



Centro de Investigación  
y Desarrollo Ecuador

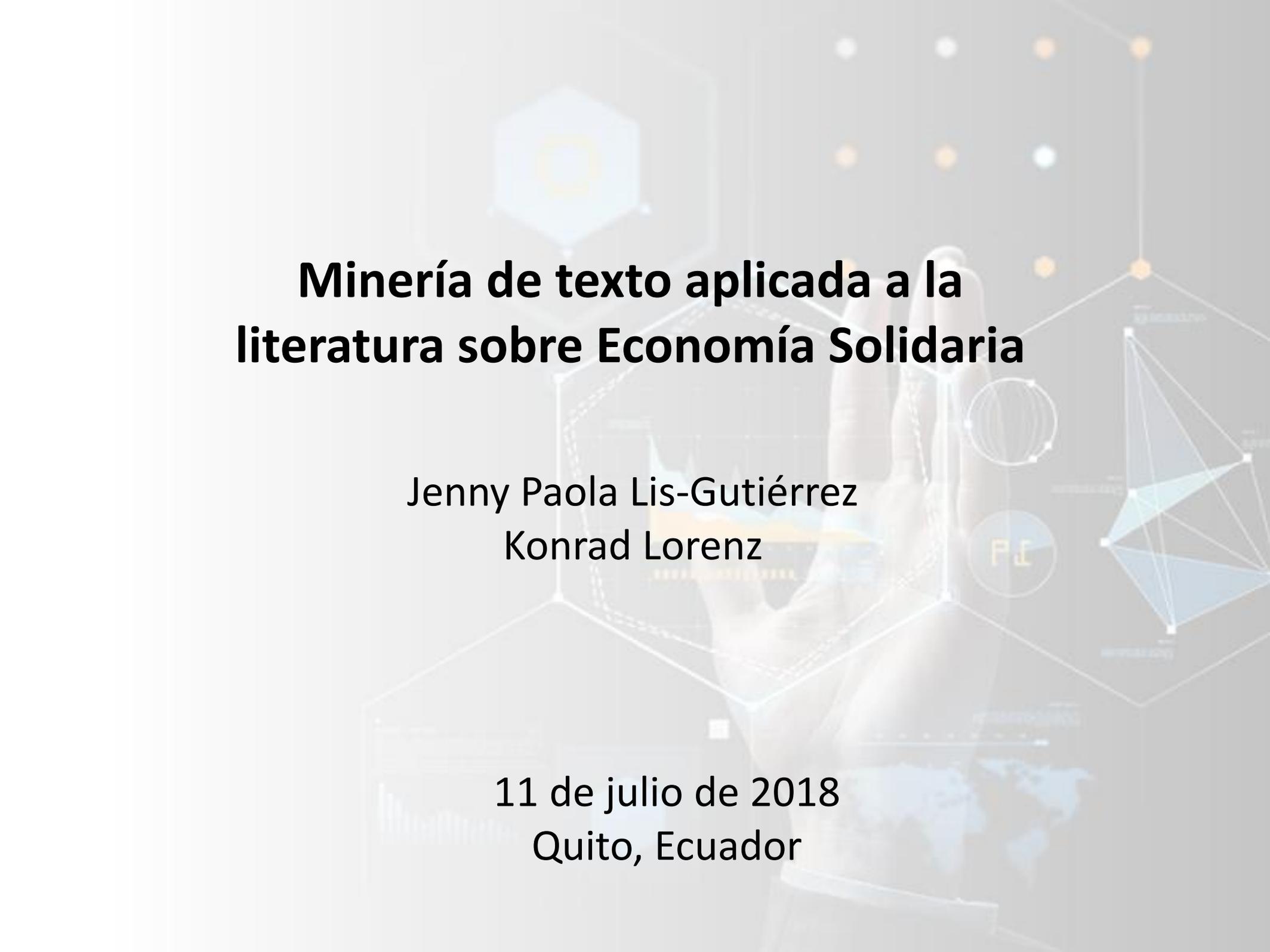
SEMINARIO  
INTERNACIONAL



# ECONOMÍA DIGITAL

POPULAR Y SOLIDARIA

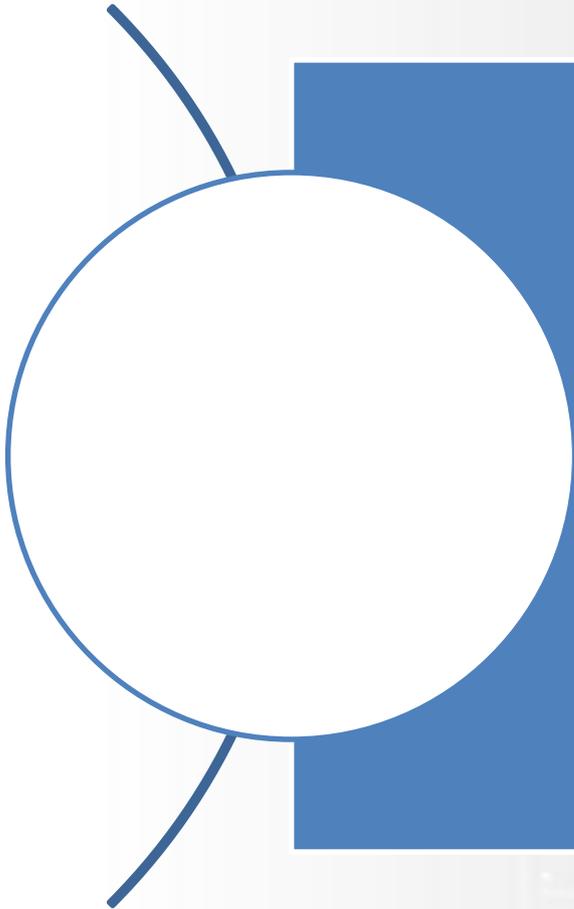




# **Minería de texto aplicada a la literatura sobre Economía Solidaria**

Jenny Paola Lis-Gutiérrez  
Konrad Lorenz

11 de julio de 2018  
Quito, Ecuador



Establecer la evolución de la producción y visibilidad de la literatura académica indexada en Scopus, sobre Economía Solidaria, empleando la minería de texto

## Minería de texto

- [E]l proceso de analizar colecciones de materiales de texto con el objeto de capturar los temas y conceptos clave y descubrir las relaciones ocultas y las tendencias existentes sin necesidad de conocer las palabras o los términos exactos que los autores han utilizado para expresar dichos conceptos (IBM, 2018).

El proceso incluye los siguientes pasos

Identificar el texto que será usado.

Aplicar los algoritmos de minería de texto.

Crear categorías y conceptos.

Analizar los datos estructurados, mediante técnicas como cluster, clasificación y modelos predictivos.

## Usos

Eskici y Koçak, 2018

- Datos del Banco Central de la República de Turquía.
- Consistencia estadística del desarrollo de los precios con las cifras de inflación del índice de precios al consumidor anual para Turquía

Rodríguez-Esteban  
(2017)

- Curación semiautomatizada de bases de datos biológicas, la farmacovigilancia, el descubrimiento de biomarcadores, la construcción de vías de señalización y la predicción de la función proteica

