph's connected components, we can always determine a unique (up to scaling) solution that guarantees the desirable requirement 3.5, as well as non-negativity of the solution. Therefore, we propose an approach of finding the exports $\mathbf{x}$ and then retrieving the imports, as well as the flows across each edge. Our method contains the following steps:

1. Given $\mathcal{G}$, we verify that it does not contain more than one strongly connected component of size greater than one. If this check fails, we stop.

2. We build $\widetilde{\mathbf{WW^T}}$ from $\mathcal{G}$ and find the eigenvector associated to the eigenvalue 1.

3. We normalize the eigenvector and re-scale it using some predetermined scale parameter α.

4. We use the weight matrix $\mathbf{W}$ to retrieve the flows across each edge.

Regarding the fixing of α, the scale of the exports, it is enough to estimate the size of the exports of one country. In particular, throughout our experiments we will be using Colombia's exports, since this country is the largest producer of cocaine in the world and there is an abundance of data available relating to its production [18].

The *Exp-Est* algorithm obtains the normalized Perron-Frobenius eigenvector of $\widetilde{\mathbf{W}}\mathbf{W}^{\mathbf{T}}$

Algorithm 1 Exp-Est

Input: $\mathcal{G} = (\mathcal{V}, \mathcal{E}), \alpha > 0$
Output: $\alpha\bar{\mathbf{v}} \in \mathbb{R}^n$
Require: $\mathcal{G}$ contains no more than one strongly connected component of size greater than one.
Build $\mathbf{W}^{\mathbf{T}}$ and $\widetilde{\mathbf{W}}$ from $\mathcal{G}$;
Find $\mathbf{v}$ such that $\widetilde{\mathbf{W}}\mathbf{W}^{\mathbf{T}}\mathbf{v} = \mathbf{v}$;
Set $\bar{\mathbf{v}} := \mathbf{v}/\sum_i v_i$;
return $\alpha\bar{\mathbf{v}}$;

Algorithm 2 Flow-Est

Input: $\mathcal{G} = (\mathcal{V}, \mathcal{E}), \alpha > 0$
Output: $\mathbf{F} \in \mathbb{R}^{n\times n}$
Require: $\mathcal{G}$ contains no more than one strongly connected component of size greater than one.
Set $\mathbf{x}$:= Exp-Est$(\mathcal{G}, \alpha)$;
Set $\mathbf{F} = \mathbf{O}_{n\times n}$;
for i $\in$ {1, 2, ..., n} **do**
    Set $F_{i,j} = W_{i,j} \cdot x_i$;
end for
return $\mathbf{F}$;

and then re-scales it. *Flow-Est* subsequently retrieves the flow matrix $\mathbf{F}$, completing step 3 of our methodology. The theorems provided in the previous sections ensure the correctness and stability of these algorithms.

3.7 Empirical Results

Having provided a theory, as well as estimation methods for recovering the flows in our network, we now empirically test our techniques. Firstly, as expected, the spectrum of $\widetilde{\mathbf{W}}\mathbf{W}^{\mathbf{T}}$ contains a unique eigenvalue equal to one, as displayed in figure 3.1. We thus apply algorithm 1 and compare the normalized exports $\mathbf{x}/||\mathbf{x}||_1$ and imports $\widetilde{\mathbf{x}}/||\widetilde{\mathbf{x}}||_1$ of the top

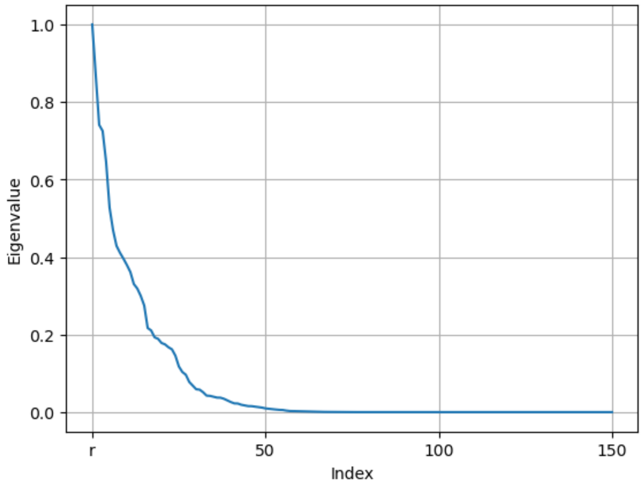

Figure 3.1: Eigenvalues of $\widetilde{\mathbf{W}}\mathbf{W}^{\mathbf{T}}$ sorted in descending order

