



UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
MAESTRÍA EN DATA MINING & KNOWLEDGE DISCOVERY

Técnicas de Ciencia de Datos aplicadas al comercio internacional

provisorio

Tesis de Maestría

Diego Kozlowski

Director: Dra. Viktoriya Semeshenko

Buenos Aires, 2019

TÉCNICAS DE CIENCIA DE DATOS APLICADAS AL COMERCIO INTERNACIONAL

PROVISORIO

Abstract
pendiente

Palabras claves: Comercio Internacional, Grafos, Ciencia de Datos, Dirección Latente de Dirichlet, Aprendizaje no Supervisado.

DATA SCIENCE TECHNIQUES APPLIED TO WORLD TRADE

PROVISORIO

abstract pendiente

Keywords: World Trade, Graphs, Data Science, Latent Dirichlet Allocation, Unsupervised Learning.

AGRADECIMIENTOS

pendiente

Índice general

1. introducción	1
1.1 Fuentes de información	2
1.2 Software utilizado	3
2. Comercio a nivel países	4
2.1 Introducción	4
2.2 Análisis Exploratorio de Datos	4
2.3 Metodología	5
2.3.1 Construcción del grafo	5
2.3.2 El comercio internacional como una red compleja	10
2.3.3 Determinación del punto de corte	12
2.4 Resultados	14
2.4.1 Correlación entre la representación de las exportaciones y las importaciones	14
2.4.2 Evolución de mediano y largo plazo del comercio internacional	17
Análisis mediano plazo	17
Análisis de largo plazo	18
2.5 Conclusiones	23
3. Comercio a nivel producto	26
3.1 introducción	26
3.2 Análisis exploratorio de datos	26
3.3 Metodología	31
3.3.1 Latent Dirichlet Allocation Models	31
Definiciones	34
Proceso generativo e inferencial	35
3.3.2 Grafo Bipartito	37
3.4 Resultados	37
3.4.1 Latent Dirichlet Allocation Models	37
3.4.2 Grafo Bipartito	39
4. Conclusiones	40
Apéndice	41
.1 Tabla de referencia: Descripción cadenas	42
.2 Tabla de referencia: Descripción subcadenas	42
.3 Tabla de referencia: Descripción Usos	44

1. INTRODUCCIÓN

El análisis del comercio internacional es una de las áreas de estudio más importantes de la investigación económica. Desde los comienzos de la economía política clásica constituye un tema de preocupación por el efecto del mismo en el desarrollo económico de los países Ricardo (2007). Por su parte, el registro de la información referente al comercio entre países también se remonta en el tiempo.

La cantidad de vínculos comerciales que se establecen entre entidades ubicadas en distintos países implica la imposibilidad de estudiar el fenómeno de forma directa, y plantea la necesidad de elaborar medidas de resumen que permitan apropiar la información subyacente al conjunto de los contratos comerciales existentes.

Sin embargo, el análisis tradicional de la información generada carece de las herramientas necesarias para hacer frente a los grandes volúmenes de datos generados por el comercio internacional en la actualidad. Históricamente, los indicadores sintéticos del área se resumen en volumen y masa de dinero comerciada, partiendo del total mundial, hacia desagregaciones por región, país y sector económico en cuestión WTO (2017). De aquí se desprende la potencialidad del análisis que se basa en técnicas que logren captar en indicadores sucintos la complejidad del objeto de estudio.

A la vez que aumenta el volumen de información persistida respecto del comercio internacional, también se facilita el acceso a técnicas de análisis de datos que requieren un mayor poder de cómputo, y que por lo tanto eran impensadas como herramientas de estudio en épocas anteriores. En este sentido, surge la posibilidad de complementar el análisis tradicional del comercio mundial con técnicas de mayor riqueza en términos cuantitativos.

El presente trabajo se propone utilizar las nuevas técnicas que provee el análisis de grafos para caracterizar el comercio internacional. Su modelización como una red compleja permite la construcción de medidas de resumen que, sin abandonar una mirada holística de la problemática, logren dar cuenta de una mayor complejidad que las métricas tradicionales de dicha área temática.

Una de las posibles formas en que se puede desarrollar dicho análisis es considerar a cada país como la unidad fundamental a partir de la cual se forma una red basada en sus relaciones comerciales con otros países-nodos. Este tipo de análisis permite tener una mirada general de la posición de los países en el comercio internacional, caracterizar su importancia relativa, así como también inferir propiedades estructurales del comercio internacional en su conjunto. En el capítulo dos se desarrolla el estudio siguiendo esta premisa.

Sin embargo, existe aún un nivel mayor de complejidad que puede ser analizado. Es posible descomponer el comercio bilateral entre dos países en los múltiples contratos comerciales individuales entre entidades de los diferentes países. Si bien ese nivel de desagregación es el máximo posible en el objeto de estudio, dicha información no se encuentra disponible al público. Lo que sí es accesible al público es la información de los volúmenes comerciados de cada tipo de producto entre cada par de países, siguiendo un nomenclador estandarizado como el 'Standard International Trade Classification' (?) o el 'Harmonized System' (Weerth, 2008). Este nivel de desagregación implica que si se continúa considerando a los países como la unidad fundamental de análisis, entre cada par de nodos puede

Nivel	Definición
Sección II	Productos Vegetales
Capítulo 10	Cereales
Partida 10.06	Arroz
Subpartida 1006.30	Arroz semiblanqueado o blanqueado, incluso pulido o glaseado

Tab. 1.1: Ejemplo de desagregación para el Sistema Harmonizado

existir una multiplicidad de aristas, potencialmente una por cada producto definido por el nomenclador. Esto implica la construcción de un grafo bipartito entre países y productos. Dicho modelo se analiza en el capítulo tercero de esta tesis.

Por su parte, los nomencladores mencionados tienen la propiedad de ser sistemas jerárquicos en los cuales un mismo producto final pertenece a una serie de categorías intermedias, la tabla 1.1 muestra un ejemplo de los distintos niveles de desagregación que presenta el arroz. Esta estructura de datos permite realizar un análisis exploratorio de la información a través de los diferentes niveles pero a su vez plantea la cuestión de cual es la mejor estructura jerárquica para el análisis de la información. Frente a esto, la multiplicidad de nomencladores, tanto oficiales como realizados ad-hoc para análisis específicos (Molinari & de Angelis, 2016) plantean que no existe un único sistema de estandarización de la información. De aquí surge la posibilidad de utilizar técnicas de reducción de dimensionalidad que utilicen la información disponible para agrupar los productos según grupos de pertenencia. En el tercer capítulo de este trabajo también se avanza en esta dirección proponiendo la utilización del modelo de Latent Dirichlet Allocation (Blei et al., 2003) para explotar la información disponible en la elaboración de un sistema de clustering difuso.

ver que pasa
con grafos
bipartitos

1.1. Fuentes de información

La modelización del comercio internacional como un grafo requiere de los datos del flujo anual de comercio bilateral para el total de las mercancías, para la mayor cantidad posible de países, idealmente todos ellos. Dado que los datos del comercio bilateral son realizados por cada país involucrado, es necesaria una base de datos en la cual se haya recolectado la información provista por cada país, y que a su vez haya sido consolidada frente a las posibles, y de hecho abundantes, contradicciones que se presentan entre los reportes oficiales de los países involucrados. En particular, dado que toda importación para un país es una exportación de otro, y que el registro de los datos se realiza de forma nacional, existe una duplicación formal de la información, que no siempre es coincidente. Por esto, se recurrió a una base previamente consistida por un organismo internacional oficial, la Organización Mundial del Comercio (de aquí en más OMC)¹, y la API que dicho organismo provee para descarga de datos. La información analizada en las secciones tercera y cuarta, del período 1996-2016, proviene dicho organismo.

Los datos utilizados para el análisis de largo plazo, entre los años 1948 y 2000 proviene del trabajo realizado por Gleditsch (2002), que también es utilizado en otros trabajos de modelización del comercio internacional mediante redes complejas Fagiolo et al. (2010).

¹ <https://comtrade.un.org/>

1.2. Software utilizado

Para el presente trabajo se utilizó el lenguaje de programación estadística R(R Core Team, 2017), junto con sendas extensiones del mismo. Aquellas que resulta importante mencionar son `igraph`(Csardi & Nepusz, 2006) para la construcción de los grafos y las medidas de resumen de los mismos; `tidyverse`(Wickham, 2017) para la manipulación de la información y elaboración de los gráficos. Para esto último, también se hizo uso de (Wilke, 2017; Arnold, 2017; Neuwirth, 2014; Slowikowski, 2017; Vu, 2011), como complementos de `ggplot`. Por su parte, los códigos de los países provienen de (Arel-Bundock, 2017).

Para el análisis del capítulo 3 también se utilizó el lenguaje Python (van Rossum, 1995), y específicamente la librería `scikit-learn` (Pedregosa et al., 2011) para la estimación de los modelos. A su vez, para esta sección también se elaboraron dos tableros dinámicos para el análisis de resultados, sobre la base de la librería `shiny` (Chang et al., 2018), y se hostearon en www.shinyapps.io/ (Allaire, 2019)

2. COMERCIO A NIVEL PAÍSES

2.1. Introducción

En la presente sección se buscará captar los rasgos generales de la estructura del comercio internacional, con independencia de los productos específicos que los países comercian. Si bien el rol que ocupan los países en la Nueva División Internacional del Trabajo determina a la vez qué tipo de mercancías concentran su producción para el mercado mundial y la importancia que tienen los mismos en dicha red (Fröbel et al., 1978), esto no quita el valor de analizar la centralidad de los países, con independencia de las ramas en que cada país concentra sus exportaciones e importaciones. Naturalmente, los resultados de dicho análisis son representan una caracterización que puede resultar útil como complemento para estudios que indaguen respecto de las causas de dicha distribución.

En la literatura reciente se realizaron diversos enfoques desde esta perspectiva, ya sea desde una mirada de la teoría de la información Bhattacharya et al. (2008), como una herramienta de modelización de los fenómenos económicos Fan et al. (2014), para analizar las relaciones de centro-periferia Fagiolo et al. (2010), o bien para realizar una descripción del estado del comercio internacional Chow (2013). Sin embargo, dada la amplitud de los temas a abordar, tanto desde la perspectiva económica, como desde la teoría de grafos, se presenta como un campo abierto para la investigación, con sendas aristas aún por recorrer.

El capítulo se estructura de la siguiente manera: En la sección siguiente se realiza un análisis exploratorio de datos, mientras que la sección tercera describe las fuentes de información utilizadas y realiza un pequeño resumen de los elementos de la teoría de grafos utilizados, así como las particularidades que surgen al aplicar este marco metodológico al comercio internacional. La sección cuarta presenta los resultados, con detalle en el año 2016, así como también un análisis de mediano y largo plazo de la evolución de la red. Finalmente, se presentan las conclusiones específicas a este capítulo.

2.2. Análisis Exploratorio de Datos

En primer lugar, en la figura 2.1(a), analizamos la frecuencia de las interacciones entre países en función de la proporción del comercio que representan para el país importador. La distribución tiene una asimetría a derecha, que asemeja una distribución de chi-cuadrado. En la figura 2.1(a) se puede observar que si bien la gran mayoría de las relaciones bilaterales de comercio representan una baja proporción de las importaciones de un determinado país, existen ciertas relaciones comerciales que se podrían caracterizar como de alta dependencia para el país importador. En particular, se observan cuatro relaciones comerciales de alta dependencia para el año 2016: La venta de mercancías de Estados Unidos a las islas de San Cristóbal y Nieves y Bermudas ; las exportaciones de Dinamarca a Groenlandia y de la India a Nepal.

En la figura 2.1(b) se puede observar aquellas relaciones que representan más de un 70 % de las importaciones del país importador, para el período 1996-2017. Allí se destaca con el tamaño de fuente el Valor comerciado en cada dupla para cada año en particular. La mayoría de las relaciones de tipo dependiente se repiten año a año y constituyen volúmenes bajos de comercio, entre países de tamaños, económicos poblacionales y territoriales, muy

diferentes entre sí. En particular destaca Estados Unidos en su relación comercial con pequeñas islas del Caribe, y Sudáfrica con países de menor tamaño en su mismo continente, como Namibia o Botswana. Otra relación de comercio dependiente que se repite en la serie es la exportación de productos desde Israel a Palestina. Sin embargo, todas estas relaciones implican un monto menor en términos del comercio mundial. La relación de tipo dependiente más importante es la exportación de productos desde Estados Unidos hacia México durante la década del noventa, en el proceso conocido como la "maquila mexicana" (Carrillo & Hualde, 1998)

Por su parte, en la figura 2.2(a) se puede observar la distribución de los montos comerciados en el año 2016, destacando aquellas relaciones de mayor volumen. Al igual que en la figura 2.1(a), también se observa una fuerte asimetría a derecha, siendo las relaciones más importantes en el continente asiático y Americano. En particular, Las importaciones de Estados Unidos desde México, Canadá y China. También destacan las exportaciones de China, además de la ya mencionada, a Japón y Hong-Kong. Estas relaciones no corresponden a caracterizarlas de tipo dependiente en un sentido de intercambio desigual, dado que se trata de volumen, y aquellas que destacan resultan de importancia tanto para exportador como para el importador. Se podría decir que estas son relaciones de alta interdependencia.

En este sentido, la figura 2.2(b) muestra las cinco relaciones más importantes de cada año. Nuevamente lo que se observa es una repetición en las duplas Importador-Exportador entre año y año. Destaca sobre todo el rol de Estados Unidos como importador de China en primer lugar, y Canadá y México en menor volumen. Es importante observar que las exportaciones de China a Estados Unidos toman una importancia creciente en el tiempo y para el final de la serie se encuentran por fuera de la nube de puntos.

Esto también se puede apreciar en relación al total de las relaciones comerciales, en la figura 2.3 donde se incluye la totalidad de las relaciones comerciales para la serie entre 1996 y 2017. Allí se observa que la absoluta mayoría de las relaciones comerciales constituyen montos menores, y que una proporción menor se realiza por montos de hasta cien mil millones de dólares anuales. También se observa un número creciente de operaciones por montos arriba de los cien mil millones de dólares anuales. Finalmente, también cabe destacar nuevamente la anomalía que representa respecto del resto de los datos las exportaciones desde China a Estados Unidos en la segunda mitad de la serie.

2.3. Metodología

2.3.1. Construcción del grafo

Un grafo (o red) se define como un conjunto de nodos y aristas, donde cada nodo representa una entidad, mientras las aristas representan la existencia de una relación entre dos entidades. La representación algebraica del grafo se realiza mediante la denominada matriz de adyacencia, donde las filas representan los nodos de salida, mientras que las columnas representan el nodo de entrada. Por lo tanto, el elemento a_{ij} de dicha matriz representa a la arista que va desde i hacia j .

