

Análisis empírico del comercio internacional a partir de la segunda mitad del siglo XX.

Propuestas metodológicas basadas en teoría de grafos y modelos generativos bayesianos

Diego Kozlowski

Universidad de Buenos Aires

*Master en Data Mining & Knowledge Discovery
Supervisor: Viktoriya Semeshenko*

Tema

Introducción

- Los datos de comercio internacional son flujos de importaciones y exportaciones entre países, para un período de tiempo
- Es un problema de alta dimensionalidad: país-año-producto-tipo de flujo
- el objetivo de esta tesis es construir un sistema de representación de la información que comprenda las múltiples dimensiones.
- Esto es realizado utilizando técnicas de Análisis exploratorio de datos, Redes complejas y modelos generativos bayesianos.

Fuentes de información

- Comercio agregado por país y año. 1948-2016¹
- Comercio desagregado a nivel producto, año y país.
1997-2016²
- Comercio desagregado a nivel producto, año y país.
1962-2016³

¹Gleditsch <http://ksgleditsch.com/extradegdp.html>

²Organización mundial del comercio <https://comtrade.un.org/>

³The Atlas of Economic Complexity

<http://www.atlas.cid.harvard.edu.>

Comercio agregado entre países

Metodología



- Grafo dirigido no ponderado de comercio bilateral por año y tipo
 - Análisis de puntos de corte y medidas de centralidad
 - Centralidad de los países

Definición del grafo

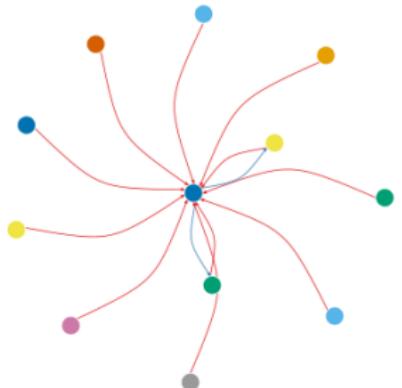
$$a_{ij} = \begin{cases} 1 & \text{si } \frac{x_{ij}}{x_{i\cdot}} \geq u \\ 0 & \text{sino} \end{cases}$$

x_{ij} : Comercio total entre los países i-j

$x_{i\cdot}$: Comercio total del país i

- ➊ Dos países están conectados si existe suficiente comercio entre ellos.
- ➋ Puede ser tanto exportación como importación.
- ➌ El grafo es dirigido, por lo que $a_{ij} \neq a_{ji}$.

Productores & Consumidores



- Desde el punto de vista de un país, no puede haber muchos países relevantes, ∴ no hay muchas aristas de salida
 - Pero puede ser relevante desde el punto de vista de muchos otros países. ∴ puede recibir muchas aristas de entrada
 - Con datos de expo, un nodo central es importante como consumidor
 - Con datos de impo, un nodo central es importante como productor

Resultados

Determinación del Punto de corte. En 1% Se maximiza el clustering y minimiza la densidad (dos objetivos deseables para el análisis)

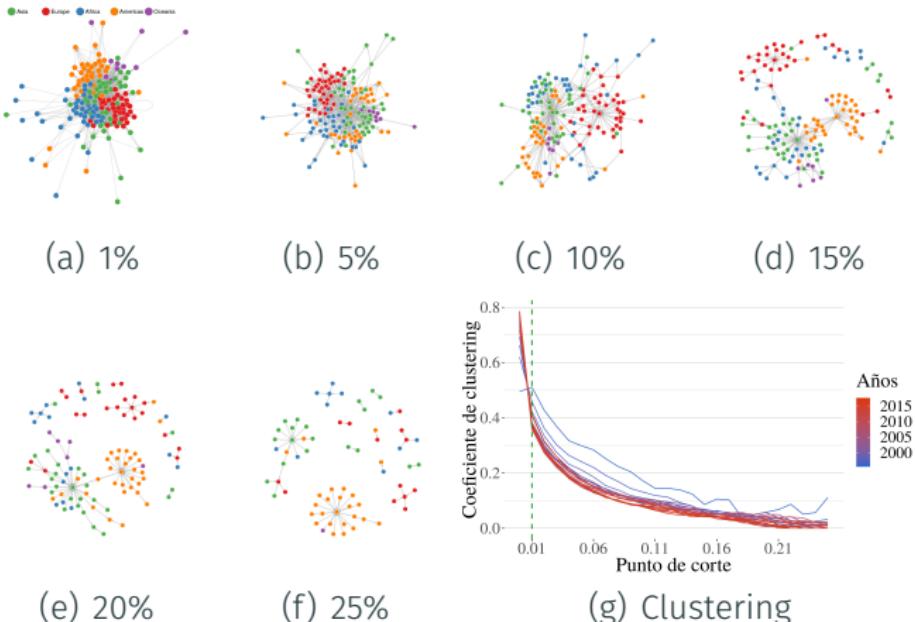
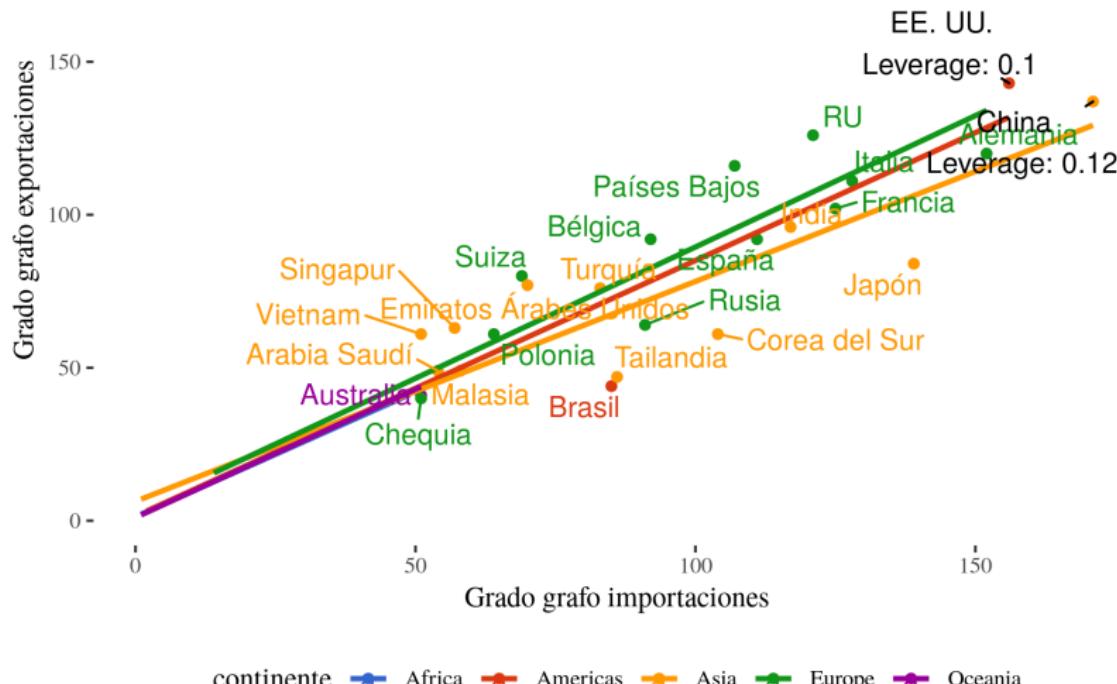