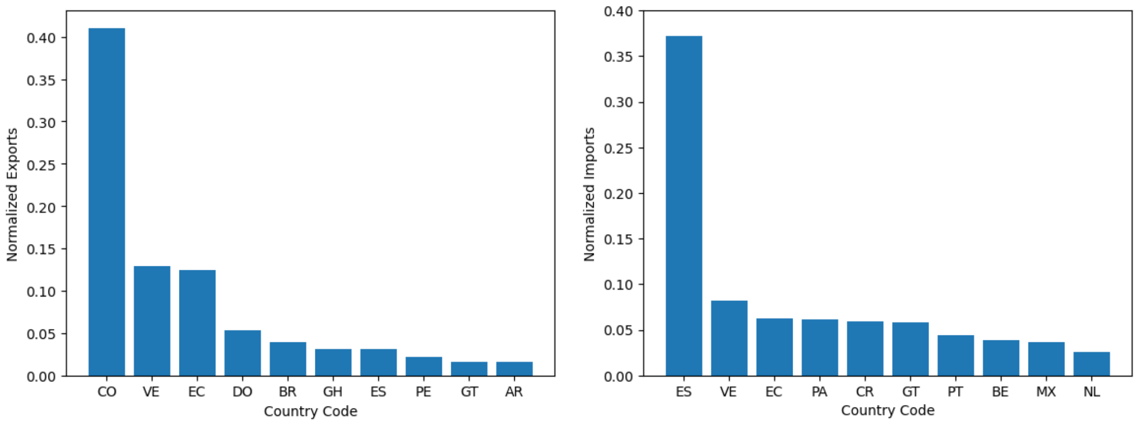

Figure 3.2: Top 10 Largest Normalized Exports and Imports

10 exporters and importers in figure 3.2.

Interestingly, four countries are part of both groups: Venezuela, Ecuador, Guatemala, and Spain. The first three are key transit countries which play an essential role in shipping the cocaine to the United States, Europe, and Asia. Spain, on the other hand, is the largest hub in the network, receiving and distributing significant amounts of cocaine. In the top exporting group, Ghana is the only country besides Spain that is not located in South America. This does not prevent it from acting as a major transit point for the cocaine produced in Colombia and its neighbours. Panama, Costa Rica and Mexico are countries located in Central America that import large quantities of cocaine from South America and ship it further to the United States and Europe, thus acting as downstream transit nodes. Finally, Portugal, Belgium, and the Netherlands serve as important gateways into

Country	Market (exogenous estimation)	Market (flow estimation)
Spain	15245.82	258950.41
Portugal	2787.61	25817.89
Belgium	448.94	22694.03
Netherlands	838.04	18740.54
United Kingdom	4004.55	1535.65
United States	14941.36	14760.96
Italy	3139.54	11430.67
France	2429.32	7502.73

Table 3.1: Comparison of Market Estimates

the European market. While not at the same level as Spain, these countries represent entry points for significant quantities of cocaine in the EU.

We subsequently recover the markets from our flow estimation procedure. As anticipated from 3.3.3, the sum of the markets is equal zero, meaning supply and demand are perfectly balanced. We compare the values for some of the largest markets in table 3.1. While for the EU countries there appear to be large, up to an order of magnitude differences w.r.t. to the values from step 2, estimates for the United Kingdom and the United States are much closer. This may be due to the issues we pointed out in section 2.2, as step 2 tends to produce market estimates which mostly reflect a country's population and seizures and are less connected to the country's consumption patterns and overall role within the global network. On the other hand, Spain's estimated market of over 2.5 hundred tons could be related to the Schengen area and the free flow of goods in the EU. We believe that once drugs reach Spain without being seized, it is significantly more unlikely that they will be seized when travelling within the EU. Since our methodology mainly relies on the seizures to identify connections and estimate flows, it could be the case that Spain's market approximately represents the entire EU market. This is an interesting aspect which could be investigated in a future study.