En el caso de las redes pesadas, una arista, $a_{ij} \in \mathbb{R}$, denota la intensidad de la relación entre los nodos i y j . Por su parte, en las redes binarias, $a_{ij} \in \{0, 1\}$ define cualitativamente la relación, dónde $a_{ij} = 1$ si existe la relación y $a_{ij} = 0$ sino. Este último tipo de grafos sera el utilizado para el presente trabajo, dado que se cuenta con una mayor cantidad de métricas para caracterizar los grafos de este tipo. A su vez, las redes pueden ser tanto dirigidas como no dirigidas. En el primer caso, la relación no tiene un sentido definido,

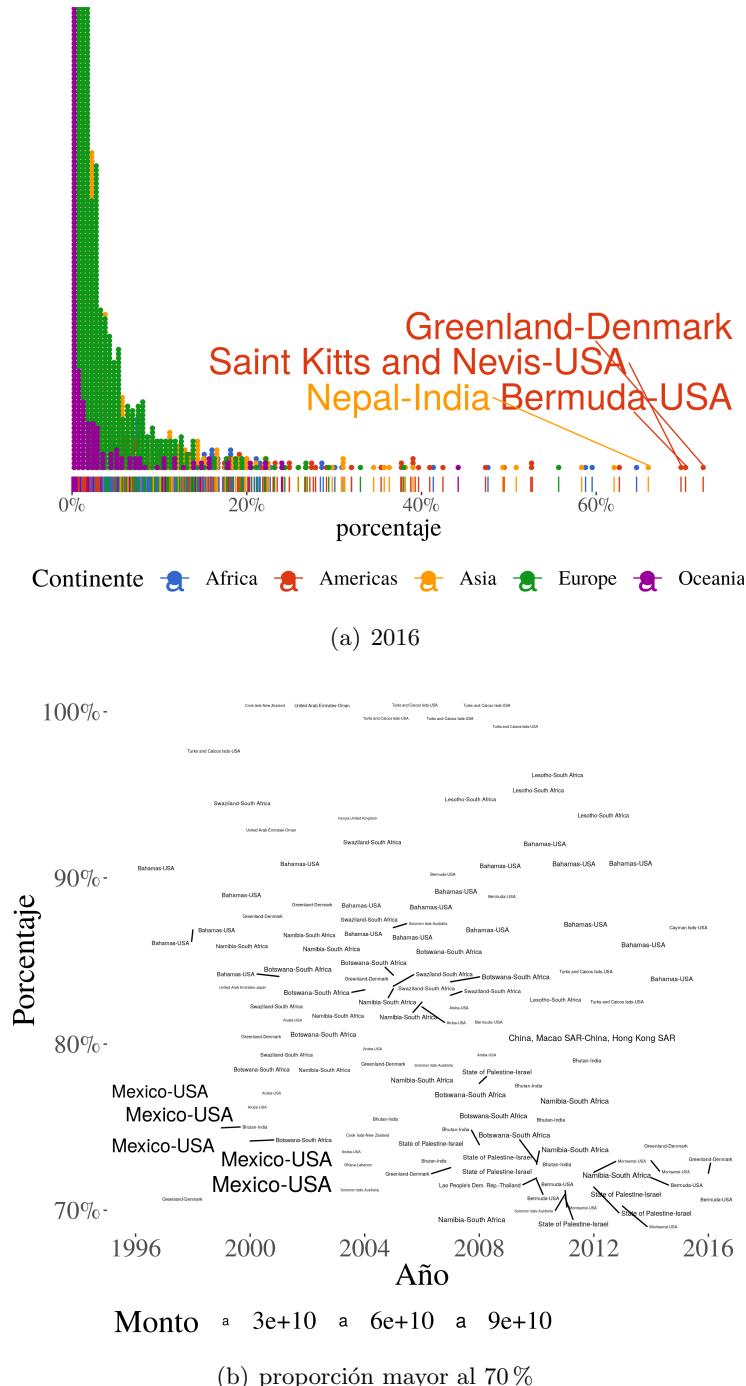


Fig. 2.1: Interacciones Según porcentaje de importaciones que representa al país importador. Des- tacado par Importador-Exportador.1996-2016

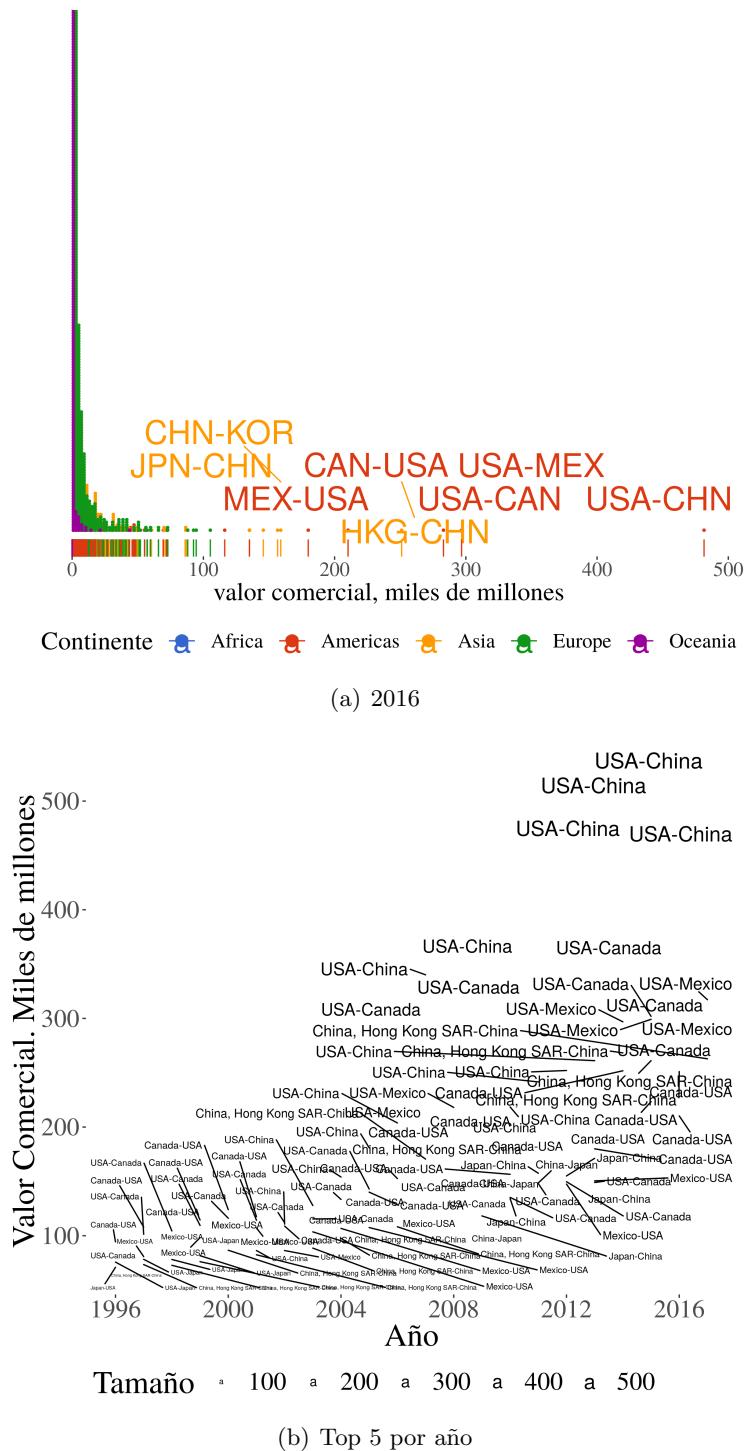


Fig. 2.2: Interacciones. Seg n valor comercial. Destacado par Importador-Exportador. 1996-2016

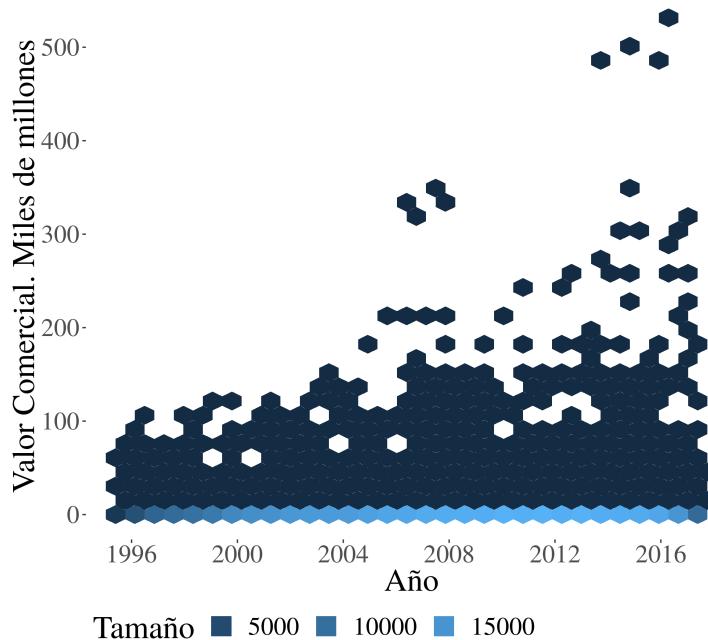


Fig. 2.3: Interacciones. Según valor comercial. 1996-2016

y por lo tanto $a_{ij} = a_{ji}$, mientras que en el segundo caso tal igualdad no es necesaria, y la presencia de una relación tiene un sentido definido. Si bien en la bibliografía existen autores que plantean que la simetría del intercambio comercial es fuerte, y por lo tanto no es necesario utilizar un grafo dirigido para su representación Fagiolo et al. (2007), en el presente trabajo se sostiene que esto depende fuertemente de la forma de construcción de la red, en particular si se trabaja con un grafo no ponderado, y por lo tanto hay una riqueza interpretativa que se pierde si se utiliza una red no dirigida.

Sobre una red se puede calcular una serie de medidas de resumen que permiten caracterizarla. Éstas pueden referirse a un atributo propio del grafo, o bien de cada nodo o arista. Si bien las posibles métricas son muchas, en el presente trabajo nos limitamos a aquellas que se definen a continuación.

Dos nodos se dicen conectados si existe un recorrido entre ellos, es decir, una secuencia de aristas y nodos entre estos. De un grafo se dice que está conectado, si todos sus elementos lo están.

El diámetro d_m de una red es la máxima distancia entre cualquier par de nodos. En teoría de grafos se considera la distancia entre un par de nodos como el camino más corto que se puede realizar entre éstos. Es decir, el diámetro de un grafo no pesado se considera como la cantidad de aristas que hay entre los dos elementos más lejanos del grafo, considerando el recorrido más corto entre ambos.

$$d_m = \max(d_{uv})$$

donde d_{uv} es la distancia entre los nodos u y v, definida como:

$$d_{uv} = \min(L(P))$$

con $P = (v_1, v_2, \dots, v_n)$, un conjunto de n vértices, donde v_i es adyacente a v_{i+1} , y siendo

$L(P)$ la longitud del camino P, definida como:

$$L(P) = \sum_{i=1}^n w_i$$

En el caso de una red no ponderada, $w_i = 1$ y por lo tanto, la longitud de un camino es la cantidad de elementos que lo componen.

La densidad De_m de una red es la relación entre la cantidad de aristas de la red y la máxima cantidad de aristas potenciales de la misma.

$$De_m = \frac{|A|}{|V|(|V| - 1)}$$

Siendo $|A|$ el número de aristas y $|V|$ el número de vértices, o nodos, del grafo. $De_m \in [0, 1]$. Aquellas redes con una densidad cercana a 0 se las considera dispersas, mientras que aquellas con una densidad cercana a 1 se las considera densos.

A su vez, existe una serie de medidas de centralidad de nodos que caracterizan la importancia de los mismos en la red. Las utilizadas en el presente trabajo son la intermediación, cercanía y autovalor.

La centralidad de intermediación mide el rol de un nodo i , $C_{BET}(i)$, como puente entre los demás nodos, considerando sus caminos más cortos. Se define como:

$$C_{BET}(i) = \sum_{j,k \neq i} \frac{b_{jik}}{b_{jk}}$$

donde b_{jk} son todos los caminos más cortos entre los nodos j y k , y b_{jik} son los caminos más cortos entre j y k , que pasan por i .

La centralidad de cercanía es la distancia promedio que tiene un nodo con los demás nodos con los que está conectado. Formalmente:

$$C_{cer}(i) = \frac{1}{\sum_{j \neq i} d_{i,j}}$$

La centralidad de grado se define como la cantidad de aristas de un nodo. En el caso de los grafos dirigidos, se puede considerar un grado de entrada, así como un grado de salida, si las aristas van en dirección al grafo o desde el grafo, respectivamente. La centralidad de autovalor, por su parte, caracteriza la importancia de un nodo considerando la de sus nodos vecinos. Si asumimos linealidad:

$$\begin{aligned} x'_i &= \sum_{j=1}^N a_{ij} x_j \\ x' &= Ax \end{aligned}$$

Siendo A la matriz de adyacencia de la red. si x' es el vector de centralidades de la red, entonces podemos plantear $Ax^* = \lambda x^*$ donde x^* es la solución y λ es el autovalor asociado al mayor autovector. Por lo tanto, podemos definir la medida de centralidad de autovalor como

$$x_i^* = \frac{1}{\lambda} \sum_i A_{ij} x_j^*$$

Esta medida tiene la propiedad de considerar la importancia de un nodo a partir de la importancia de aquellos nodos con los que esta conectado, y con los que éstos están conectados, etc. En el presente trabajo también se considera a la centralidad de autovalor ponderado como una alternativa, donde el peso esta definido por el volumen de dinero comerciado entre los países. Para calcularlo, se multiplica la matriz de adyacencia por una matriz de pesos. Por lo tanto

$$x_i^{pond} = \frac{1}{\lambda} \sum_i W A_{ij} x_j^{pond}$$

Estas medidas de centralidad pueden ser resumidas para el grafo en su conjunto mediante una agregación de algún tipo, como el promedio, máximos y mínimos, etc. En el presente trabajo se utilizan de ambas maneras, tanto para definir la importancia de un país en particular en el grafo, como para caracterizar a la red en su conjunto.

El Coeficiente de clustering estima la cohesividad local, midiendo la probabilidad de que dos nodos que comparten un vecino estén conectados. Para calcularlo, se mide en primer lugar el grado de agrupamiento de cada nodo, tomando la vecindad del mismo, es decir, todos aquellos nodos con los que esta conectado, y se calcula el cociente entre las aristas existentes y la máxima cantidad de aristas posibles:

$$C_i = \frac{|A_i|}{|V_i|(|V_i| - 1)}$$

siendo $|A_i|$ las aristas en el subgrafo de la vecindad de i , y $|V_i|$ los vértices del subgrafo de la vecindad de i . Nótese que esta medida no representa lo mismo que la densidad, si bien las ecuaciones sean similares, dado que en el clustering se analiza solamente la vecindad de un nodo, mientras que la densidad se refiere a toda la red. Si quisieramos caracterizar a toda la red según el coeficiente de clustering, podemos calcular el promedio del mismo para todos los nodos, también conocido como transitividad del grafo, que se define como el promedio del grado de agrupamiento de los nodos. Es decir:

$$\bar{C} = \frac{1}{n} \sum_i C_i$$

Por último, también se considera la correlación de grado entre las aristas. Esta se calcula como la correlación de Pearson entre el grado de nodos adyacentes. Si la correlación es fuerte y positiva, esto implica que la red es de tipo selectiva, lo que indica que los nodos de mayor grado tienden a conectarse con otros nodos de mayor grado. Si la correlación es negativa y fuerte, esto indica que la red es no selectiva o heterogénea, lo que implica que los nodos más importantes de la red tienden a conectarse con nodos poco centrales.

2.3.2. El comercio internacional como una red compleja

La modelización del comercio internacional como un grafo conlleva una serie de simplificaciones de la información original. Nos encontramos en primer lugar que toda compra es a su vez una venta, y que por lo tanto la información se puede interpretar de ambas maneras. Por su parte, dado que lo que se busca representar es existencia de un vínculo

comercial entre dos países, considerar toda compra o toda venta como la existencia de este vínculo resultaría exagerado. Por lo tanto, es necesario establecer un punto de corte a partir del cual se considere que existe una relación comercial entre la dupla de países en cuestión. Sin embargo, tal punto de corte no se debe considerar en término absolutos, como el monto comerciado, ya que dicho monto está sumamente determinado por el tamaño de los países en cuestión. En otros términos, una determinada masa de dinero es significativa para un país en función del valor total comerciado por el mismo. Por lo tanto, es necesario considerar el umbral en términos relativos respecto al tamaño de los mismos.

En este punto se abre una serie de consideraciones respecto al denominador a utilizar. Si se normaliza por el tamaño de la economía, el Producto Bruto Interno (PBI), se está considerando no sólo la importancia relativa de ese vínculo comercial respecto del total comerciado, sino también el grado de apertura de dicha economía. En ese sentido, si se quisiera analizar el rol de un país en el comercio internacional, independientemente del tamaño real de su economía, lo más correcto es normalizar por el volumen total comerciado, y no por su PBI. De esta forma, se puede considerar un punto de corte como un tanto por ciento de las importaciones o exportaciones del país que reporta la transacción.

Dado que se considera que la información de las importaciones suele ser más confiable que la de las exportaciones (Fan et al., 2014), en el presente trabajo se utiliza en la generalidad de los casos el punto de corte como un porcentaje del total de las importaciones del país importador. No obstante, dada la construcción de dicho punto de corte, la interpretación de los resultados varía. Si se utiliza la información provista por el importador, la dirección de la arista resultante es desde el importador hacia el exportador, el sentido contrario al de las mercancías, es decir, el sentido del dinero. Esta arista sólo existe si resulta significativa para el país importador.

Podemos entonces, utilizar un grafo binario dirigido. Dados N países, definimos la matriz de adyacencia NxN, donde:

$$a_{ij} = \begin{cases} 1 & \text{si } \frac{x_{ij}}{x_{i\cdot}} \geq u \\ 0 & \text{sino} \end{cases}$$

Donde u representa el umbral elegido, y x representa el volumen comerciado. Para el caso de las importaciones, x_{ij} representa el monto total comerciado desde el país j hacia el país i , mientras que $x_{i\cdot}$ representa el total de las importaciones del país i . En el grafo de exportaciones, al contrario, x_{ij} representa el monto total comerciado desde el país i hacia el país j , mientras que $x_{i\cdot}$ representa el total exportado por el país i .