Figure 1: Grafos. 2016, según punto de corte

Correlación Expo-Impo.

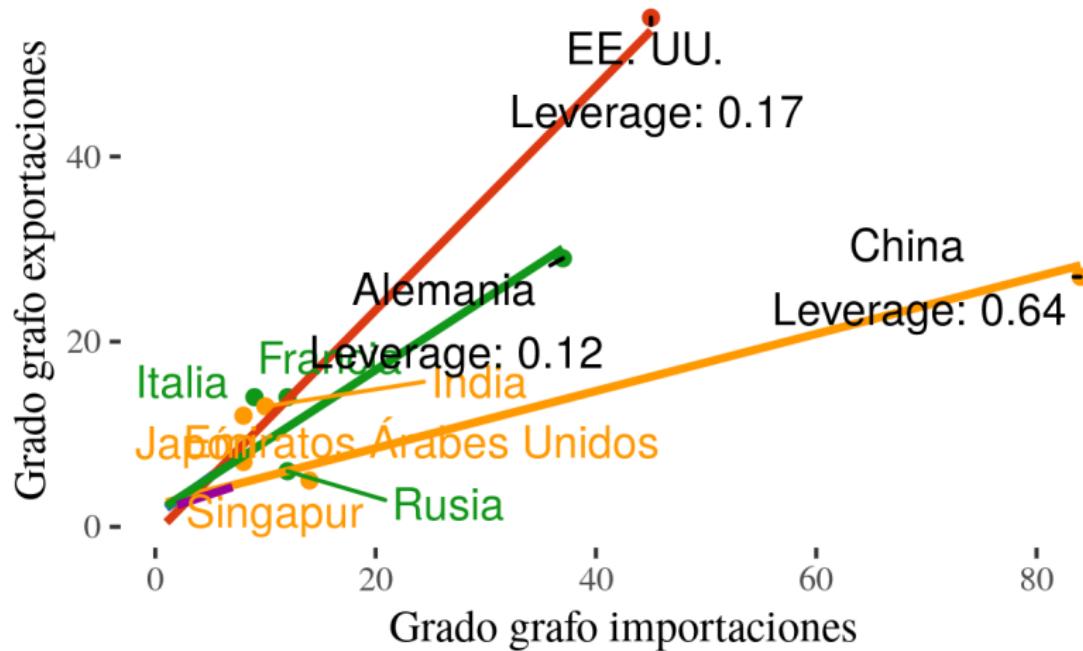
punto de corte: 1%. Coeficiente de Pearson= 0.96



- Por la forma de construcción, una mayor centralidad en el grafo de las importaciones denota una importancia del país como productor para el mercado mundial, y viceversa.
- Podemos ver que hay una correlación alta entre ambos grafos, pero también que países como Bélgica o Suiza son más importantes como consumidores de mercancías.
- Corea del Sur, Japón y Brasil, por su parte, son más importantes como productores de mercancías.

Correlación Expo-Impo.

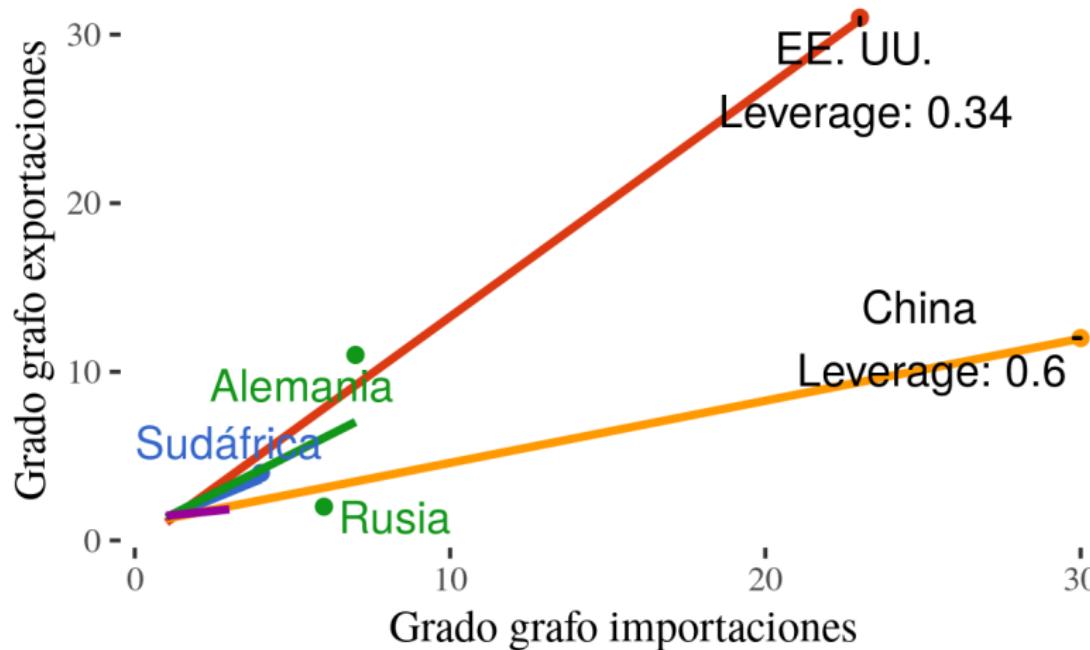
punto de corte: 10%. Coeficiente de Pearson= 0.77