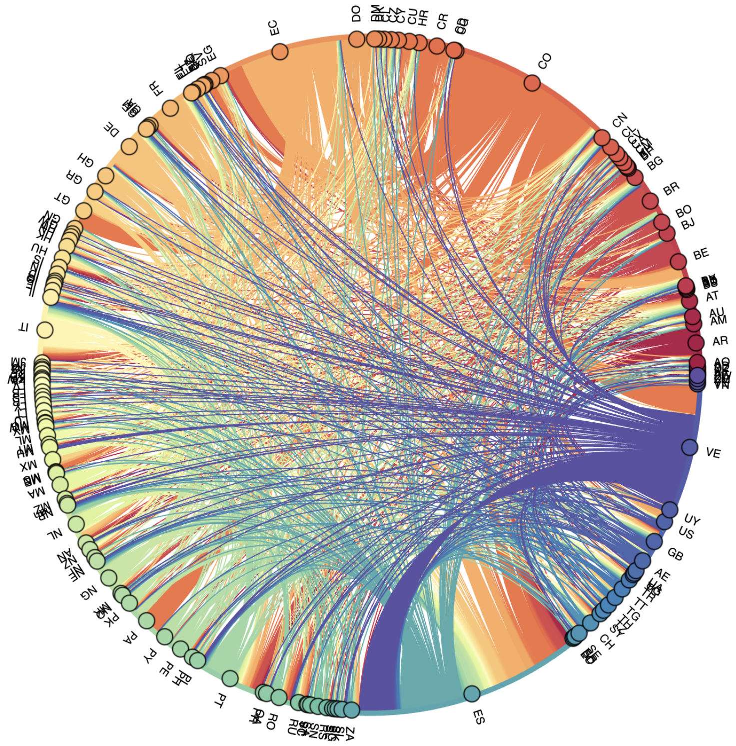

Figure 3.3: Chord Plot of Estimated Flows

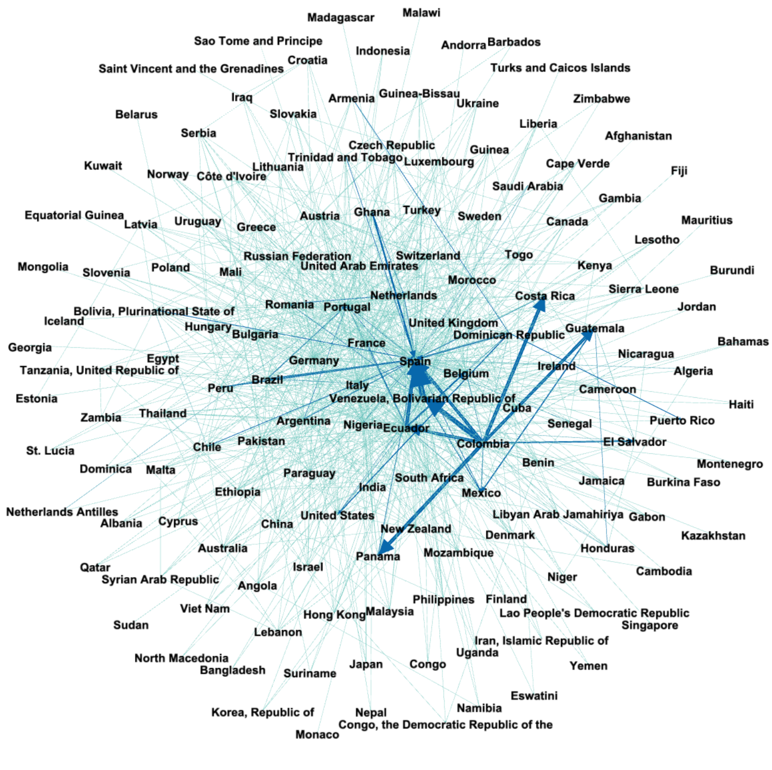

Figure 3.4: Cocaine Network with Estimated Flows (Fruchterman Reingold Layout [4])

Chapter 4

Conclusions

4.1 Findings

Our work has shed light upon several important aspects of flow estimation within transnational drug trafficking networks. Firstly, we pointed out the fact that under the current model, the systems are poorly conditioned, landing attempts to directly solve them highly problematic. Secondly, we determined the conditions under which reasonable solutions exist and proposed an approach that achieves good results in the empirical setting. This allowed us to observe interesting patterns, such as the overestimation of certain markets like Spain. This in turn highlights the limitations related to estimating relative weights solely based on seizure data, while prompting us to hypothesize that this is a consequence of the EU's policies and, in particular, the Schengen area.