Dado que la restricción se construye por la importancia relativa para el nodo de origen, en general sólo existirán una pequeña cantidad de aristas que salgan de cada nodo, aunque no están limitadas las aristas en dirección hacia el nodo. Más precisamente, la máxima cantidad posible de aristas de salida es igual a $\frac{1}{u}$. Esta cantidad se obtiene sólo cuando el país comercia de forma uniforme con exactamente $\frac{1}{u}$ países. Como se verá más adelante, la distribución del comercio dista mucho de ser uniforme, por lo cual la cantidad de aristas de salida de un nodo se verá fuertemente limitada. Sin embargo, las aristas de entrada al nodo sólo se limita formalmente por la cantidad de nodos en el grafo.

Por lo tanto, para el grafo de importaciones, cuando se consideren medidas de centralidad de los nodos, se deberá considerar en particular las aristas que están dirigidas en dirección al nodo, y éstas marcarán la importancia de tal nodo como un país productor del mercado mundial, ya que lo que se mide es si los productos que exporta dicho país resultan significativos o no considerados en las importaciones de los demás países. Por

su parte, si se consideran los datos provistos por los países exportadores, las medidas de centralidad de los nodos, estarán reflejando la importancia de tal país como consumidor global, ya que su consumo resulta significativo como exportación para los países que le venden dichas mercancías. Dado que, por la forma de construcción de los grafos, resulta distinta la interpretación de la información de las exportaciones y las importaciones, se considerarán alternativamente ambas redes según qué resulte más interesante.

2.3.3. Determinación del punto de corte

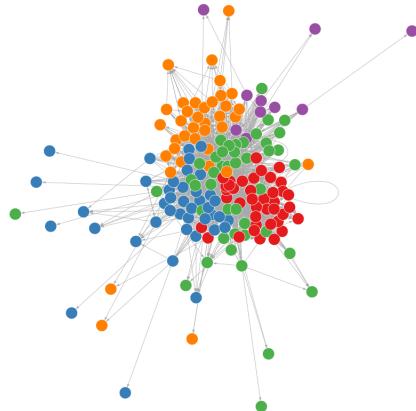
El punto de corte a utilizar, dada la metodología propuesta para discretizar la información del volumen comerciado, opera fuertemente sobre el grafo resultante. La figura 2.4 muestra el grafo según si se utiliza un punto de corte de 1 %, 5 %, 10 %, 15 %, 20 % o 25 %.

Se puede observar como hasta un punto de corte del 10 % la red se mantiene unida, aunque disminuye la cantidad de nodos. A partir de este punto, la red deja de estar débilmente conectada y, a la vez que sigue disminuyendo la cantidad de nodos, se observa mejor la importancia de ciertos nodos de alto grado. En particular, uno del continente americano (Estados Unidos) y otro del continente africano (Sudáfrica).

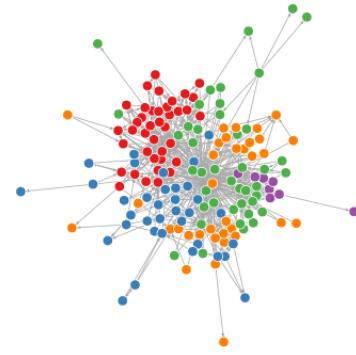
De esto se desprende la importancia de este parámetro en los resultados del modelo. Dado que no se trata de un modelo supervisado, no es posible optimizar el parámetro según alguna medida de bondad de ajuste del modelo. Por el contrario, debemos elegir el punto de corte que resulte más conveniente para los fines del análisis.

En el sentido de esto último, contamos con varias métricas que permiten caracterizar las propiedades del grafo, así como conocimiento previo respecto a propiedades conocidas y ampliamente aceptadas del comercio internacional. De esta forma, para determinar el punto de corte se propone como metodología utilizar una medida de centralidad que describe un atributo previamente conocido del comercio mundial, y buscar aquél punto de corte en que dicho atributo se expresa más plenamente. Una característica ampliamente estudiada del comercio internacional es su regionalismo (Das, 2004), es decir, la tendencia de los países de comerciar en mayor medida con sus vecinos (Head & Mayer, 2014). Si bien no contamos en este trabajo con información respecto de la distancia entre países, de esta característica emerge un fenómeno medible: el nivel de clustering de la red. En este sentido un valor alto en el clustering resulta deseable desde el punto de vista del objeto de estudio. En la figura 2.5(a) se observa el coeficiente de clustering para la red según el punto de corte utilizado, para cada año de la serie entre 1996 y 2016. Allí se observa que dicha métrica tiene una relación negativa no lineal con el punto de corte. Por lo tanto utilizar un punto de corte bajo resulta conveniente para captar el fenómeno. Sin embargo el máximo se encuentra en el punto de corte igual a cero, esto quiere decir que el regionalismo se expresa más cuando consideramos todos los intercambios, independientemente del monto comerciado. Como se observa en la figura 2.5(b), la cantidad de aristas del grafo es significativamente más alta para este punto de corte, y por lo tanto el grafo es sustancialmente más grande y denso, tal como se observa en la figura 2.5(b). En este sentido, si bien en la sección siguiente se continúa estudiando diferentes puntos de corte para ver el efecto que esto produce sobre el grafo y sus posibles interpretaciones, se decidió elegir un punto de corte por defecto de 1 % para el resto del trabajo, dado que en dicho valor existe un punto de inflexión en la densidad y número de aristas, lo cual permite resaltar aquellas relaciones comerciales más significativas, a la vez que se maximiza el coeficiente de clustering.

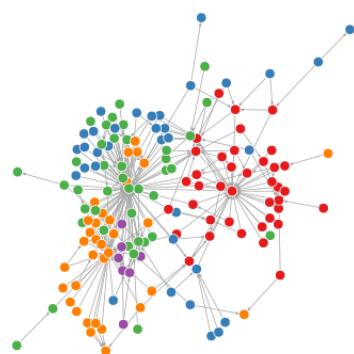
● Asia ● Europe ● Africa ● Americas ● Oceania



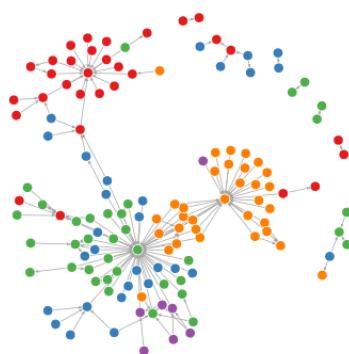
(a) 1 %



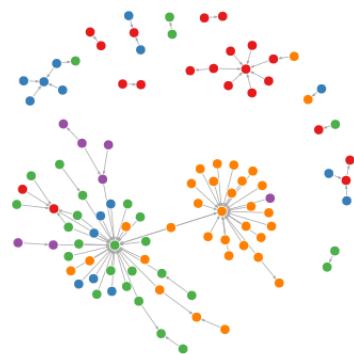
(b) 5 %



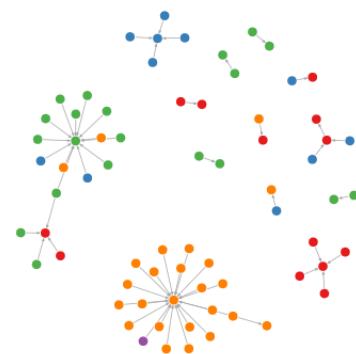
(c) 10 %



(d) 15 %



(e) 20 %



(f) 25 %

Fig. 2.4: Grafos. 2016, según punto de corte

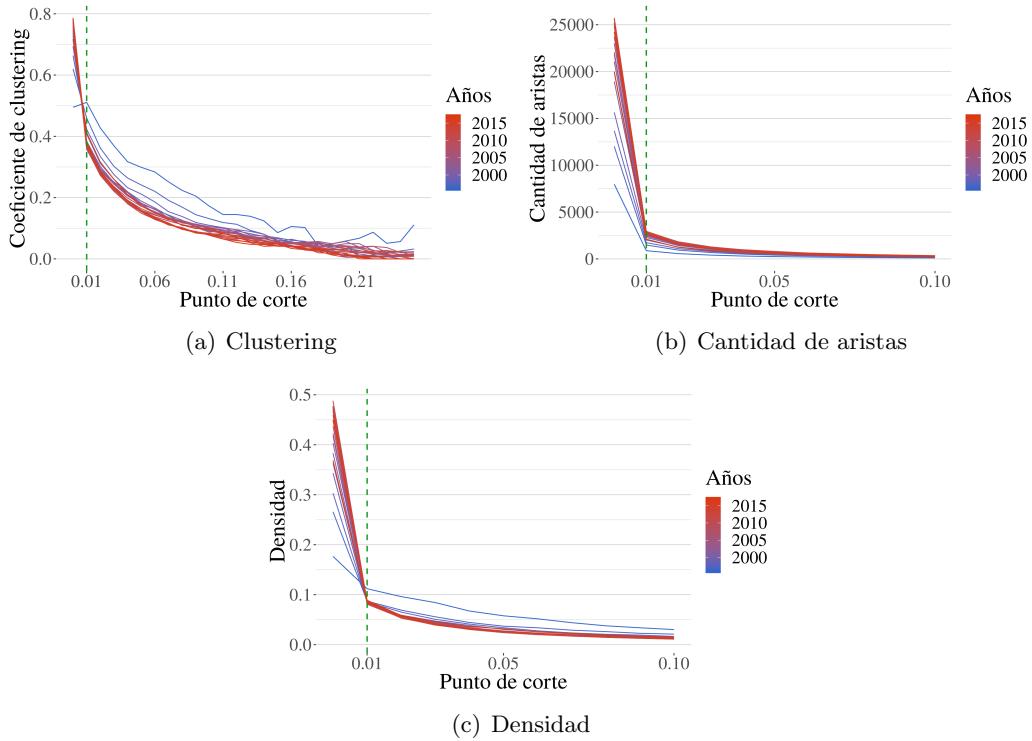


Fig. 2.5: **Medidas de centralidad.** Importaciones, 2016

2.4. Resultados

2.4.1. Correlación entre la representación de las exportaciones y las importaciones

Dada la potencial diferencia en el análisis si se tiene en cuenta las exportaciones y las importaciones, se incorporó esta disyuntiva como una nueva dimensión de estudio. Para ello, se analizó el grado total de cada nodo, para sucesivos puntos de corte, tanto para el grafo que surge de considerar las exportaciones, como para el grafo de las importaciones. En la figura 2.6 se pueden observar los resultados. Vale mencionar que sólo se grafican los nodos de mayor grado, aunque se haya considerado al conjunto de nodos para el cálculo de la tendencia lineal por continente y el coeficiente de correlación de Pearson.

La correlación entre los grafos resultantes oscila entre el 0.72 y 0.96 si se consideran puntos de corte de entre el 1 % y el 25 %, por lo que se puede considerar que al observar las importaciones no se pierde demasiada información. Esto último es especialmente válido para el punto de corte del 1 % que se eligió como estándar.

Sin embargo, surgen algunas diferencias interesantes. La regresión lineal para cada continente, muestra el diferente rol de éstos en el mercado mundial, donde Asia se presenta en el rol de productor de mercancías, mientras que Europa lo hace como un gran consumidor del mercado mundial. El caso americano tiene por un lado el rol preponderante de Estados Unidos, con un comportamiento similar al de Europa (con independencia del umbral del 0.25), mientras que el resto del continente, junto con Oceanía y África, juegan un rol secundario. En el caso europeo, se observa como Rusia tiene un comportamiento más acorde al continente asiático, mientras que en este último se destaca en particular

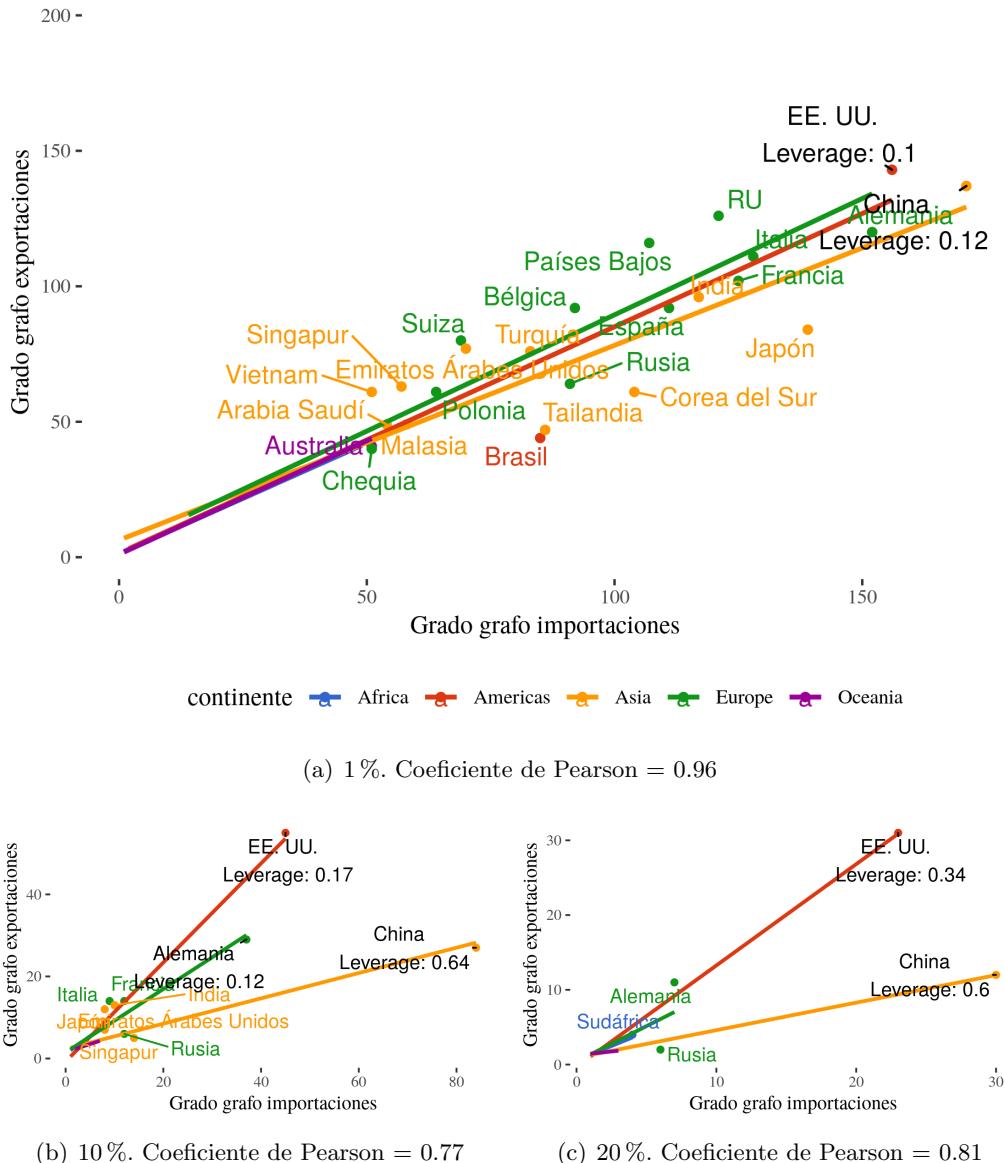


Fig. 2.6: Grado total. Multigrafo exportaciones e importaciones. Año 2016

Tab. 2.1:

Orden	Grafo Importaciones		Grafo Exportaciones	
	País	G °entrada	País	G °entrada
1°	CHN	147	USA	124
2°	USA	139	CHN	114
3°	DEU	130	GBR	103
4°	JPN	117	DEU	98
5°	ITA	109	NLD	98

China como gran productora.

Sin embargo, el rol de cada país varía también respecto al umbral considerado. En el caso de Japón, en puntos de corte bajos, donde hay relaciones comerciales de importancia moderada para los países, se presenta jugando un rol de productor más que de consumidor, incluso más que la media asiática. A medida que se aumenta el punto de corte, y por lo tanto se consideran exclusivamente relaciones de carácter más dependiente, Japón se presenta como un país consumidor, incluso más que los países europeos. Esta transición estaría indicando que Japón, productor de productos industriales y tecnológicos, provee a muchos países del mundo de este tipo de mercancías, aunque éstos no dependen en su consumo de Japón, mientras que como consumidor de materias primas, este país establece relaciones comerciales con países que sí dependen de Japón para poder colocar sus mercancías. Más interesante aún, a medida que se consideran relaciones de tipo más dependiente, aparece como destacado el rol de Estados Unidos, relegando a los países europeos y asiáticos. Por su parte, también surge Sudáfrica como un jugador importante del mercado mundial, cuando se consideran este tipo de relaciones de mayor dependencia. Esto implica que estos dos países tienen una forma de interacción en el mercado mundial marcadamente diferente a la de los países europeos y asiáticos.