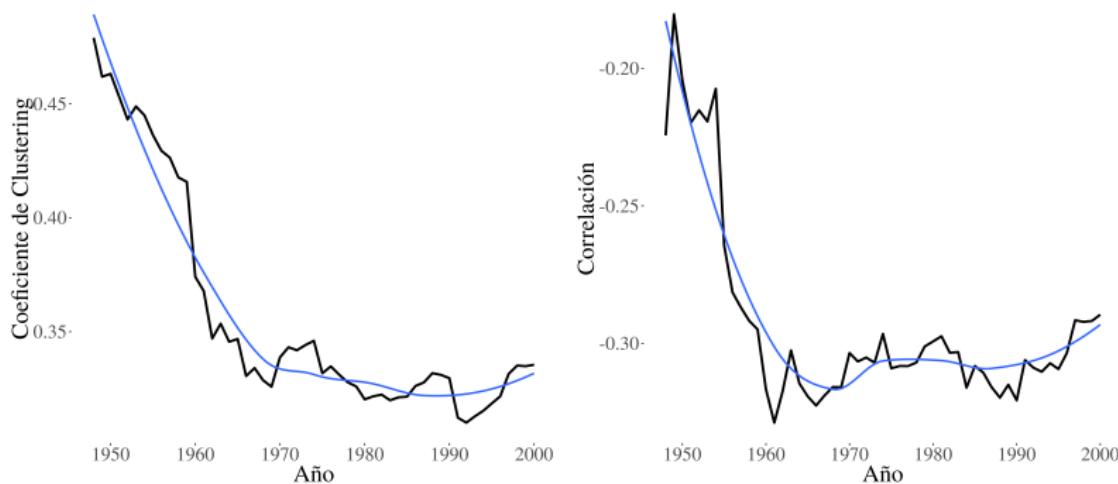
- Un punto de corte más alto implica relaciones de mayor dependencia.
- En este tipo de relaciones destacan menos países, con mayor leverage

Correlación Expo-Impo.

punto de corte: 20%. Coeficiente de Pearson= 0.81

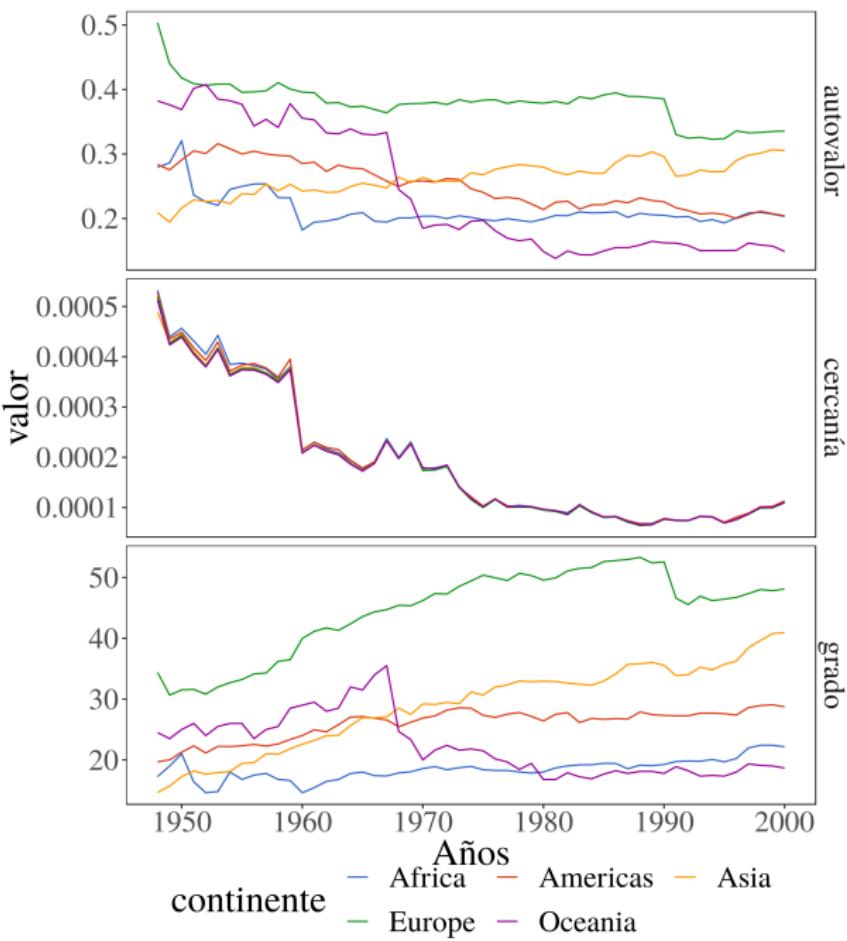


- En el extremo, en relaciones de alta dependencia destacan sobretodo Estados Unidos como consumidor, y China como productor.
- Alemania, Sudáfrica y Rusia también destacan por tener relaciones de alta dependencia.



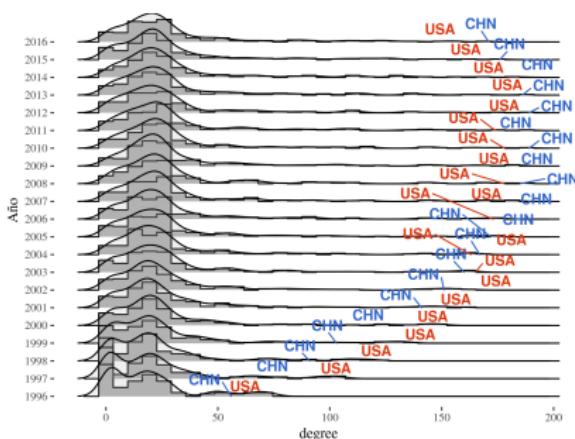
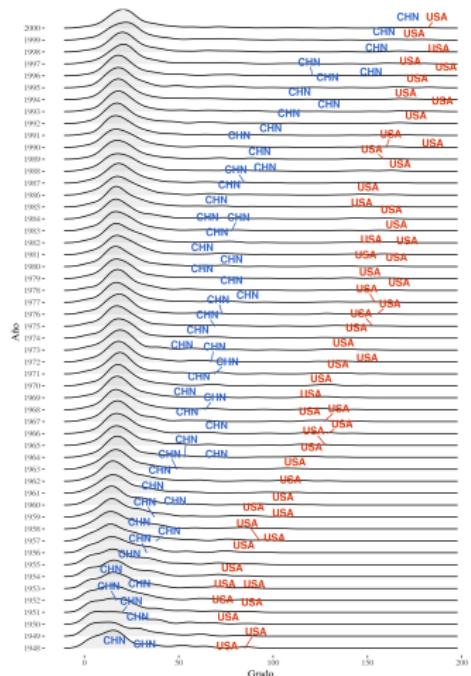
Evolución de la estructura de la red. Importaciones. Umbral 1%. 1948-2000