4.2 Future Directions

There are several ways of expanding upon our study. On the methodological side, one may try to extend our approach to a more general setting that allows for several large connected components. More importantly, one may attempt to improve the estimation procedure for the relative weights. This could involve taking into account population, travel patterns, legal trade networks, and other data that are relevant for the determining the relative magnitude of the flows.

From an empirical perspective, a natural extension of this work would be to apply our framework to another type of drug network. The most appropriate candidate is heroin, since, like cocaine, there is a limited number of large producers. Continuing the analysis of the cocaine network is another important direction. Understanding the similarity between the illicit network and its legal counterparts, such as trade networks for different product types, may indicate the means of hiding and transporting cocaine across borders. Another objective may be to delve deeper into the flows between EU countries and finding a way to better understand and adjust the relative weights. This is directly related to the case of Spain's overestimated market due to the large difference between exports and imports.

Appendices

Appendix A

Data Sources

1. *Individual Drug Seizures (IDS)*, UNODC - accessed in .xlsx format from: `https://dmp.unodc.org/`. Currently, the dataset published on the UNODC website is incomplete and does not contain information about source and destination countries for individual seizures. The version of the dataset used for this study was provided by prof. Luca Giommoni (Cardiff University).

2. *Population by 5-year age groups and sex*, UNDESA - accessed through the API, details available at: `https://population.un.org/dataportal/about/dataapi`. We considered the age group 15-64.

3. *Prevalence of drug use in the general population - national data*, UNODC - accessed in .xlsx format from: `https://wdr.unodc.org/wdr2019/en/maps-and-tables.html`.

4. *Drug purity*, UNODC - accessed in .xlsx format from: `https://www.unodc.org/unodc/en/data-and-analysis/wdr2022_annex.html`

5. *Geographic coordinates*, World Factbook - accessed in .csv format from: `https://www.cia.gov/the-world-factbook/field/geographic-coordinates/`

6. *ALPHA 2 code*, ISO - accessed in .xlsx format from: `https://www.iso.org/iso-3166-country-codes.html`

Appendix B

Country Codes

Afghanistan AF	Croatia HR
Albania AL	Cuba CU
Algeria DZ	Cyprus CY
Andorra AD	Czech Republic CZ
Angola AO	Côte d'Ivoire CI
Argentina AR	Denmark DK
Armenia AM	Dominica DM
Australia AU	Dominican Republic DO
Austria AT	Ecuador EC
Bahamas BS	Egypt EG
Bangladesh BD	El Salvador SV
Barbados BB	Equatorial Guinea GQ
Belarus BY	Estonia EE
Belgium BE	Eswatini SZ
Benin BJ	Ethiopia ET
Bolivia, Plurinational State of BO	Fiji FJ
Brazil BR	Finland FI
Bulgaria BG	France FR
Burkina Faso BF	Gabon GA
Burundi BI	Gambia GM
Cambodia KH	Georgia GE
Cameroon CM	Germany DE
Canada CA	Ghana GH
Cape Verde CV	Greece GR
Chile CL	Guatemala GT
China CN	Guinea GN
Colombia CO	Guinea-Bissau GW
Congo CG	Haiti HT
Congo, the Democratic Republic of the CD	Honduras HN
Costa Rica CR	Hong Kong HK