Por último, en la figura 2.6 también se puede observar el leverage de los países con mayor apalancamiento para la regresión del grado de las exportaciones respecto del de las importaciones, sin diferenciar por continente. Cuando se considera un punto de corte del 1 %, China tiene el mayor leverage, con un valor algo mayor que Estados Unidos, aunque relativamente bajo en términos absolutos. Sin embargo, a medida que se consideran puntos de corte más altos, este país se constituye como el país con mayor apalancamiento en la regresión, a la vez que el valor aumenta fuertemente.

En la tabla 2.1 se presenta el top 5 de países de mayor centralidad de grado para el 2016, en el grafo de importaciones y en el de exportaciones.

La tabla 2.1 muestra como China, Estados Unidos y Alemania son nodos de una centralidad importante, tanto como consumidores, como en su rol de productores. Sin embargo, China se destaca más en este último sentido, dado que su centralidad en el grafo de las importaciones es mayor, respecto a los demás países, que en el grafo de las exportaciones. Por su parte, Estados Unidos marca una importante hegemonía como consumidor respecto de los demás países, tal como se observa en su centralidad en el grafo de las exportaciones, aunque no conserva tal preponderancia como productor global.

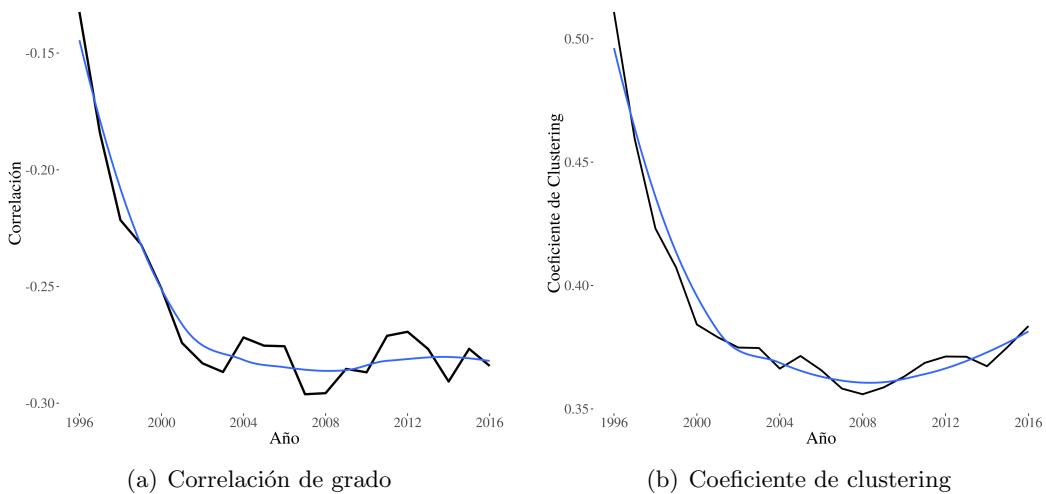


Fig. 2.7: Importaciones. umbral 1%, según año

2.4.2. Evolución de mediano y largo plazo del comercio internacional

Análisis mediano plazo

Lo que resta del análisis se considerará exclusivamente el punto de corte del 1%, incorporando la dimensión temporal. Para esto, se estudiará en primer lugar el movimiento de algunas medidas de resumen a lo largo del tiempo, para luego analizar la distribución de ciertas medidas de centralidad y el cambio en el rol de los nodos centrales en durante el período.

Medidas de resumen de la red

En la figura 2.7 se presentan los valores de algunas medidas de resumen entre el período de 1996 y 2016, considerando un punto de corte del 1% para las importaciones.

Como se puede observar, las características topológicas del grafo varían fuertemente en el período 2007-2009, durante la gran recesión económica, que se caracterizó por un aumento en el proteccionismo económico de ciertos países centrales, reflejado en el “Buy American”¹ o “Buy China”².

La figura 2.7(a) muestra la caída de la correlación de grado en el año 2007, que implica una transición circunstancial hacia una economía menos selectiva. Es decir, un comercio internacional donde aumentan las relaciones de nodos de mayor grado con nodos de menor grado, respecto de las relaciones entre nodos de similar magnitud.

Por su parte, la figura 2.7(b) muestra como el coeficiente de clustering tiene una tendencia decreciente hasta 2008, para luego comenzar a crecer nuevamente. Esto se puede interpretar como una de-segmentación del comercio internacional hasta la crisis, con al cual comienza un proceso de re-segmentación.

Distribución de los nodos

A continuación se presenta la distribución de los nodos según ciertas medidas

¹ <http://edition.cnn.com/2009/POLITICS/02/05/senate.buy.american/index.html>

² <https://economix.blogs.nytimes.com/2009/06/29/buy-china/>

de centralidad a lo largo de los años, destacando aquellos países con mayores valores de centralidad.

En primer lugar, lo que se observa en la figura 2.8 es que la distribución de los nodos de acuerdo a las distintas medidas de centralidad no varía sustancialmente de año en año, ni en su rol como consumidores globales, ni en su rol como productores globales. Por su parte en 2.8(a) se observa la distribución de grado de los nodos del grafo de importaciones, donde se ve como los tres países de mayor grado tienden a alejarse del centro de masa de la distribución. Este podio se disputa entre Alemania, Japón, Estados Unidos y, a partir del 2002, China, quien a su vez tiende a alejarse de éstos posteriormente. Por su parte, la figura 2.8(b) marca el rol de consumidores globales. Aquí, mientras hasta 2007 Estados Unidos y Gran Bretaña ocupan los lugares principales en la centralidad de autovalor, a partir de este año comienza a tener un rol más destacado Francia, los países bajos, y luego China.

Es de resaltar el aumento en la centralidad de China para el comercio internacional, tanto como productor como consumidor. En particular, este país desplaza a Estados Unidos de ser el nodo de mayor centralidad en muchas medidas. Por ejemplo, la figura 2.9 para el caso del autovalor ponderador por el valor del intercambio, Estados Unidos ocupa el primer lugar hasta el año 2006, para luego ser reemplazado por China, un país que diez años antes tenía un valor de centralidad cuatro veces menor al de Estados Unidos. En ese mismo año China también pasa a ocupar el primer lugar en dicha métrica para el grafo de las importaciones. La medida de centralidad del autovalor es particularmente relevante, sobretodo todo ponderada por el valor de las transacciones, porque considera no sólo la importancia del nodo, sino también la importancia de aquellos nodos con que se relaciona. En este sentido, se destaca el rol de México, Japón, Canadá, junto con China y Estados Unidos como grandes productores mundiales, a la vez que aparece, sobre los últimos años, Corea como una nueva potencia en este aspecto.

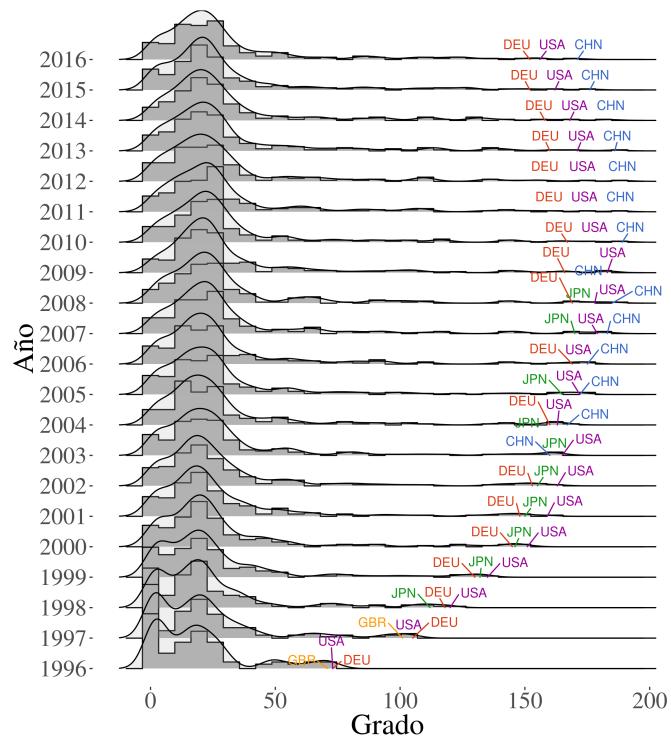
Análisis de largo plazo

Otro análisis que resulta interesante es la evolución de la red en un período de largo plazo. Esto nos permite encontrar los cambios estructurales del comercio internacional, a la vez que da cuenta de qué elementos de la topología de la red son robustos, y cuales contienen una tendencia histórica. Para esto, tomamos la serie elaborada por Gleditsch (2002), que reúne información del comercio internacional entre 1948 y 2000. A continuación se analizan los elementos estructurales de la red, para luego analizar la evolución de algunos países seleccionados dentro del grafo

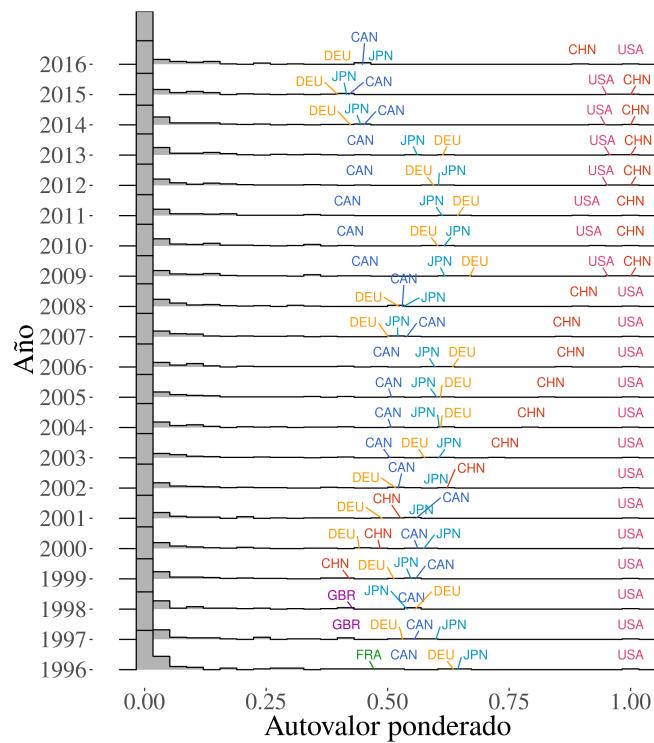
Estructura de la red

Como se observa en la figura 2.10, en la red ocurren importantes cambios en el período que va de 1948 al año 2000. La figura 2.10(a) muestra la evolución del coeficiente de clustering de la red. Este, como se mencionó previamente, indica la tendencia de la red a formar vecindarios bien delimitados. La fuerte caída durante las décadas del cincuenta y sesenta reflejan la nueva división internacional del trabajo, en dónde las relaciones comerciales se alejan del regionalismo, y el comercio inter-continental aumenta en volumen. En concreto, esta expresando la incorporación de Asia como productor de mercancías para el mercado mundial.

La figura 2.10(b) muestra la evolución del coeficiente de correlación de grado. Este indica el grado de homofilia de la red, es decir, la tendencia de los nodos de mayor grado a



(a) Grado, importaciones



(b) Autovalor, exportaciones

Fig. 2.8: Distribución medidas de centralidad según año. umbral 1 %

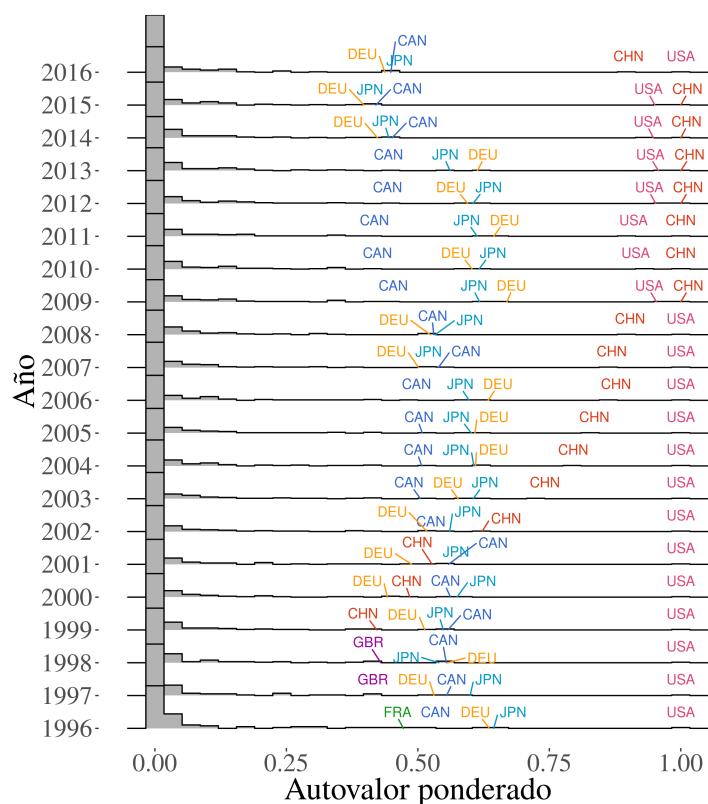


Fig. 2.9: Distribución autovalor ponderado, exportaciones según año. umbral 1 %

conectarse con otros nodos de similares características. Recordemos que al tomar valores negativos, esta correlación indica una tendencia de los países de mayor grado a relacionarse con países de menor grado. La caída de este valor en las décadas del cincuenta y sesenta expresa un incremento en esta característica de la red.

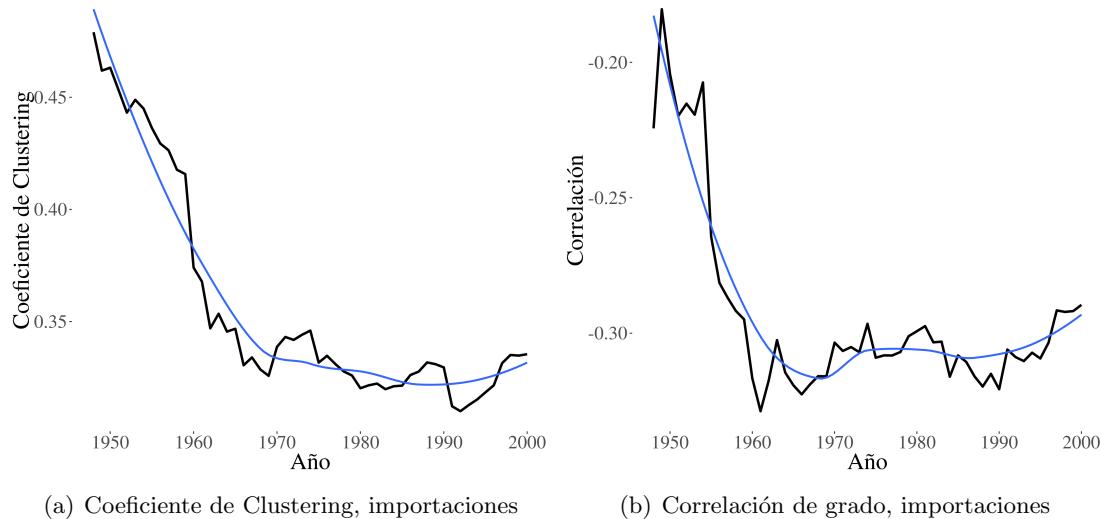


Fig. 2.10: Evolución de la estructura de la red. Impo. Umbral 1%

Estas medidas refieren a características generales del grafo. También es posible analizar, dentro del grafo, el movimiento en la centralidad de subconjuntos del mismo. En particular, en la figura 2.11 se muestra la evolución de la centralidad promedio por continente para las medidas de autovalor, cercanía y grado. En primer lugar, resalta la robustez de la centralidad de cercanía, que evoluciona uniformemente para todos los continentes. La disminución de esta medida indica un crecimiento del tamaño de la red³.

La centralidad de autovalor, por su parte, muestra un cambio importante en el rol de los países en la economía mundial. Mientras que Oceanía pasa del segundo al quinto lugar en el período, Asia hace lo opuesto, marcando la incorporación al mercado mundial de este continente como productor de mercancías. África y América, por su parte, si bien siguen un distinto recorrido a lo largo del período, comienzan y terminan el mismo con valores muy similares. Europa mantiene la mayor centralidad promedio a lo largo de todo el período, en las tres métricas. Por último, el recorrido de Asia y Oceanía es similar para la centralidad de grado.