- En la segunda mitad del S XX se observa una tendencia a la clusterización o regionalización de las relaciones comerciales.
- La caída de la correlación de grado indica un aumento de las relaciones entre países centrales y periféricos

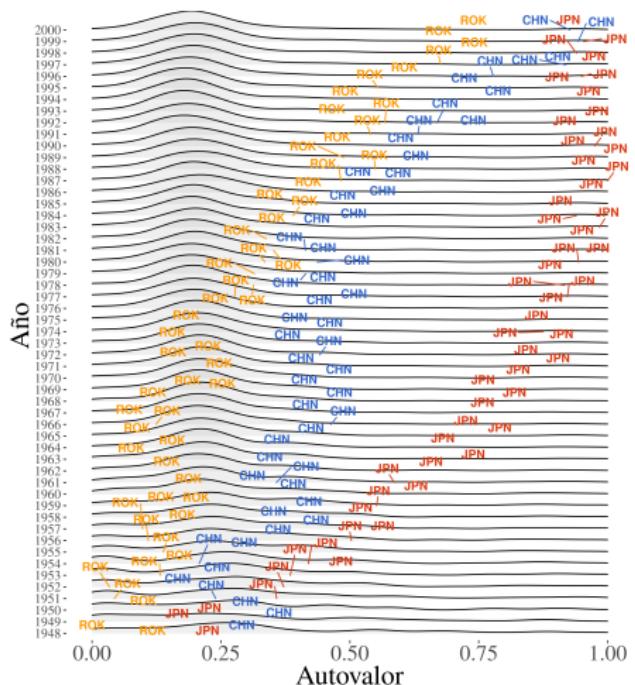


- Tanto en la centralidad de autovalor como en la de grado, Asia pasa del último lugar al segundo.
- En la centralidad de cercanía podemos observar que cae a lo largo de la serie de igual manera en todos los continentes.

Evolución de la distribución de la centralidad de grado en el tiempo. China y EEUU destacado



Evolución de la distribución de la centralidad de autovalor en el tiempo. Japón, Corea del Sur y China destacado



Estos tres países del sudeste asiático tienen una creciente importancia en el comercio mundial a partir de la segunda mitad del S XX, aunque en diferentes momentos.

Más resultados en:

[https://diegokoz.shinyapps.io/
Distribucion_nodos_wrdtrade/](https://diegokoz.shinyapps.io/Distribucion_nodos_wrdtrade/)

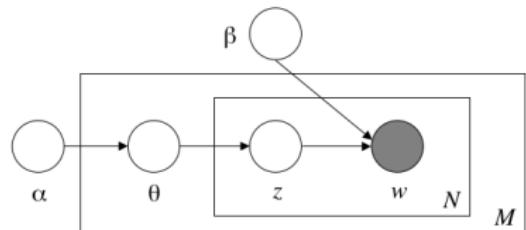
Comercio a nivel productos

- ⦿ Para caracterizar la posición de un país en el mercado mundial, no basta con conocer su importancia relativa.
 - ⦿ También es necesario entender su rol en términos de la composición de su canasta exportadora-importadora.
-

- ⌚ Se proponen dos formas distintas de organización de la información:
- ⌚ La primera busca las dimensiones latentes del comercio internacional, basado en [Blei et al., 2003]
- ⌚ La segunda se basa en Hidalgo et al. [2007], Hidalgo and Hausmann [2009] y proyecta un *espacio de productos* a partir de un grafo bipartito, para luego buscar clusters de productos.

Topic Modeling

Metodología



$\beta \sim Dir_v(\eta)$ Distribución de productos en los componentes latentes

$\theta_d \sim Dir_K(\alpha)$: Distribución de los componentes latentes por país

$z_{dn} \sim Mult(\theta_d)$: Componente latente

$w_{dn} \sim Mult(\beta_{zn})$: u\$s comerciado

Basado en Topic Modeling [Blei et al., 2003] realizamos inferencia sobre los *componentes latentes* en el comercio internacional utilizando un modelo de inferencia bayesiana

Cómo resultado, obtenemos la distribución de los productos sobre los componentes latentes, y la distribución de los componentes sobre los países

- **Input:** Matriz de dólares exportados por producto para cada país-año ($P^*Y \times N$)
- **Output:**
 - > La distribución de los productos por componente
 - > La distribución de los componentes por año-país
- Se construyó una página interactiva para explorar los resultados y asignar etiquetas a los componentes:
https://treemaps.shinyapps.io/shiny_LDA/
- Luego del análisis de componentes, se estudió la evolución de los mismos en la canasta exportadora de los países.

Ejemplo con dos componentes

	Component 1	Component 2		Search:
	Show 10 entries			
1	1	3330	Crude petroleum and oils obtained from bituminous materials	3%
2	1	9310	Special transactions, commodity not classified according to class	6%
3	1	3345	Lubricating petroleum oils, and preparations, nes	5%
4	1	3414	Petroleum gases, nes, in gaseous state	4%
5	1	3413	Petroleum gases and other gaseous hydrocarbons, nes, liquefied	4%
6	1	3222	Other coal, not agglomerated	2%
7	1	2815	Iron ore and concentrates, not agglomerated	2%
8	1	9710	Gold, non-monetary (excluding gold ores and concentrates)	2%
9	1	3344	Fuel oils, nes	1%
10	1	6821	Copper and copper alloys, refined or not, unwrought	1%