Hungary HU	New Zealand NZ
Iceland IS	Nicaragua NI
India IN	Niger NE
Indonesia ID	Nigeria NG
Iran, Islamic Republic of IR	North Macedonia MK
Iraq IQ	Norway NO
Ireland IE	Pakistan PK
Israel IL	Panama PA
Italy IT	Paraguay PY
Jamaica JM	Peru PE
Japan JP	Philippines PH
Jordan JO	Poland PL
Kazakhstan KZ	Portugal PT
Kenya KE	Puerto Rico PR
Korea, Republic of KR	Qatar QA
Kuwait KW	Romania RO
Lao People's Democratic Republic LA	Russian Federation RU
Latvia LV	Saint Vincent and the Grenadines VC
Lebanon LB	Sao Tome and Principe ST
Lesotho LS	Saudi Arabia SA
Liberia LR	Senegal SN
Libyan Arab Jamahiriya LY	Serbia RS
Lithuania LT	Sierra Leone SL
Luxembourg LU	Singapore SG
Madagascar MG	Slovakia SK
Malawi MW	Slovenia SI
Malaysia MY	South Africa ZA
Mali ML	Spain ES
Malta MT	St. Lucia LC
Mauritius MU	Sudan SD
Mexico MX	Suriname SR
Monaco MC	Sweden SE
Mongolia MN	Switzerland CH
Montenegro ME	Syrian Arab Republic SY
Morocco MA	Tanzania, United Republic of TZ
Mozambique MZ	Thailand TH
Namibia nan	Togo TG
Nepal NP	Trinidad and Tobago TT
Netherlands NL	Turkey TR
Netherlands Antilles AN	Turks and Caicos Islands TC

Uganda UG Ukraine UA United Arab Emirates AE United Kingdom GB United States US Uruguay UY Venezuela, Bolivarian Republic of VE Viet Nam VN Yemen YE Zambia ZM Zimbabwe ZW	Venezuela, Bolivarian Republic of VE Viet Nam VN Yemen YE Zambia ZM Zimbabwe ZW

The country codes used are the ISO ALPHA 2 codes [6].

Bibliography

[1] Alberto Aziani. Illicit financial flows, Jun 2018.

[2] Alberto Aziani, Giulia Berlusconi, and Luca Giommoni. A quantitative application of enterprise and social embeddedness theories to the transnational trafficking of cocaine in europe. *Deviant Behavior*, 42, 09 2019.

[3] Albert-László Barabási and Márton Pósfai. *Network science*. Cambridge University Press, Cambridge, 2016.

[4] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: An open source software for exploring and manipulating networks. 2009.

[5] Giulia Berlusconi, Alberto Aziani, and Luca Giommoni. The determinants of heroin flows in europe: A latent space approach. *Social Networks*, 03 2017.

[6] Rémi Boivin. Drug trafficking networks in the world-economy. *Crime and Networks*, 01 2013.

[7] Jonathan P. Caulkins. After the grand fracture: Scenarios for the collapse of the international drug control regime. *Journal of Drug Policy Analysis*, 10, 2017.

[8] Luca Giommoni, Alberto Aziani, and Giulia Berlusconi. How do illicit drugs move across countries? a network analysis of the heroin supply to europe. *Journal of drug issues*, 47, 12 2016.

[9] Luca Giommoni, Giulia Berlusconi, and Alberto Aziani. Interdicting international drug trafficking: A network approach for coordinated and targeted interventions - european journal on criminal policy and research. *SpringerLink*, Feb 2021.

[10] Roger V. Gould and Roberto M. Fernandez. Structures of mediation: A formal approach to brokerage in transaction networks. *Sociological Methodology*, 19:89–126, 1989.

[11] Beau Kilmer, Peter Reuter, and Luca Giommoni. What can be learned from cross-national comparisons of data on illegal drugs? *Crime and Justice*, 44:000–000, 08 2015.

[12] M. E. J. Newman. Analysis of weighted networks. *Phys. Rev. E*, 70:056131, Nov 2004.

[13] Letizia Paoli, Victoria Greenfield, and Peter Reuter. *The World Heroin Market: Can Supply be Cut?* 04 2009.

[14] P.H. Reuter. The mobility of drug trafficking. in j. collins (ed.), ending the drug wars. *Report of the LSE Expert Group on the economics of drug policy (pp. 33–40)*, 2014.

[15] M. R. F. Smyth. A spectral theoretic proof of perron-frobenius. *Mathematical Proceedings of the Royal Irish Academy*, 102A(1):29–35, 2002.

[16] G. Strang. *Linear Algebra and Learning from Data.* Wellesley-Cambridge Press, 2019.

[17] UNODC. Drug money and opiate trafficking on the balkan route, focus of new unodc report, 2015.

[18] UNODC. World drug report 2019 - methodology report, 2019.

[19] UNODC. Executive summary - world drug report 2023, 2023.