Evolución de los países

Otra característica interesante del análisis de grafos es que permite seguir la evolución de un país en particular dentro del mismo, de forma tal que se puede observar la evolución de la centralidad de algunos nodos en el tiempo.

En la figura 2.12 se muestra la distribución de las medidas de grado y de autovalor ponderado, y se marca la posición de México, Brasil y Argentina en dicha distribución. En 2.12(a) observamos la cantidad de conexiones de cada país en el grafo de las importaciones. Aquí se observa como México comienza la serie en la moda de la distribución de grado,

³ Que parcialmente puede deberse a la incorporación de nuevas fuentes de información, en particular para los años 1949, 1960 y 1991

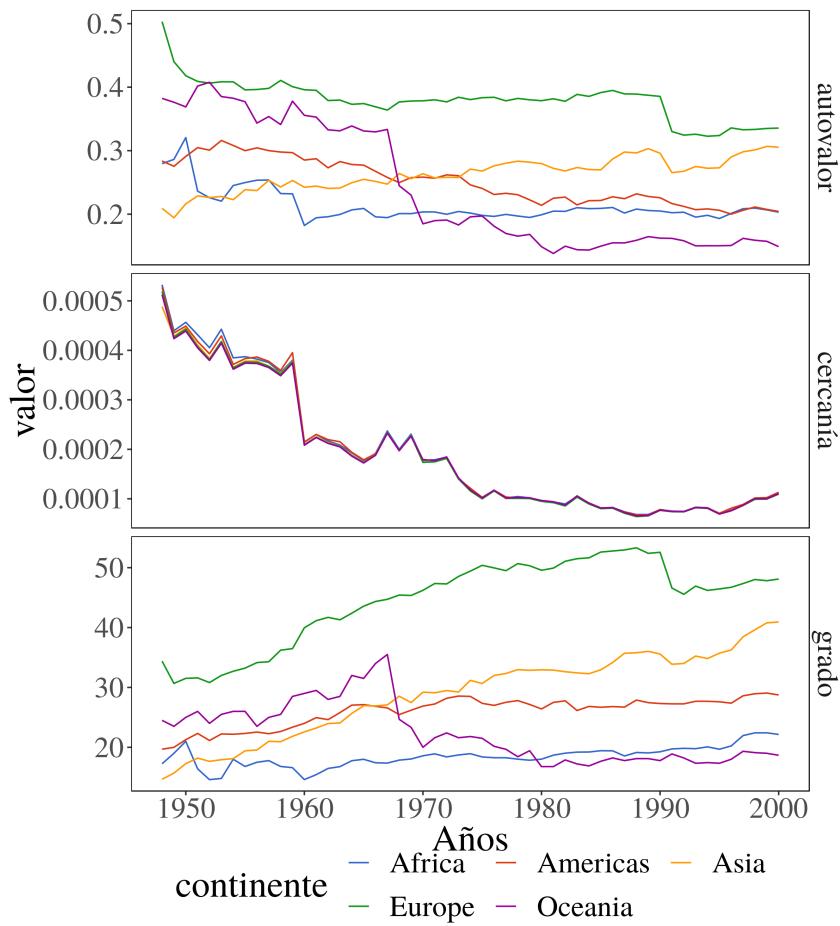


Fig. 2.11: Evolución de la centralidad promedio por continente. Impo. Umbral 1%

mientras que Brasil y Argentina se encuentran a la derecha de la misma. En la evolución se vé como la distribución se convierte en más leptocúrtica, y, por un lado Brasil incrementa la cantidad de nodos considerablemente, mientras que por otro lado Argentina mantiene a lo largo de toda la serie la misma cantidad de socios comerciales. Por su parte, México, pasa a ubicarse en la cola derecha de la distribución, más cercano a la Argentina. En 2.12(b) se observa el autovalor ponderado por el volumen comerciado. Considerando esta medida, la evolución pareciera ser diferente. Mientras Argentina se mantiene en el mismo lugar de la distribución, Brasil se repliega hacia valores similares a los Argentinos, y México se despega de la moda de la distribución a partir de los años noventa. Al considerar no solo el grado, sino también el grado de los vecinos, y el volumen comerciado, se destaca el proceso de la *Maquila Mexicana*, en dónde el comercio entre este país y Estados Unidos, un nodo sumamente importante, es muy difundido.

Por último, resulta interesante complementar el análisis de la figura 2.9, donde se veía como China supera a Estados Unidos en ciertas medidas de centralidad para el año 2006, con un análisis de dicha relación en el período de 1948 a 2000. Para ello, en la figura 2.13(a) se observa la distribución de la centralidad de autovalor, y el detalle de estos dos países.

En todo el período se observa la predominancia de Estados Unidos como un nodo

central del grafo. China, por su parte, comienza en la moda o incluso en la cola izquierda de la distribución. Para luego ubicarse del lado derecho de la misma, entre los sesentas y setentas, aunque aún con valores relativamente bajos. A partir de la década del ochenta, comienza, de forma lenta pero acelerada, a escalar posiciones de mayor centralidad dentro de la red. Por su parte, en la figura 2.13(b) se observa la evolución de China junto a Corea y Japón. Aquí se puede ver como no es solo China quién aumenta la centralidad en el grafo, sino Asia en su conjunto. Japón lo hace primero, desde el comienzo de la serie, en los cincuentas, y llega a valores cercanos a 1, el máximo valor posible de esta medida, para los ochentas. Por su parte, Corea también sigue el mismo camino, aunque comienza más tarde, y desde una posición de menor centralidad que los otros países.

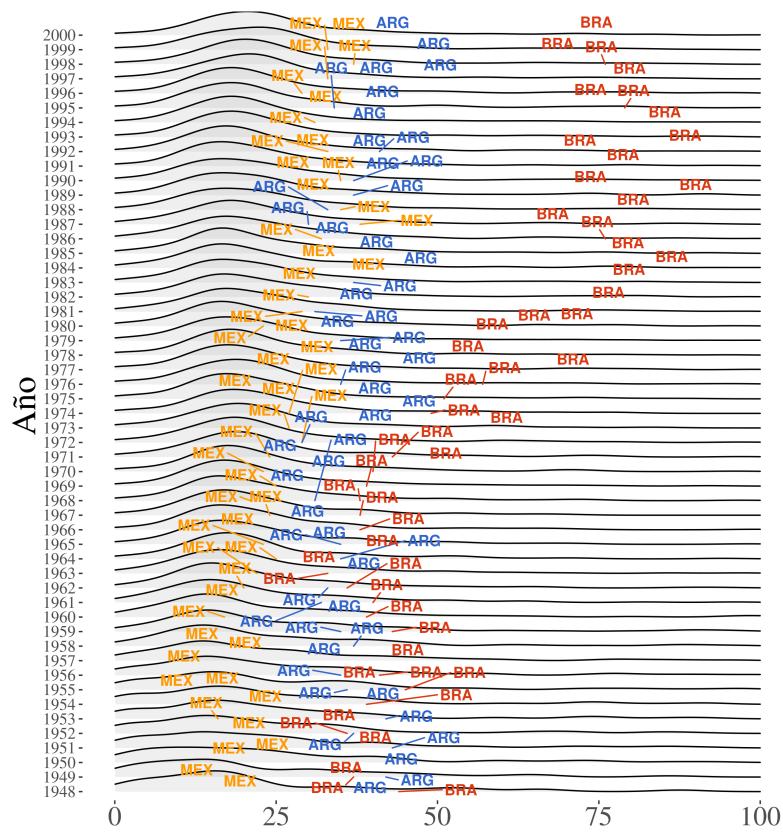
2.5. Conclusiones

En la presente sección se propuso una primera aproximación de la teoría de grafos como herramienta para la caracterización del comercio internacional, utilizando los datos agregados a nivel país. Para ello se construyó un grafo dirigido no ponderado. Se utilizó al año 2016 como punto de referencia para determinar determinar los mejores hiperparámetros del modelo. En particular, de este análisis resultó que el 1% constituye un umbral posible para el punto de corte, aunque considerar este hiperparámetro como una dimensión de estudio también permite observar las distintas formas de relación comercial entre los países, y por lo tanto se logra enriquecer la discusión.

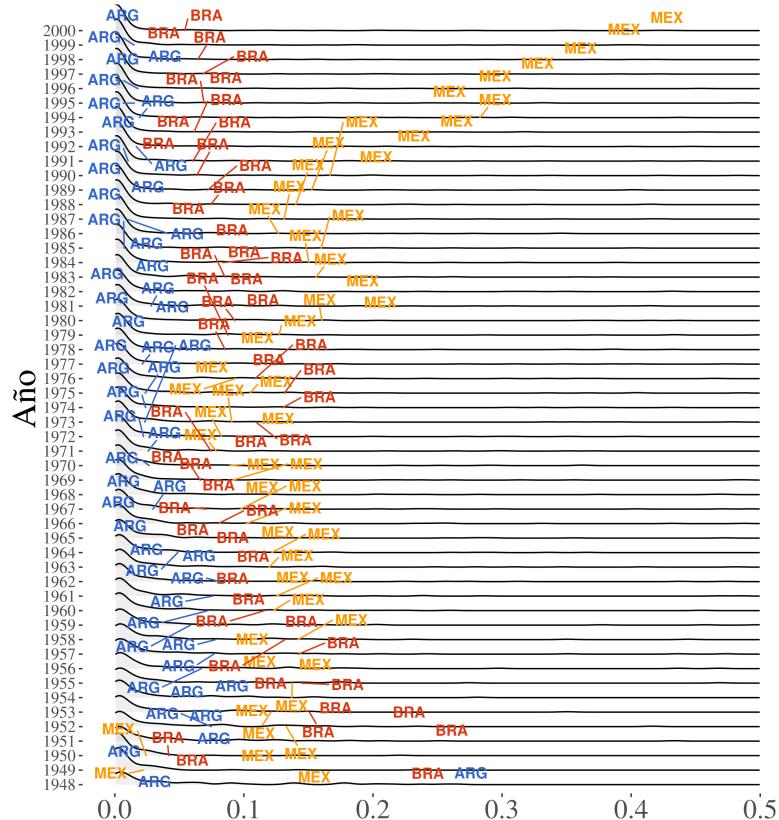
Por su parte, la dimensión temporal en un período de veinte años mostró que las medidas de resumen de un grafo pueden poseer un potencial para la descripción de la crisis del año 2009. El análisis de una serie de largo alcance, por su parte, permite ver los cambios estructurales del comercio internacional.

Se observó también que la distribución de las medidas de centralidad no pareciera variar en el mediano plazo, aunque sí existe un movimiento visible en los actores principales de la red, lo cual muestra una riqueza en el análisis para describir los cambios de la economía mundial, y los roles cambiantes que en esta juegan los distintos recortes nacionales.

Finalmente, el trabajo realizado deja una contradicción planteada respecto del rol de ciertos países como consumidores, y de otros como productores, dado que tal estructura no sería sostenible de forma prolongada.



(a) Grado



(b) Autovalor Ponderado

Fig. 2.12: Argentina, Brasil y México. Importaciones. Umbral 1%

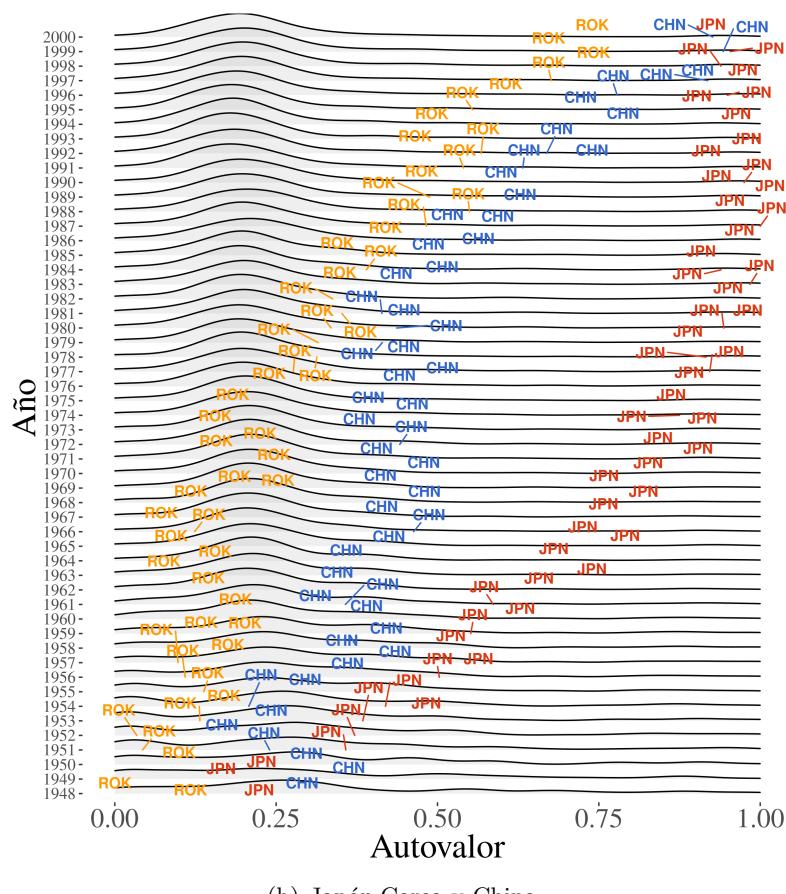
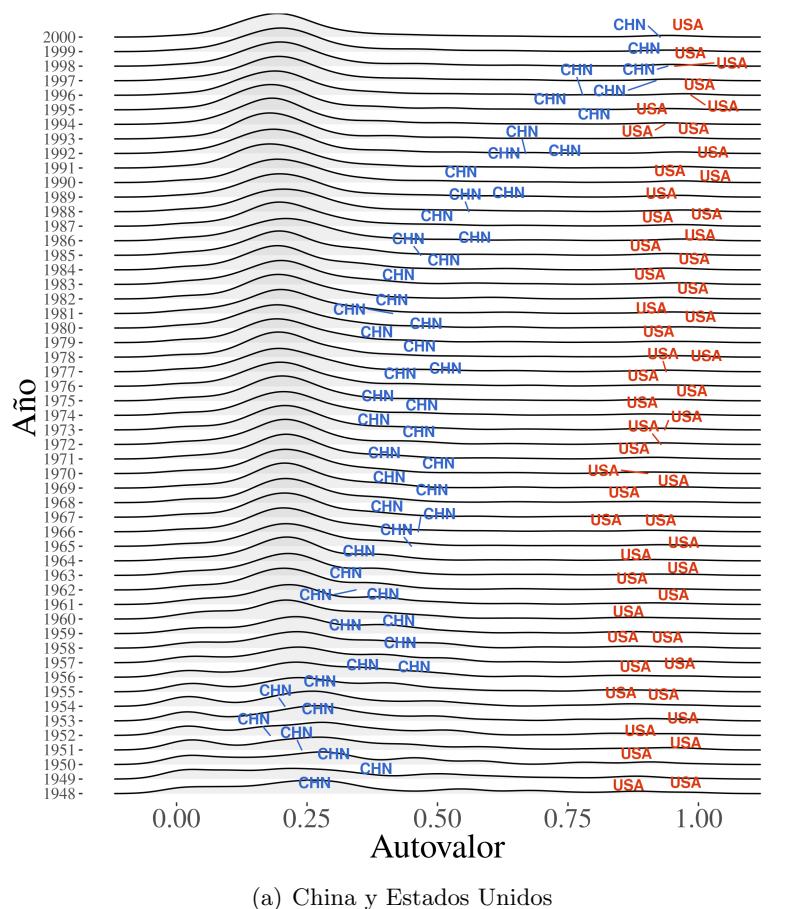


Fig. 2.13: Autovalor. Importaciones. Umbral 1 %

3. COMERCIO A NIVEL PRODUCTO

3.1. introducción

incluye esta-
do del arte

3.2. Análisis exploratorio de datos

El análisis exploratorio de la información desagregada a nivel producto implica un mayor nivel de complejidad, dado que se incorpora una nueva dimensión, de alta cardinalidad, al estudio. Dado que la información a nivel de comercio agregado entre países ya fue realizada en el capítulo previo, en este análisis exploratorio se realizará foco en comprender la composición de las canastas exportadoras e importadoras de los países.