Showing 1 to 10 of 786 entries

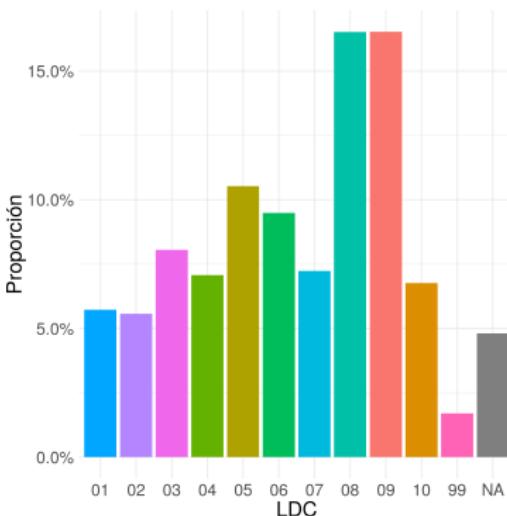
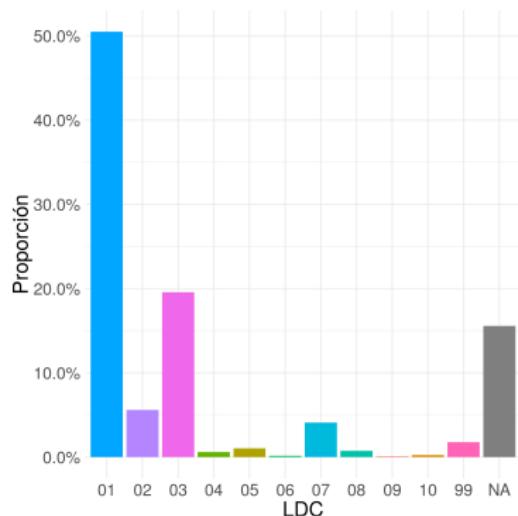
Previous 1 2 3 4 5 ... 79 Next

	Component 1	Component 2		Search:
	Show 10 entries			
1	2	7810	Passenger motor vehicles (excluding buses)	5%
2	2	7764	Electronic microcircuits	3%
3	2	9310	Special transactions, commodity not classified according to class	3%
4	2	7849	Other parts and accessories, for vehicles of headings 722, 781-783	2%
5	2	5417	Medicaments (including veterinary medicaments)	2%
6	2	3345	Lubricating petroleum oils, and preparations, nes	2%
7	2	7599	Parts, nes of and accessories for machines of headings 7512 and 752	2%
8	2	7649	Parts, nes of and accessories for apparatus falling in heading 76	1%
9	2	7643	Television, radio-broadcasting, transmitters, etc	1%
10	2	7721	Switches, relays, fuses, etc; switchboards and control panels, nes	1%

Showing 1 to 10 of 786 entries

Previous 1 2 3 4 5 ... 79 Next

Para ayudar en la caracterización también se analizó la distribución de cada componente según la clasificación de usos de Lall [2000].



De este análisis se concluyó que el mejor número de componentes era 30.

Topic Modeling

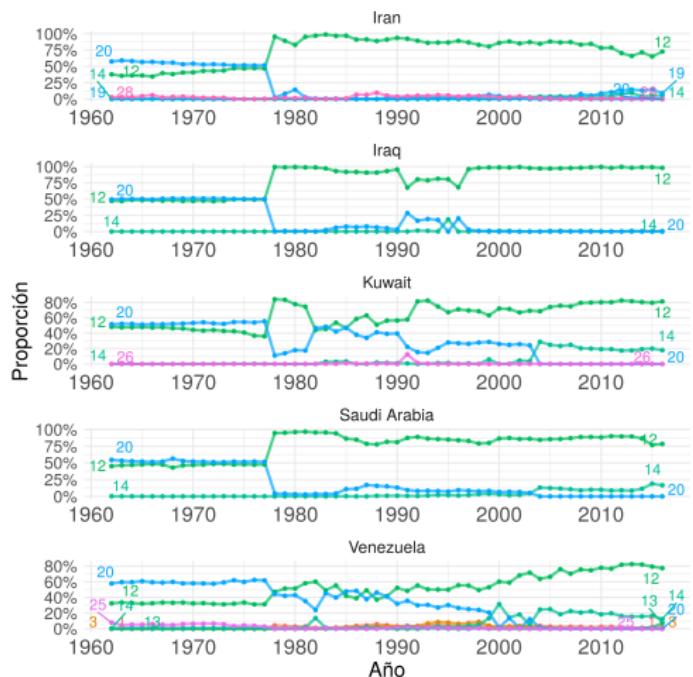
Resultados

Caracterización de los componentes según Grupo, subgrupo, Complejidad (industrial) y país característico.

Comp	Grupo	Subgrupo	Complejidad	País
1	Minerales	Carbón, hierro	-	Australia
2	Industria	Textiles, ingenieril, otros	Baja y media	San Marino
3	Industria	Vehículos y partes	Media	Bélgica
4	Industria	Textiles y juguetes, etc	Baja	Macao
5	Industria	Electrónicos no digitales. Cintas de grabación. Lineas de telefono. Papél fotográfico	Alta (hasta 70')	Checoslovaquia
6	Industria	Vehículos, barcos, maquinaria y partes	Media y alta	Japón
7	Petróleo	Estado Gaseoso	-	Turkmenistán
8	Minerales	Cobre	-	Chile
9	Agropecuario	Café, bananas, otros	-	Réunion
10	Industria	Autos y electrónicos	Media y alta	México
11	Industria	Autos, partes y componentes	Media	Alemania
12	Petróleo	Crudo	-	Sudán del Sur
13	-	Oro, relojes, joyas	-	Suiza
14	Industria	Químicos	-	Curazao
15	Minerales	Diamantes	-	Botswana
16	Industria + Agro.	Aviones, autopartes, soja y maíz	Media y alta	USA