Dado que los nomencladores de productos en su nivel desagregado implican una alta cardinalidad, el estudio de su distribución resulta inabarcable. Por ello, tradicionalmente el análisis económico recurre a diferentes niveles de agregación. Para realizar el análisis exploratorio de esta sección recurrimos al nomenclador elaborado por Molinari & de Angelis (2016) a partir del concepto de *cadenas productivas*. Este nomenclador tiene dos niveles de agregación: Cadenas y Subcadenas (ver Apéndice).

La figura 3.1 muestra la distribución de las exportaciones según Cadenas, Subcadenas y Usos, para el promedio mundial durante el período 1996-2016. En 3.1(a) se puede ver que las dos cadenas más importantes son las de Bienes de Capital y Otras Manufacturas, dentro de las cuales las subcadenas que destacan son los equipos eléctricos y de transporte. Les sigue la cadena de insumos difundidos, donde destacan los metales y químicos. También son importantes las subcadenas de autopartes, autos, y dentro de la cadena de combustibles, el gas y petróleo. En la figura 3.1(b) las cadenas se subdividen según el uso que se le da a los productos: Productos Primarios, Bienes de Capital, etc (ver Apéndice). Vale mencionar que la cadena Bienes de Capital no incluye exclusivamente productos de este tipo, tal como se observa en la figura, dado que las cadenas hacen referencia a las Cadenas de valor (Humphrey & Schmitz, 2000), es decir a la rama de la producción a la que pertenecen los productos del final de la cadena. Allí podemos observar que en la cadena de Otras manufacturas destacan los productos semiterminados y bienes de consumo, mientras que en las cadena de Bienes de capital destacan su homónimo y las partes y componentes. A su vez en la cadena de insumos difundidos destacan los productos semiterminados ampliamente, mientras que en la industria automotriz se exportan mayoritariamente partes y componentes.

La figura 3.2 muestra la distribución de las exportaciones e importaciones según continente y país exportador, para el promedio 1996-2016. Allí destaca el mayor volumen de exportaciones que importaciones de Asia, y el mayor volumen de importaciones que exportaciones de Estados Unidos, siguiendo las mismas conclusiones que el capítulo 2.

Sin embargo, la riqueza en el análisis a nivel producto se encuentra en comprar la distinta composición de la balanza comercial de los países. Ya sea la diferencia entre la composición de sus exportaciones respecto de sus importaciones, como de estos elementos respecto de otros países o regiones. Dado que dicho análisis implica la comparación de una multiplicidad de treemaps, se decidió elaborar una herramienta interactiva para el análisis.



(a) Cadenas y Subcadenas



(b) Cadenas y Usos

Fig. 3.1: Treemaps por tipos de productos. Exportaciones. 1996-2016. Total mundial

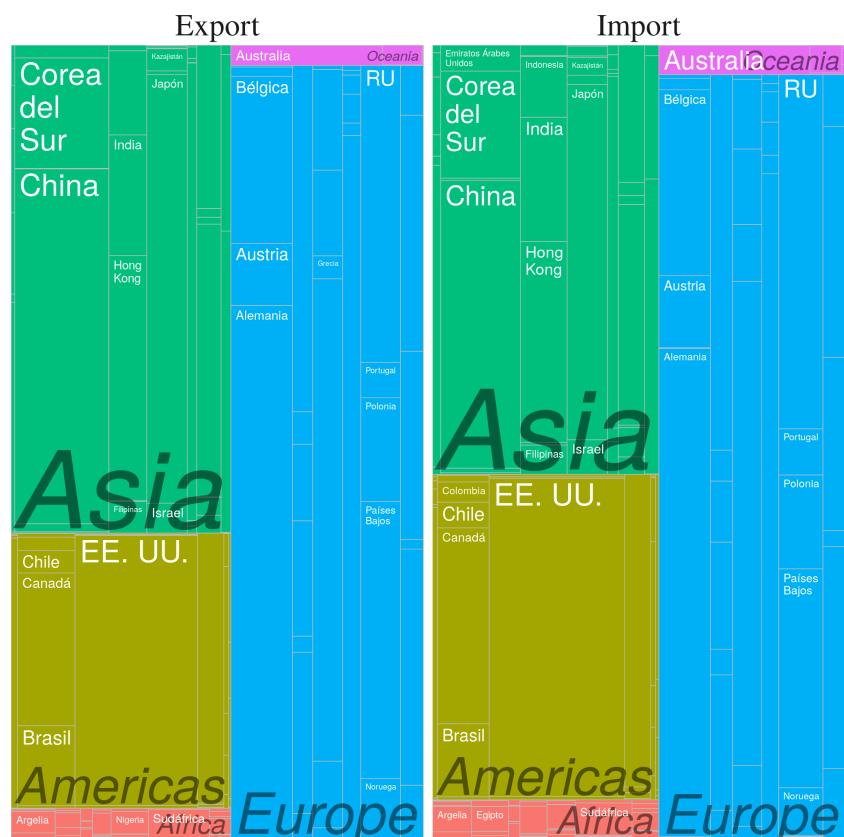
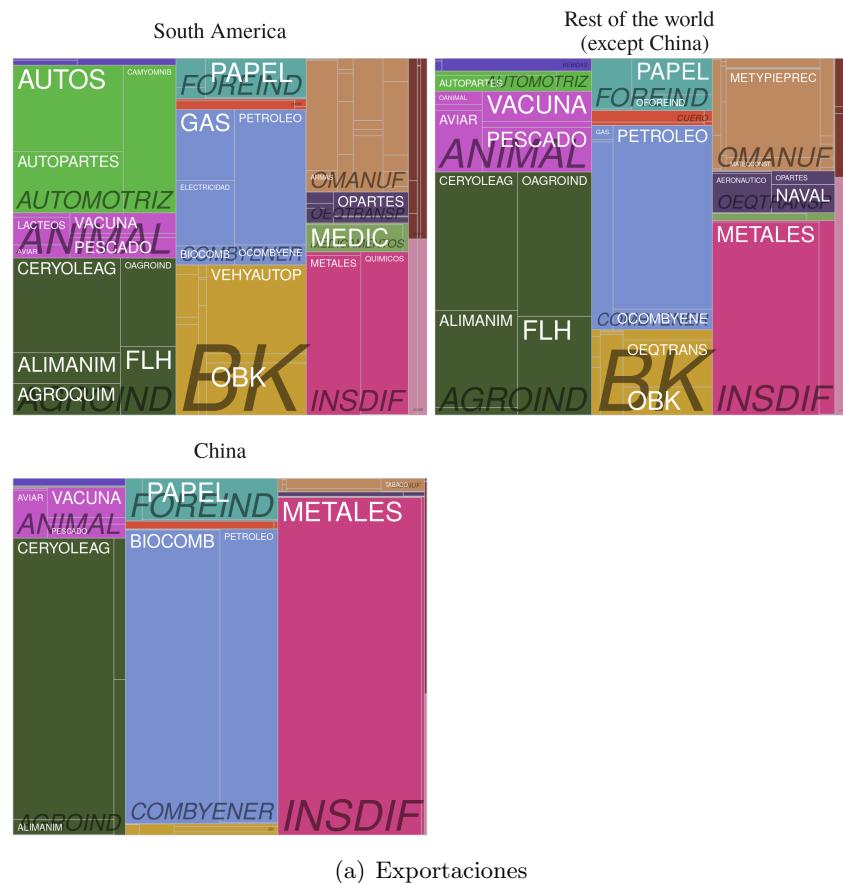


Fig. 3.2: Treemaps por Continentes y países. 1996-2016. Total mundial

La misma fue elaborada utilizando la librería *shiny* (Chang et al., 2018) y se encuentra disponible en <https://treemaps.shinyapps.io/treemaps/>. Dados los límites del servidor y el interés de estudiar la herramienta para un caso específico, el análisis se centra en el caso latinoamericano, tanto para el subcontinente en conjunto como para los países que lo integran. El objeto de análisis es estudiar las posibilidades de integración regional de la producción, y para ello analizar el comportamiento diferencial de las canastas exportadoras e importadoras de los países latinoamericanos cuando comercian entre sí respecto de su comercio con el resto del mundo. A su vez, dada la importancia del comercio de esta región con China, se decidió dividir la información del resto del mundo exceptuando a China y este país de forma individual.

En la figura 3.3 se pueden observar los treemaps de cadenas y subcadenas para 2016 del total de los países sudamericanos, según su destino. En la figura 3.3(a) se ve como la canasta exportadora de los países latinoamericanos varía según si su destino se encuentra dentro o fuera de la región, y en particular si son exportaciones hacia China. En particular, el comercio intra-regional tiene un componente importante de bienes de capital, y del sector automotriz, ya sean autos terminados o autopartes. Por su parte, en el comercio con el resto del mundo estos componentes cumplen un rol secundario, mientras que se destacan las cadenas agroindustriales, de combustibles e insumos industriales. Respecto del resto del mundo, el comercio con China resalta por las exportaciones de la subcadena de biocombustibles, cereales y oleaginosas y Metales. Las importaciones de estos mismos países se pueden observar en la figura 3.3(b). Naturalmente los treemaps de las exportaciones e importaciones son muy similares, ya que la única diferencia es el país que registra la operación. Sin embargo, en las importaciones del resto del mundo destacan las cadenas automotriz y de bienes de capital. A su vez, resulta de interés el cambio de composición de las cadenas: Mientras que las exportaciones de insumos industriales hacia el resto del mundo son mayoritariamente metales, en las importaciones destacan los productos químicos. Por su parte, mientras que en la cadena de otras manufacturas se exportan hacia el resto del mundo metales y piedras preciosas, se importan dentro de esta cadena equipos eléctricos. A su vez, del resto del mundo se importan medicamentos, aunque este rubro no aparece en el treemap de las exportaciones. Cabe destacar que la región exporta e importa petróleo hacia el resto del mundo, aunque el comercio intrarregional de este producto es menor. Existe, por lo tanto, una potencialidad de integración comercial dentro de la región para este producto en particular. El comercio con China está particularmente orientado a la venta de materias primas y la compra de productos industriales. Destacan especialmente las exportaciones de cereales y oleaginosas, biocombustibles y metales; mientras que se importan equipos eléctricos, bienes de capital, calzados y químicos.

La figura 3.4 muestra los treemaps exclusivamente según los usos de los productos, para el total de la región, según su destino, en el año 2016. En 3.4(a) se puede observar las exportaciones. Los usos de los productos exportados varían fuertemente según su destino. Los productos semiterminados constituyen la mayoría de las exportaciones tanto hacia el interior de la región, como con el resto del mundo. Sin embargo, mientras que los bienes de consumo resultan de particular importancia dentro de la región, los productos primarios lo son respecto del resto del mundo. El comercio con china destaca esta tendencia, donde la mayoría de las exportaciones provienen de productos primarios. En 3.4(b) se observa la distribución de las importaciones según su uso. Allí se ve que las importaciones de productos semiterminados desde el resto del mundo supera a las exportaciones, si se lo compara con 3.4(a). A su vez, prácticamente no se exportan productos primarios, pero sí partes



(a) Exportaciones



(b) Importaciones

Fig. 3.3: Treemaps de Cadenas y Subcadenas.2016. Total Latinoamérica

y componentes, y bienes de capital. El comercio con China se encuentra equitativamente distribuido entre bienes de capital, productos semiterminados, de consumo y partes y componentes. Es notorio que no se importa prácticamente productos primarios, aunque estos constituyen la base de las exportaciones a dicho país.

En la figura 3.5 se observa la distribución de los socios comerciales de Sudamérica en el 2016. Esto es, los países destino de las exportaciones y origen de las importaciones. En 3.5(a) se excluye al resto del mundo y China, para observar la distribución interna del comercio. Naturalmente Brasil es el mayor socio comercial interno, para ambos tipos de flujo comercial, aunque su importancia es mucho mayor como exportador que como importador, dado que es el origen del 40 % de las importaciones de los demás países, y el destino del 25 % de las exportaciones. En la figura 3.5(b) se puede observar el mismo gráfico, pero incluyendo al resto del mundo. Allí se ve que tanto las exportaciones como importaciones hacia el resto del mundo exceptuando a China constituyen más de un 60 % del comercio, y China exclusivamente representa casi un 20 % del comercio de la región, superando ampliamente a Brasil.

De las figuras 3.3, 3.4 y 3.5 se desprende que la región de Sudamérica tiene una baja integración regional en tanto una pequeña proporción de su comercio se realiza entre países de la región. A su vez, mientras el comercio intrarregional se realiza fundamentalmente sobre productos de alta complejidad y valor agregado, el comercio con el resto del mundo se encuentra desbalanceado: mientras las exportaciones se concentran en productos primarios de baja complejidad, aquellos productos que llegan desde el resto del mundo tienen un mayor grado de tecnificación. Estas conclusiones son particularmente válidas para el comercio con China, que constituye una quinta parte del comercio total de la región. Vale mencionar que la importancia de dicho país ha crecido de forma sostenida en el período analizado, siendo que en 1996 representaba menos de 2 % del comercio ¹

3.3. Metodología

Del análisis exploratorio se desprende el potencial de la información desagregada a nivel producto para caracterizar la inserción en el mercado mundial de un país o región. Sin embargo, dada la alta cardinalidad de la información y la múltiples dimensiones de estudio, resulta dificultoso su estudio de forma general, sin hacer eje en un determinado país, o sin agregar la información según un nomenclador. En particular, el uso de agrupaciones jerárquicas de los productos resulta esencial para el análisis de resultados. La elaboración de los mismos constituye una extensa tarea por parte de expertos en las temáticas sectoriales de los diferentes tipos de productos, y el nomenclador resultante está fuertemente determinado por los objetivos con que será utilizado. Es por ello que en la presente sección se propone una elaboración alternativa de niveles jerárquicos de agrupamiento de productos.

agregar grafo
bipartito

3.3.1. Latent Dirichlet Allocation Models

El objetivo de la presente sección es elaborar un agrupamiento automático de los productos, basado en la información disponible. Este problema puede ser concebido desde dos puntos de vista: por un lado, se puede pensar como un problema de *clustering* donde lo

¹ ver <https://treemaps.shinyapps.io/treemaps/>

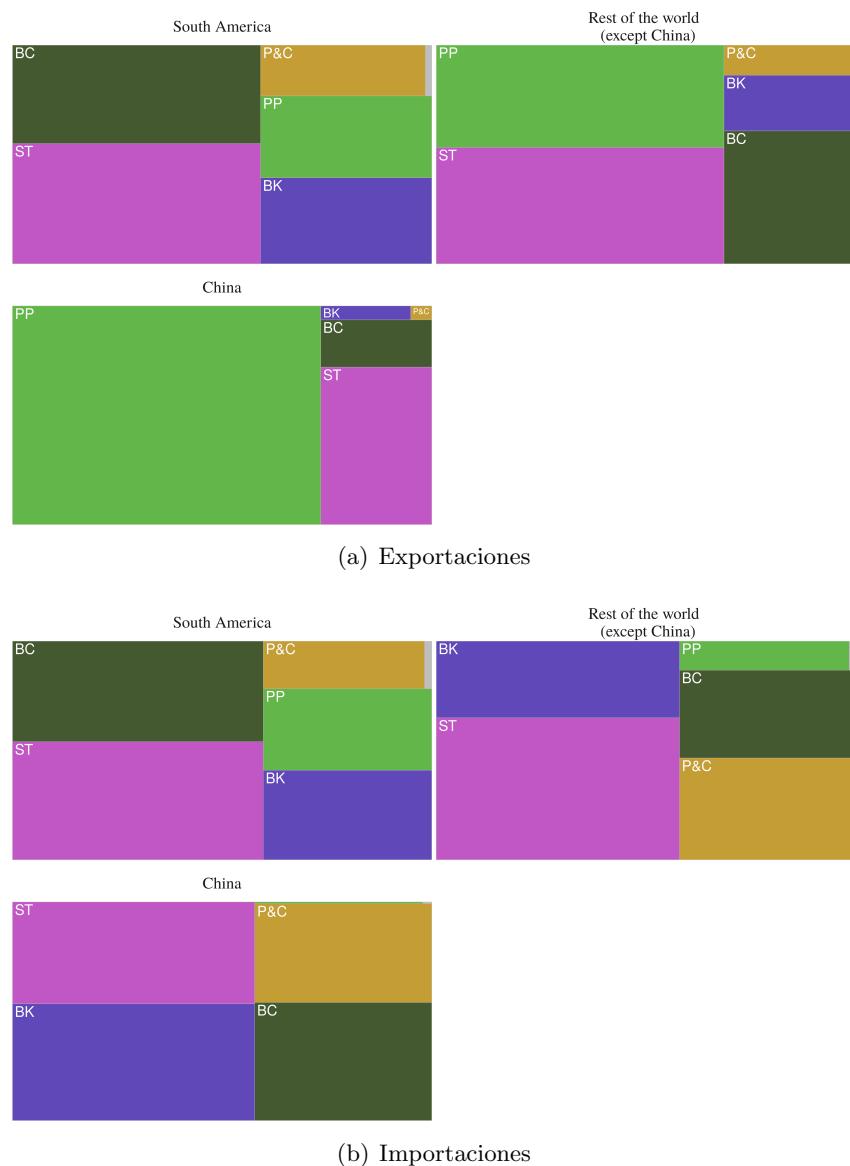
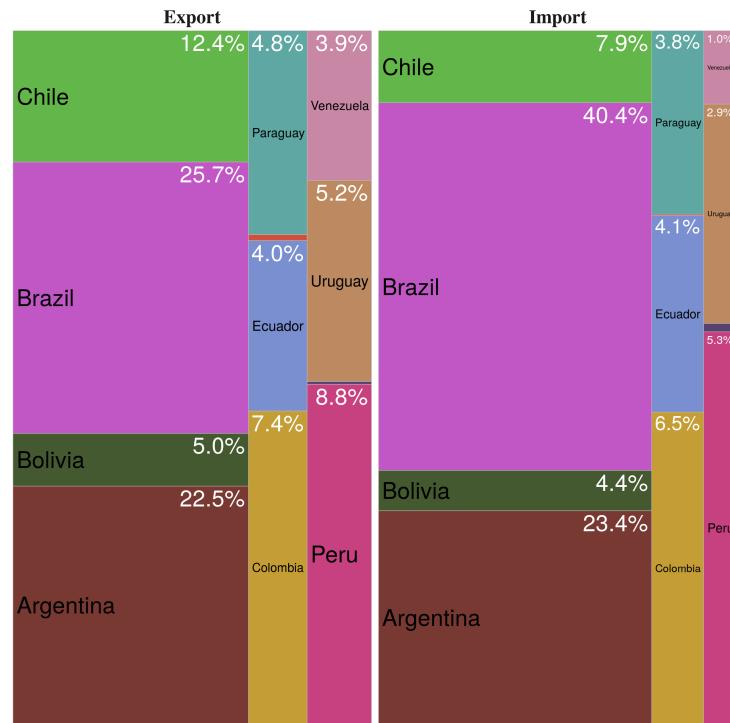
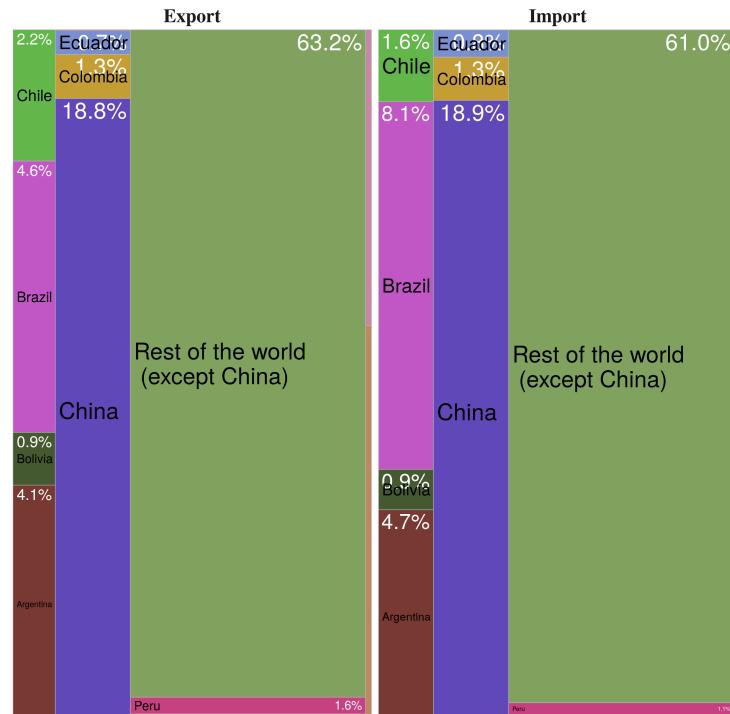


Fig. 3.4: Treemaps de Cadenas y Usos.2016. Total Latinoamérica



(a) Excluyendo Resto del Mundo



(b) Incluyendo Resto del mundo

Fig. 3.5: Treemaps según socio comercial. 2016. Total Latinoamérica

que se busca es crear grupos de productos con similares características. Por otro lado, es un problema de reducción de dimensionalidad, donde lo que se busca es encontrar un espacio de menor dimensión al que existen originalmente los datos. Esto sería posible dado que existe la posibilidad de explotar las similitudes y diferencias entre los distintos productos.

Sin embargo, las técnicas de clustering tradicionales encuentran problemas en espacios de alta dimensionalidad (Aggarwal et al., 2001). A su vez, siguiendo el trabajo de Molinari & de Angelis (2016), se entiende que los grupos no deberían ser excluyentes, dado que un mismo producto puede ser utilizado en diferentes formas, como producto intermedio o final. En este sentido, el problema se puede especificar como de *clustering difuso*.

En concreto, la dimensionalidad del problema se puede pensar como el siguiente espacio:

$$\mathcal{R}^{N*P*Y*2}$$

Es decir la interacción de N países, P productos, Y años, tanto para las exportaciones como las importaciones. La propuesta es utilizar la técnica propuesta por Blei et al. (2003) conocida como *Latent Dirichlet Allocation Models* o *Topic Modeling*. En su versión original, esta técnica se propone como una forma de encontrar los tópicos presentes en un corpus, y la distribución de dichos tópicos sobre cada texto. Este problema es análogo al que se busca en el presente trabajo: Allí se busca una dimensión latente de k tópicos, embebidos en un diccionario de alta dimensionalidad (las palabras presentes en el corpus), que se distribuyen a lo largo de los textos que componen dicho corpus. En el presente problema, se busca una dimensión latente de k *componentes*, embebidos en un nomenclador de alta dimensionalidad, que se distribuyen a lo largo de los países. Esta técnica a su vez puede ser pensada como un problema de clustering difuso, en tanto cada palabra (en el contexto original) puede pertenecer a más de un tópico.

A continuación se realizará una descripción del modelo propuesto por Blei et al. (2003), adaptado al presente dominio.

Definiciones

- Un **producto** es la *unidad básica discreta de los datos*, se define como un ítem de un nomenclador (SITC). Se representa como un vector unitario. Definimos el superíndice i del vector como el i-ésimo producto del nomenclador y el i-ésimo elemento del vector. El V-ésimo producto del nomenclador es el vector w , tal que $w^v=1$ y $w^u=0$, $u \neq v$
- Un **país** es una secuencia de \mathbf{N} productos, definido como $W = (w_1, w_2, \dots, w_N)$
- Nuestro **corpus** es la colección de \mathbf{M} países, definido como $D = (d_1, d_2, \dots, d_M)$
- Un **componente** es una dimensión latente sobre el corpus, y suponemos una cantidad fija k de los mismos.
- Nuestro objetivo es obtener:
 1. Una distribución de componentes sobre cada país.
 2. Una distribución de los productos sobre los componentes.

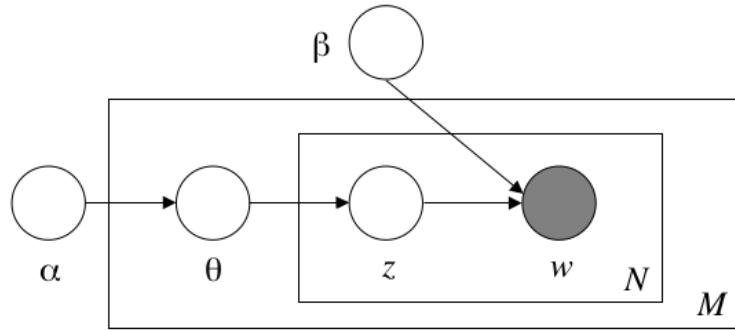


Fig. 3.6: fuente: Blei et al. (2003)

Proceso generativo e inferencial

Intuitivamente, suponemos el siguiente proceso generativo de datos:

- Para cada país del corpus, imaginamos que las exportaciones surgen de un proceso de dos etapas:
 - Elegimos aleatoriamente una distribución sobre los componentes
 - Para cada dólar exportado:
 - Elegimos aleatoriamente el componente al que pertenece, dada la distribución definida en el paso anterior
 - Elegimos aleatoriamente un producto de la distribución correspondiente a dicho componente
1. Para cada Componente $k \in \{1, 2, \dots, K\}$
 - Generar una distribución sobre los componentes $\beta \sim Dir_v(\eta)$ con $\eta \in \mathcal{R}_{>0}$ un parámetro fijo
 2. Para cada país $d \in \{1, 2, \dots, D\}$
 - generar un vector de proporciones de componentes $\theta_d \sim Dir_K(\alpha)$ con $\alpha \in \mathcal{R}_{>0}^K$ un parámetro fijo
 - Para cada dólar exportado:
 - I. generar una asignación del componente $z_{dn} \sim Mult(\theta_d)$
 - II. asignar el producto $w_{dn} \sim Mult(\beta_{zn})$

En la figura 3.6 se puede observar el proceso generativo de forma gráfica. Aquí, cada nodo representa una distribución de probabilidad. Las aristas significan que la distribución de salida define los parámetros de la distribución de entrada. Los recuadros significan replicación: El recuadro interior representa que el proceso se realiza para cada dólar exportado en el país. El recuadro exterior representa que el proceso se realiza para cada país en el corpus.

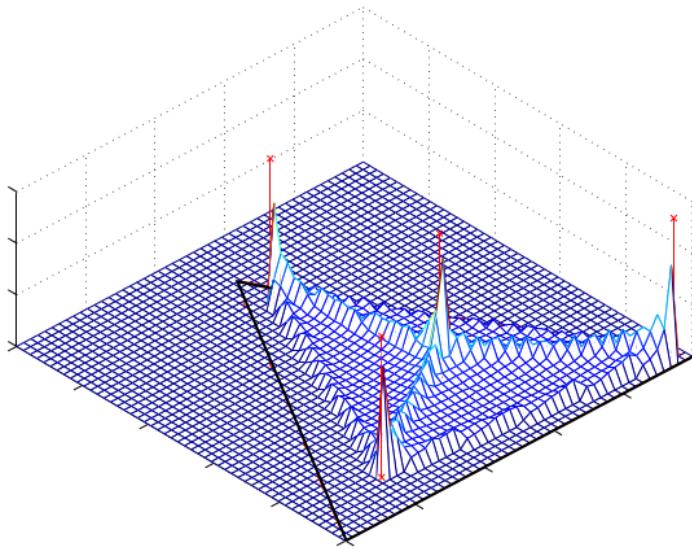


Fig. 3.7: fuente: Blei et al. (2003)

Un Proceso de Dirichlet es una familia de procesos estocásticos donde las realizaciones son ellas mismas distribuciones de probabilidad. Es decir, el rango de esta distribución (así como en una normal son los reales) son distribuciones de probabilidad. Para interpretarlo geométricamente, la figura 3.7 muestra un ejemplo de la distribución de densidad para 3 productos y 4 componentes. El triángulo representa todas las distribuciones (multinomiales) posibles sobre los tres productos. Cada uno de los vértices del triángulo es una distribución de probabilidad que asigna una probabilidad de 1 a uno de los productos. El punto medio de cada lado, es una distribución con probabilidad 0.5 a dos componentes. El cuarto componente, el centroide del triángulo, asigna probabilidad de $\frac{1}{3}$ a cada producto. Los cuatro puntos marcado con x son las distribuciones multinomiales de $p(w|z)$ para cada uno de los cuatro componentes. La altura en el eje z es una posible distribución de densidad sobre el simplex, es decir, sobre las distribuciones de densidad multinomiales, dada por LDA.

Cuando observamos los datos, no contamos con los tópicos ni con su distribución, sino con los productos y países. El objetivo es realizar inferencia sobre las variables latentes, mediante el Teorema de Bayes:

$$p(\theta, z|w, \alpha, \beta) = \frac{p(\theta, z, w|\alpha, \beta)}{p(w|\alpha, \beta)}$$

Nuestra función objetivo es:

$$\ell(\alpha, \beta) = \sum_{d=1}^M \log p(w_d|\alpha, \beta)$$

El problema es que esta ecuación es intratable, por la interacción entre θ y β . Por ello, la inferencia se realiza sobre una familia de modelos que se sabe que son una cota inferior de probabilidad, y que son tratables. Estos modelos tienen parámetros variacionales, que se ajustan para obtener el modelo que más se acerca a la cota inferior. La forma de obtener una familia de modelos tratables es considerar algunas modificaciones sobre el modelo

gráfico original, removiendo nodos y aristas (Hoffman et al., 2013). En la figura 3.8 se puede observar la solución propuesta por (Blei et al., 2003) que, dado que utilizamos la implementación del modelo que se encuentra en Pedregosa et al. (2011) es también la del presente trabajo.

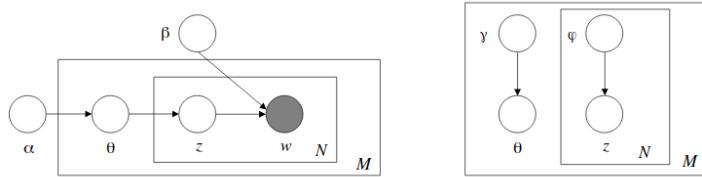


Fig. 3.8: fuente: Blei et al. (2003)

La estimación de los parámetros se realiza a través del proceso de *variational Expectation Maximization* (EM):

- **paso E:** Optimizamos los parámetros variacionales γ, φ
- **paso M:** Para los valores fijos γ, φ , maximizamos la cota inferior respecto a los parámetros del modelo, α, β

Estos dos pasos se alternan hasta converger. Por último, se realiza un suavizado sobre las probabilidades que un componente asigna a un producto, para que sean siempre mayores a 0.

3.3.2. Grafo Bipartito

[ver si llego](#)

3.4. Resultados

3.4.1. Latent Dirichlet Allocation Models

Los resultados obtenidos de los modelos de Topic Modelling para el análisis de textos normalmente se analizan en dos etapas: En primer lugar se etiquetan los tópicos obtenidos sobre la base de las palabras más salientes de cada tópico, y luego se analiza su distribución en los textos. El etiquetado de los tópicos es una tarea subjetiva donde lo que se busca es un concepto generalizador de aquellas palabras que componen al tópico, donde es posible que esta tarea no se pueda realizar por falta de coherencia dentro del tópico. El hiperparámetro k , es decir la cantidad de tópicos, juega un rol fundamental en este punto, dado que con una cantidad baja de tópicos estos tenderán a reflejar conceptos amplios, mientras que si k es mayor que la cardinalidad del espacio latente que se busca, esto puede generar tópicos repetidos. Lo anterior no escapa al presente dominio sino que se refuerza, dado que la búsqueda, subjetiva, de un concepto abarcador entre productos puede resultar más compleja que la de un concepto generalizador de un grupo de palabras. Un problema que no existe en el presente dominio es el de la polisemia, dado que todos los significantes, índices del nomenclador, hacen referencia a un único significado no ambiguo. Sin embargo, nuevos problemas aparecen respecto a este punto, como qué nomenclador de base utilizar

y en qué nivel de desagregación. Por motivos de comparabilidad con los resultados de Molinari & de Angelis (2016) se decidió utilizar el nomenclador SITC a 4 dígitos (ONU, 2006). A su vez, se realizaron pruebas para varios valores de k : 2,4,6,8,10,20,30,40, 100 y 200. Para el etiquetado de los componentes se observó que la práctica usual de observar los primeros diez elementos de la distribución no bastaba para encontrar una etiqueta generalizadora, por lo que se elaboró un tablero dinámico para estudiar la distribución y su función acumulada, a la vez que se gráfico la distribución en función de un nomenclador de complejidad tecnológica (Lall, 2000):

Código	Descripción
01	Primary products
02	Resource-based manufactures: agro-based
03	Resource-based manufactures: other
04	Low technology manufactures: textile, garment and footwear
05	Low technology manufactures: other products
06	Medium technology manufactures: automotive
07	Medium technology manufactures: process
08	Medium technology manufactures: engineering
09	High technology manufactures: electronic and electrical
10	High technology manufactures: other
99	Unclassified products

En la figura 3.9 se puede observar la tabla utilizada para caracterizar el cuarto componente en el modelo con $k = 4$. Allí se observa que la mayoría de los componentes pertenecen al rubro textil, como pulloveres, remeras y pantalones. También se aparece el té, que no guarda coherencia con le resto del rubro.