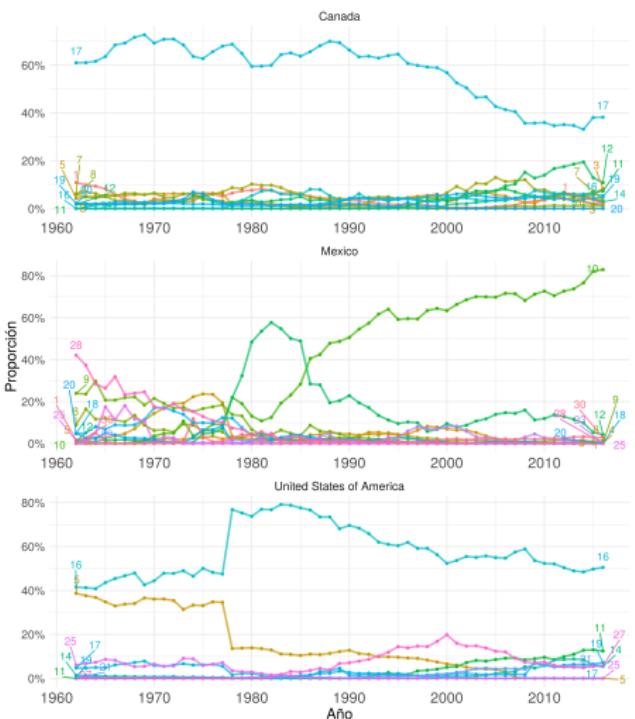
Introducción	Comercio agregado entre países oooooooooooooooooooo	Comercio a nivel productos ooooooo●ooooo○oooooooooooo	Conclusiones oooooo	References
Comp	Grupo	Subgrupo	Complejidad	País
17	Industria + Agro.	Autopartes, madera y derivados	Media	Finlandia
18	Industria + Agro.	Productos primarios y textiles	Baja	Isla de Navidad
19	-	Transacciones especiales, no clasificadas	-	Isla San Martín
20	Combustibles	Fuel oil, gasolina, etc.	-	Rep. Dem. Pop.del Yemen
21	Industria	Medicamentos e instrumentos médicos.	Alta	Irlanda
22	Petróleo + Agro	Hidrocarburos, aceite de palma,cacao, etc.	-	Ghana
23	Industria	Procesadores, microcircuitos, juguetes y calzado.	Alta y baja	China
24	Industra + Agro.	Barcos, carnes, pescados, lácteos	Media	Islandia
25	Minerales + Agro.	Soja y derivados, Hierro	-	Paraguay
26	Industria + Agro.	Aviones, perfumería, vino	Alta	Francia
27	Industria	Microcircuitos	Alta	filipinas
28	Industria + Agro.	Arroz, algodón, textiles, goma, etc.	Baja	Pakistán
29	Industria + Agro.	maquinaria, flores, quesos.	Alta	Países Bajos
30	Industria	Vehículos, partes y medicamentos	Media y alta	Reino Unido

Distribución de componentes en países I. Países petroleros.



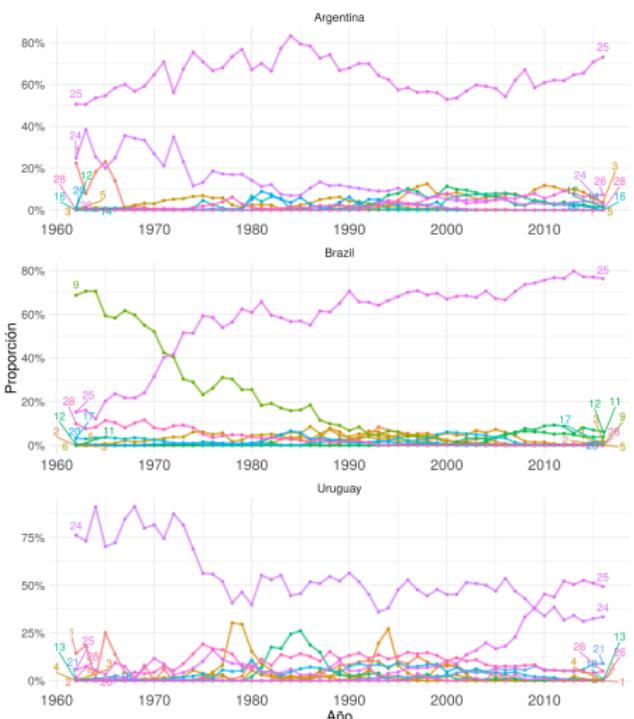
A partir de la crisis del petróleo del 70'
reprimarizan su producción
y pasan de derivados del petróleo, *Fuel Oil*, a Petróleo crudo.

Distribución de componentes en países II. Canadá, México y Estados Unidos



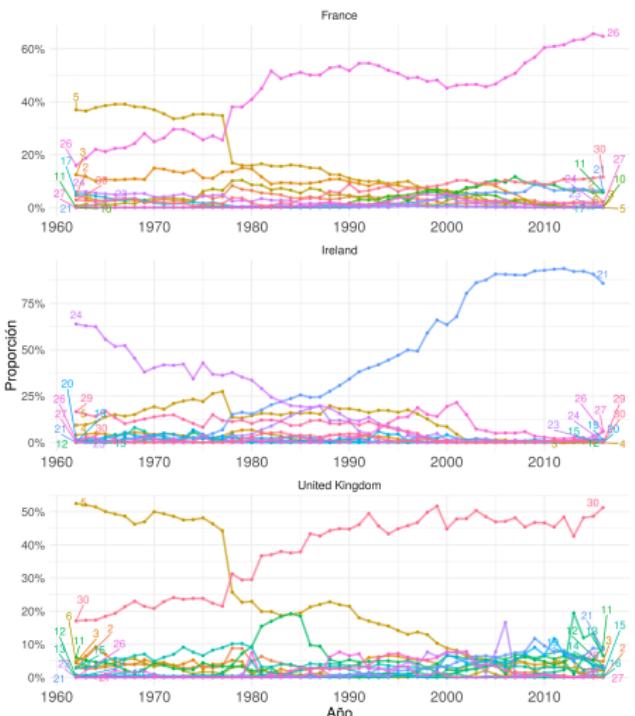
- En Canadá domina el componente 17: Industria más producción maderera y derivados. Sin embargo, desde mediados de los 90' crecen en importancia el petróleo y derivados (comp. 12 y 7)
- En México se observa un boom petrolero entre mediados de los 70' hasta mediados de los 80'. A partir de allí se observa el fenómeno conocido como la Maquila mexicana.
- Estados Unidos exporta en el componente 5, de alta tecnología hasta los 70'. Luego toma su lugar el componente 16: aviones y autopartes, pero también maíz y soja

Distribución de componentes en países III. Argentina, Brasil y Uruguay



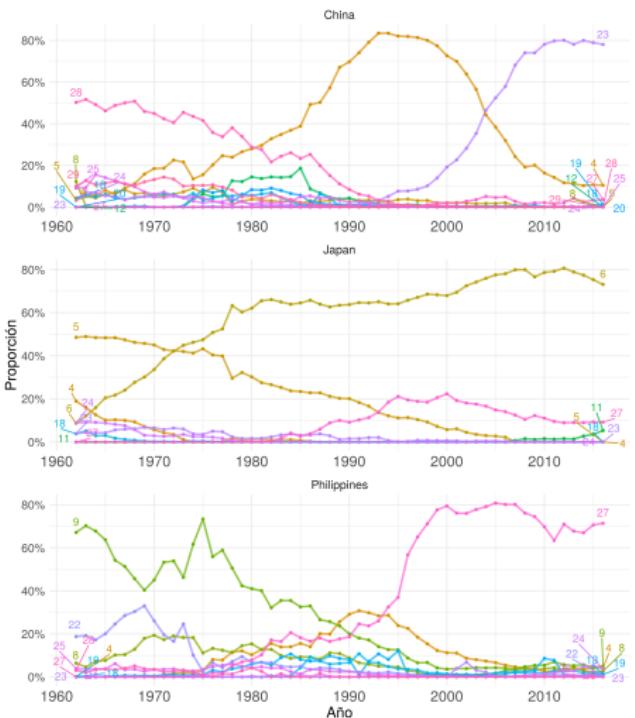
- ⌚ En los tres países el componente más importante al finalizar la serie es el 25 (soja).
- ⌚ En Argentina al principio de la serie tiene importancia el comp. 24 (Ganadería). En Uruguay ese componente es el más importante hasta casi el final de la serie.
- ⌚ En Brasil La producción sojera reemplaza al componente 9(café, bananas,etc.), dominante en los 60' y 70'.

Distribución de componentes en países IV. Francia, Irlanda y Reino Unido



- ➊ Francia y Reino Unido también exportaban sobre el comp. 5 hasta fines de los 70' y luego pasan a un componente característico.
- ➋ Francia exporta en el comp. 26 (Aviones, perfumería y vino) y UK en el 30(Vehículos, partes y medicamentos).
- ➌ A mediados de los 80' Irlanda pasa del comp.24 (ganadería) al 21, de industria de alta complejidad.

Distribución de componentes en países IV. China, Japón y Filipinas



En China se pueden observar tres etapas:
Hasta los 80' el componente 28
(Arroz, algodón, textiles) era predominante.
Entre 1980-2004 domina el componente 4
(textiles y juguetes), de
industria de baja complejidad. A partir de
allí, el componente 23
(procesdaores, microcircuitos, etc.).

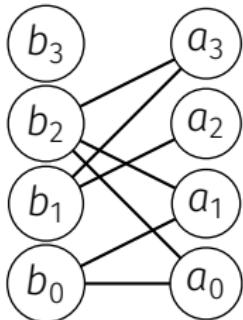
Japón, al igual que EUA, Francia y UK
comienza la serie en el comp. 5, pero la caída
en este caso es tendencial, y se reemplaza
por el 6 (vehículos, barcos, maquinarias y
partes).

En el caso de Filipinas, hasta los 90' se
especializa en productos agropecuarios (Café,
bananas, etc.) pasa brevemente por el
componente 4, y rápidamente toma su lugar
el 27(microcircuitos).

En muchos países del sudeste asiático se ven, con matices, particularidades y
distintas temporalidades, las tres etapas vistas en China y Filipinas.

Grafo Bipartito

Metodología



- ➊ Grafo bipartito entre países y productos
- ➋ Se puede proyectar a un grafo de productos, donde la relación de cercanía se define a partir de qué países los exportan
- ➌ También se puede proyectar a un grafo de países, donde la relación de cercanía surge por la similitud de sus canastas exportadoras.

$$RCA(c, i) = \frac{\frac{x(c, i)}{\sum_i x(c, i)}}{\frac{\sum_c x(c, i)}{\sum_{c,i} x(c, i)}}$$

- ★ Para definir el punto de corte, se utilizó el concepto de *Relative Comparative Advantages* basado en Hidalgo and Hausmann [2009].
- ★ Esto es, la proporción que representa un producto en la canasta exportadora de un país, respecto la proporción que representa en la canasta exportadora promedio mundial.

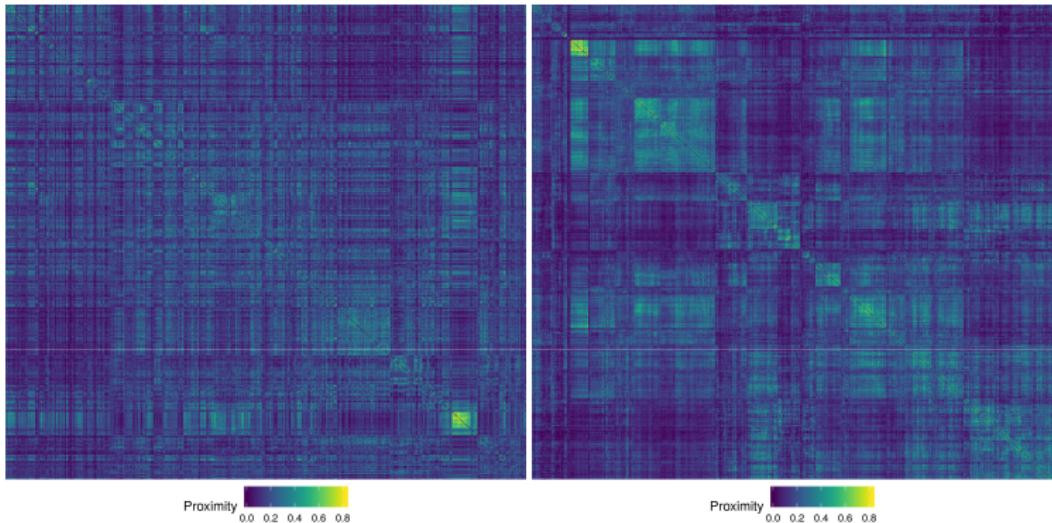
$$\phi_{ij} = \min(P(RCA_i > 1 / RCA_j > 1), P(RCA_j > 1 / RCA_i > 1)),$$

- ★ Para el análisis de los productos, nos basamos en el concepto de *proximidad* de Hidalgo and Hausmann [2009].
- ★ $P(RCA_i / RCA_j)$ es la probabilidad condicional de exportar el producto i dado que exporta el producto j .
- ★ Φ es por lo tanto una matriz de distancias entre los productos. Esto resulta ideal para analizar los datos a través de la técnica de clustering K-medoids [Kaufman and Rousseeuw, 1987]. El producto característico sirve para interpretar el cluster.

Para el análisis del grafo proyectado de países, se utilizó el *clustering de Louvain* [Blondel et al., 2008] para detectar las comunidades de países que comparten una canasta exportadora similar.