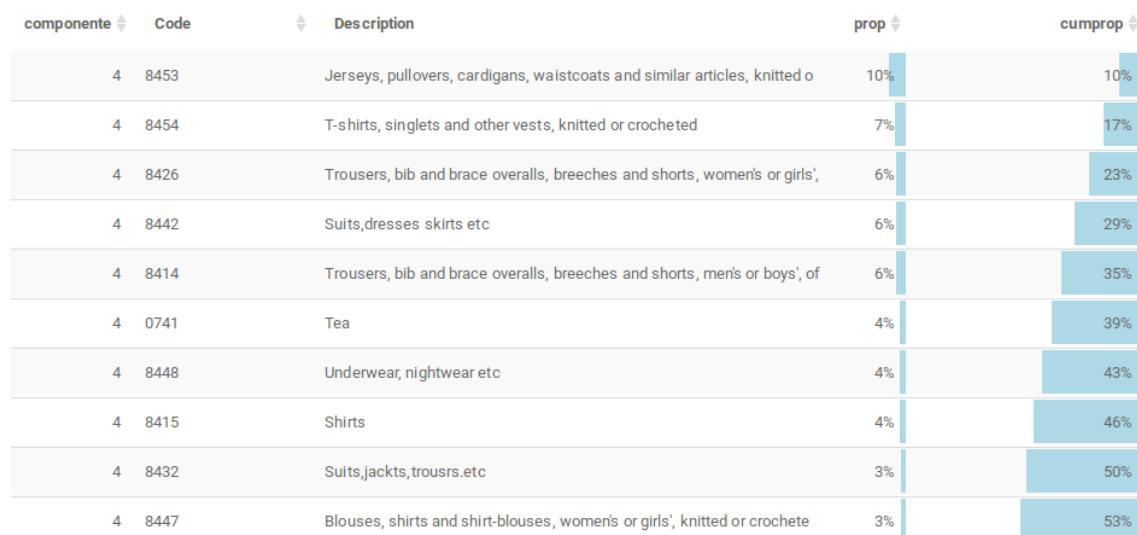


Fig. 3.9: Distribución del Componente 4. K = 40

La figura muestra la distribución siguiendo los grupos de Lall (2000). Aquí se puede confirmar que la distribución se concentra fuertemente en la categoría 4, textiles.

seguir aca:
Granularidad, casos
problemáticos. 40 es el
mejor. Mostrar etiquetas
finales.

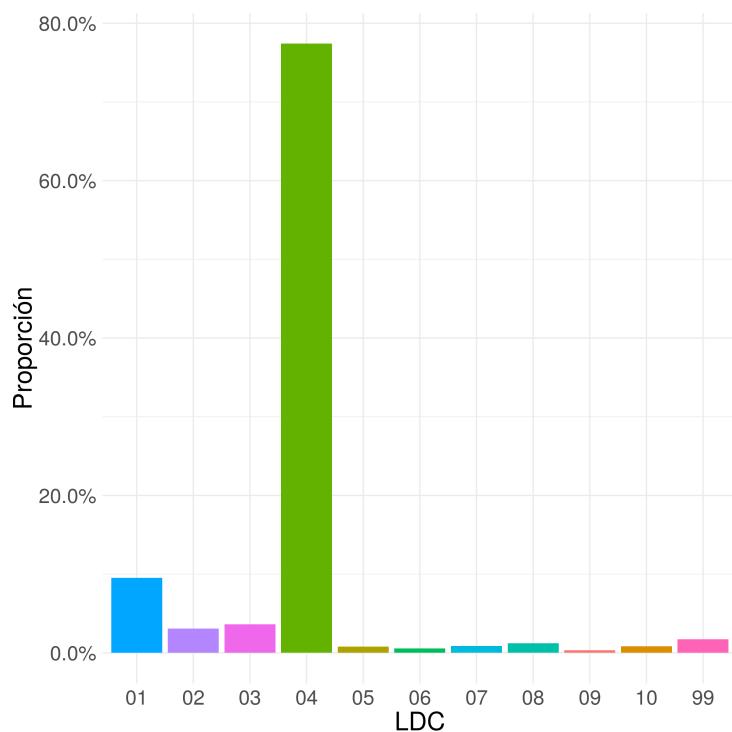


Fig. 3.10: Distribución del Componente 4. K = 40 según Lall

3.4.2. Grafo Bipartito

[ver si llego](#)

4. CONCLUSIONES

Apéndice

La nomenclatura de Cadenas y Subcadenas utilizada en el presente trabajo proviene de Molinari & de Angelis (2016). A su vez la clasificación en Usos proviene de Flóres (2008)

.1. Tabla de referencia: Descripción cadenas

Cadena	Descripción
Agroind	Agroindustria
Animal	Manufacturas De Origen Animal
Automotriz	Automotriz
Bebidas	Bebidas
Bk	Bienes De Capital
Combyene	Combustibles Y Energía
Cuero	Marroquinería, Calzado Y Otras Manufacturas De Cuero Y Pieles
Foreind	Foresto-Industria
Insdif	Insumos Difundidos
Medic	Medicamentos
Oeqtransp	Otro Equipo De Transporte
Texteind	Textil E Indumentaria
Omanuf	Otras Manufacturas
Quim	Otros Químicos

.2. Tabla de referencia: Descripción subcadenas

Cadena	Subcadena	Descripción
Agroindustria	Agroquim	Agroquímicos
Agroindustria	Alimanim	Alimentos Para Animales
Agroindustria	Ceryoleag	Cereales Y Oleaginosas
Agroindustria	Flh	Frutas, Legumbres Y Hortalizas
Agroindustria	Oagroind	Otros Agroindustria
Animal	Aviar	Aviar
Animal	Lácteos	Lácteos
Animal	Oanimal	Otros Cadena Animal
Animal	Pescado	Pescado
Animal	Porcina	Porcina
Animal	Vacuna	Vacuna
Automotriz	Autopartes	Autopartes
Automotriz	Autos	Autos
Automotriz	Camyonib	Camiones Y Ómnibus
Automotriz	Oautom	Otros Automotriz
Bebidas	Alcohol	Alcohólicas
Bebidas	Noalco	No Alcohólicas
Bebidas	Vino	Vino
Bienes De Capital	Eqelec	Equipos Eléctricos
Bienes De Capital	Eqmedico	Equpamiento Médico
Bienes De Capital	Maqagr	Maquinaria Agrícola

Cadena	Subcadena	Descripción
Bienes De Capital	Maqalim	Maquinaria Para Alimentos
Bienes De Capital	Maqcomb	Maquinaria Para Combustibles
Bienes De Capital	Maqfor	Maquinaria Forestal
Bienes De Capital	Maqmet	Maquinaria Para Metales
Bienes De Capital	Maqtextycuero	Maquinaria Para Textiles Y Cuero
Bienes De Capital	Obk	Otros Bienes De Capital
Bienes De Capital	Oeqtrans	Otro Equipo De Transporte
Bienes De Capital	Vehyautop	Vehículos Y Autopartes
Combustibles Y Energía	Biocomb	Biocombustibles
Combustibles Y Energía	Electricidad	Electricidad
Combustibles Y Energía	Gas	Gas
Combustibles Y Energía	Nuclear	Nuclear
Combustibles Y Energía	Ocombyene	Otros Combustibles Y Energía
Combustibles Y Energía	Petroleo	Petróleo
Combustibles Y Energía	Quimicos	Químicos
Marroquinería, Calzado Y Otras Manufacturas De Cuero Y Pieles	Calzado	Calzado
Marroquinería, Calzado Y Otras Manufacturas De Cuero Y Pieles	Cueros	Cueros
Marroquinería, Calzado Y Otras Manufacturas De Cuero Y Pieles	Ocuero	Otros Cueros
Marroquinería, Calzado Y Otras Manufacturas De Cuero Y Pieles	Quimcuero	Químicos Para Cueros
Foresto-Industria	Grafica	Gráfica
Foresto-Industria	Madymueb	Madera Y Muebles
Foresto-Industria	Oforeind	Otros Foresto-Industria
Foresto-Industria	Papel	Papel
Insumos Difundidos	Metales	Metales
Insumos Difundidos	Minerales	Minerales
Insumos Difundidos	Oinsdif	Otros Insumos Difundidos
Insumos Difundidos	Plasticos	Químicos
Medicamentos	Medic	Medicamentos
Medicamentos	Omedic	Otros Medicamentos
Medicamentos	Pactivos	Principios Activos
Otro Equipo De Transporte	Aeronautico	Aeronáutico
Otro Equipo De Transporte	Ferroviario	Ferroviario
Otro Equipo De Transporte	Motos	Motos
Otro Equipo De Transporte	Naval	Naval
Otro Equipo De Transporte	Oeqtrans	Otro Equipo De Transporte
Otro Equipo De Transporte	Opartes	Partes Para Otro Equipo De Transporte
Textil E Indumentaria	Fibras	Fibras
Textil E Indumentaria	Hilados	Hilados
Textil E Indumentaria	Indyacces	Indumentaria Y Accesorios
Textil E Indumentaria	Otexteind	Otros Textiles E Indumentaria
Textil E Indumentaria	Primtext	Primarios Textiles
Textil E Indumentaria	Quimtext	Químicos Para Textiles

Cadena	Subcadena	Descripción
Textil E Indumentaria	Tejidos	Tejidos
Otras Manufacturas	Animal	Otras Manufacturas De Origen Animal
Otras Manufacturas	Armas	Armamento
Otras Manufacturas	Cauyplast	Manufacturas De Caugo Y Plástico
Otras Manufacturas	Cineyfoto	Manufacturas Para Cine Y Fotografía (no Bk)
Otras Manufacturas	Corcho	Corcho Y Sus Manufacturas
Otras Manufacturas	Electrodom	Electrodomésticos
Otras Manufacturas	Eqeletc	Otro Equipo De Eléctrico (no Bk)
Otras Manufacturas	Herramientas	Herramientas
Otras Manufacturas	Instrumentos	Aparatos E Instrumentos Varios (no Bk)
Otras Manufacturas	Juguetes	Juguetes
Otras Manufacturas	Mateqconst	Materiales Y Equipos De Construcción
Otras Manufacturas	Metylprec	Metales Y Piedras Preciosas
Otras Manufacturas	Muebles	Otros Muebles
Otras Manufacturas	Oarte	Obras De Arte
Otras Manufacturas	Oartmet	Otros Artículos De Metal
Otras Manufacturas	Oficina	Artículos De Oficina
Otras Manufacturas	Omateriales	Otros Materiales
Otras Manufacturas	Oomanuf	Otras Manufacturas
Otras Manufacturas	Quimicos	Otros Químicos Y Sus Manufacturas
Otras Manufacturas	Tabaco	Tabaco Y Sus Manufacturas
Otras Manufacturas	Vegetal	Otras Manufacturas De Origen Vegetal
Otros Químicos	Caucho	Caucho
Otros Químicos	Cosmeticos	Cosméticos
Otros Químicos	Ooquim	Otros Químicos
Otros Químicos	Quimicos	Químicos

3. Tabla de referencia: Descripción Usos

Uso	Descripción
PP	Productos Primarios
P&C	Partes Y Componentes
ST	Bienes Semiterminados
BK	Bienes De Capital
BC	Bienes De Consumo

Bibliografía

- Charu C Aggarwal, Alexander Hinneburg, and Daniel A Keim. On the surprising behavior of distance metrics in high dimensional space. In *International conference on database theory*, pages 420–434. Springer, 2001.
- J J Allaire. *rsconnect: Deployment Interface for R Markdown Documents and Shiny Applications*, 2019. URL <https://cran.r-project.org/package=rsconnect>.
- Vincent Arel-Bundock. countrycode: Convert Country Names and Country Codes, 2017.
- Jeffrey B Arnold. ggthemes: Extra Themes, Scales and Geoms for 'ggplot2', 2017.
- K. Bhattacharya, G. Mukherjee, J. Saramäki, K. Kaski, and S. S. Manna. The international trade network: Weighted network analysis and modelling. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(2), 2008. ISSN 17425468. doi: 10.1088/1742-5468/2008/02/P02002.
- David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.
- Jorge Carrillo and Alfredo Hualde. Third generation maquiladoras? The Delphi-General Motors case. *Journal of Borderlands Studies*, 13(1):79–97, 1998.
- Winston Chang, Joe Cheng, J J Allaire, Yihui Xie, and Jonathan McPherson. shiny: Web Application Framework for R, 2018. URL <https://cran.r-project.org/package=shiny>.
- William Chow. An Anatomy of the World Trade Network. pages 1–20, 2013.
- Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research, 2006.
- Dilip K Das. *Regionalism in global trade*. Edward Elgar Publishing, 2004.
- Giorgio Fagiolo, Javier Reyes, and Stefano Schiavo. The Evolution of the World Trade Web Giorgio. 2007.
- Giorgio Fagiolo, Javier Reyes, and Stefano Schiavo. The evolution of the world trade web: A weighted-network analysis. *Journal of Evolutionary Economics*, 20(4):479–514, 2010. ISSN 09369937. doi: 10.1007/s00191-009-0160-x.
- Ying Fan, Suting Ren, Hongbo Cai, and Xuefeng Cui. The state's role and position in international trade: A complex network perspective. *Economic Modelling*, 39:71–81, 2014. ISSN 02649993. doi: 10.1016/j.econmod.2014.02.027. URL <http://dx.doi.org/10.1016/j.econmod.2014.02.027>.
- Renato G Jr. Flôres. The World Fragmentation of Production and Trade Concepts and Basic Issues. In *International Workshop Integração Produtiva-Lições da Ásia e Europa para o MERCOSUL*, CEPAL/Brasil, page 7, 2008.

- Folker Fröbel, Jürgen Heinrichs, and Otto Kreye. The new international division of labour. *Information (International Social Science Council)*, 17(1):123–142, 1978.
- Kristian S Gleditsch. Expanded Trade and GDP Data. *Journal of Conflict Resolution*, 46:712–24, 2002.
- Keith Head and Thierry Mayer. Gravity Equations: Workhorse, Toolkit, and Cookbook. *Handbook of International Economics*, 4:131–195, jan 2014. ISSN 1573-4404. doi: 10.1016/B978-0-444-54314-1.00003-3. URL <https://www.sciencedirect.com/science/article/pii/B9780444543141000033?via%3Dihub>.
- Matthew D Hoffman, David M Blei, Chong Wang, and John Paisley. Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1):1303–1347, 2013.
- John Humphrey and Hubert Schmitz. *Governance and upgrading: linking industrial cluster and global value chain research*, volume 120. Institute of Development Studies Brighton, 2000.
- Sanjaya Lall. The Technological structure and performance of developing country manufactured exports, 1985–98. *Oxford development studies*, 28(3):337–369, 2000.
- Andrea Molinari and Jesica Yamila de Angelis. Especialización y complementación productiva en el MERCOSUR: un Enfoque de Cadenas Productivas de Valor. 2016.
- Erich Neuwirth. RColorBrewer: ColorBrewer Palettes, 2014.
- ONU. Standard International Trade Classification Revision 4. *Statistical Papers Series M No*, 34, 2006.
- F Pedregosa, G Varoquaux, A Gramfort, V Michel, B Thirion, O Grisel, M Blondel, P Prettenhofer, R Weiss, V Dubourg, J Vanderplas, A Passos, D Cournapeau, M Brucher, M Perrot, and E Duchesnay. Scikit-learn: Machine Learning in Python . *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- R Core Team. R: A Language and Environment for Statistical Computing, 2017.
- David Ricardo. *Principios de economía política y tributación*. Claridad, Buenos Aires, 2007.
- Kamil Slowikowski. ggrepel: Repulsive Text and Label Geoms for 'ggplot2', 2017.
- G van Rossum. Python tutorial. Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, may 1995.
- Vincent Q Vu. ggbiplot: A ggplot2 based biplot, 2011.
- Carsten Weerth. Basic Principles of Customs Classifications under the Harmonized System. *Global Trade & Cust. J.*, 3:61, 2008.
- Hadley Wickham. tidyverse: Easily Install and Load the 'Tidyverse', 2017.
- Claus O Wilke. ggridges: Ridgeline Plots in 'ggplot2', 2017.
- WTO. World Trade Statistics Review 2017. Technical report, 2017.