Grafo Bipartito

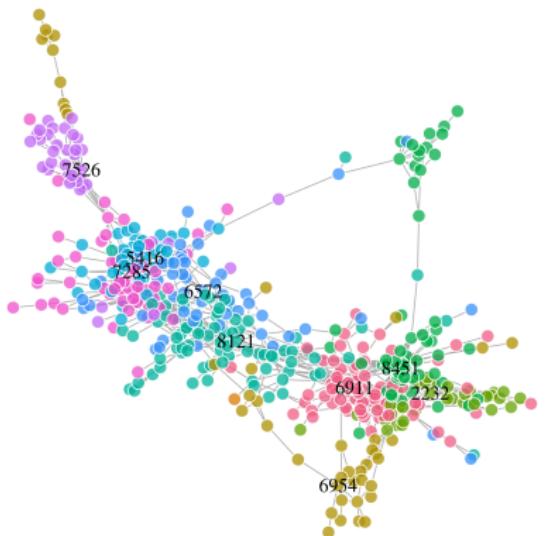
Resultados



(a) Ordenado por nomenclador
SITC (b) Ordenado por clustering

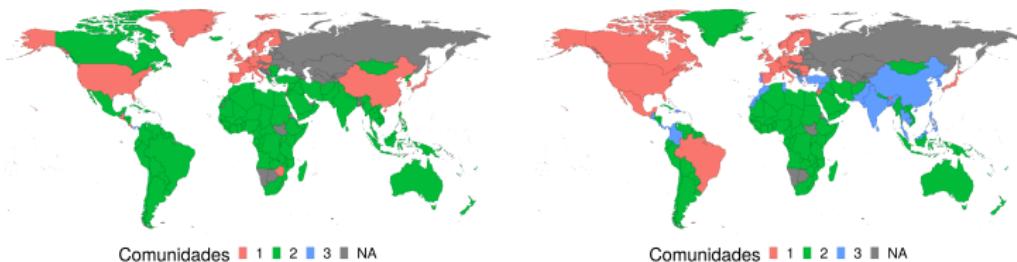
Figure 3: Heatmap proximidad

Componente gigante. Grafo de proximidad y clustering por K-medoides. Exportaciones. Promedio 1996-2017. 10 medioides



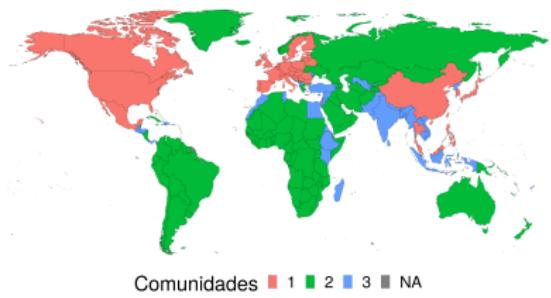
mediodide	Description
6911	Metal structures,parts
2831	Copper ores and concentrates
6954	Hand tools,etc. nes
2232	Palm nuts and kernels
8451	Babies'garments,clths acc
8121	Boilrs.radiatrs,etc.n.el
5416	Glycosides; glands etc.
6572	Non-wovens, whether or not impregnated
7526	... units for automatic data processing
7285	Parts public wrk mach etc

- El ordenamiento por proximidad permite una mejor organización que por el nomenclador tradicional (SITC[WTO, 2017])
- En la proyección del espacio de productos se observan algunos clusters bien diferenciados, como el de microcircuitos(7526) y el de herramientas(6954). Sin embargo, los demás no son claramente separables, y su interpretación, al estar atada a un solo producto, es más compleja.



(a) 1966

(b) 1986



(c) 2016

Figure 4: Proyección a países del grafo bipartito. Clustering Louvain.
Exportaciones

- En la proyección al grafo de países se pueden ver, a grandes rasgos, los cambios en la división internacional del trabajo:
 - > En 1966 el mundo tiene dos grandes grupos: industriales y agropecuarios
 - > En 1986 se constituye un tercer cluster de países del Sudeste Asiático (China, Pakistán, India, Filipinas, Tailandia, Malasia, etc.)
 - > en 2016, en el Sudeste Asiático se suman países al tercer cluster, mientras que otros (China, Tailandia, Malasia, Filipinas) pasan al cluster de Europa, EEUU y Japón.

Conclusiones

Comercio Agregado

- El grafo simple de comercio agregado entre países nos permite caracterizar las relaciones asimétricas en el comercio internacional
- Según el punto de vista de exportaciones e importaciones, se puede ver el rol de los países en tanto productores y consumidores de mercancías en el mercado mundial.

Comercio Agregado

- ★ Por el método de construcción, al integrar el punto de corte como variable, podemos observar relaciones de dependencia entre países.
- ★ Las medidas de clustering y correlación del grafo nos permiten caracterizar la evolución de las formas generales del comercio mundial.
- ★ La evolución en el tiempo de las medidas de centralidad nos permite observar los cambios en los grandes jugadores del comercio internacional.

LDA

- Al utilizar la información desagregada a nivel producto encontramos que el análisis se enriquece, pero también que la multiplicidad de dimensiones de estudio complejizan la lectura de la información.
- La metodología propuesta de Topic Modeling implica la construcción de un nomenclador *agnóstico*, exclusivamente basado en los datos.

LDA

- LDA Logra dar cuenta de la especificidad de la canasta exportadora de los países, y su evolución, con un grado de detalle que permite corroborar una multiplicidad de fenómenos reconocidos en la literatura
- Lograr resumir en un único sistema de indicadores una caracterización del comercio mundial con cierto nivel de detalle permite evitar sesgos de selección de indicadores.

Grafo bipartito

- Como contraste a LDA, se elaboró una variación de la línea de trabajo de Hidalgo et al. [2007] basada en grafos bipartitos, y las nociones de *RCA* y *proximity*.
- Con las proyecciones al espacio de productos y países se obtuvieron resultados análogos a los de LDA, aunque con un menor nivel de detalle.

Referencias

References

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

C. A. Hidalgo, B. Winger, A. L. Barabási, and R. Hausmann. The product space conditions the development of nations. *Science*, 317(5837):482–487, 2007. ISSN 00368075. doi: [10.1126/science.1144581](https://doi.org/10.1126/science.1144581).

César Hidalgo and Ricardo Hausmann. The building blocks of economic complexity. *Proceedings of the National Academy of the Sciences of the United States of America*, 106(26):10570–10575, 2009. ISSN 0027-8424. doi: [10.1073/pnas.0900943106](https://doi.org/10.1073/pnas.0900943106).

Leonard Kaufman and Peter J Rousseeuw. Clustering by means of medoids. Statistical Data Analysis based on the L1 Norm. *Y. Dodge, Ed*, pages 405–416, 1987.

Sanjaya Lall. The Technological structure and performance of developing country manufactured exports, 1985-98. *Oxford development studies*, 28(3):337–369, 2000.

WTO. World Trade Statistics Review 2017. Technical report, 2017.