Su & Chen, 2018

- Reconocimiento de escritura no segmentada y reconocimiento de voz

## Usos

Cheng y Jin (2019)

- Atributos que influyen en las experiencias de los usuarios de Airbnb. Análisis de comentarios de revisión en línea a través del proceso de minería de textos y análisis de sentimientos.

Zhao, Xu y Wang (2019)

- Muestra de 127.629 reseñas de tripadvisor.com, para predecir la satisfacción general del cliente

Westergaard (2018)

- Extracción de textos de la literatura científica (15 millones de artículos de texto completo en inglés, 1823-2016).

Van Eck & Waltman (2011).

- Minería de texto y visualización usando VosViewer

TITLE-ABS-KEY ( "solidarity economy" OR " économie solidaire" OR " solidarity-based economy" OR " economía solidaria" OR "economía de solidaridad" OR " economia solidaria" OR "economia de solidaridad" OR cooperativismo OR cooperativism )

La delimitación de las interrelaciones entre las palabras clave resaltadas se llevará a cabo de acuerdo con el siguiente fórmula planteada a partir del trabajo de Lulewicz-Sas (2017):

$$P(A|B) = \sum_{i=1}^n p_i(A|B)$$

$$p_i(A|B) = \begin{cases} 1 & \text{if } A \in art_i \wedge B \in art_i \\ 0 & \text{if } A \notin art_i \wedge B \notin art_i \end{cases}$$

donde:

$p_i(A|B)$ : probabilidad de selección del artículo  $i$  a partir de las condiciones  $A$  y  $B$ .

$i$ : Número de artículos científicos ( $i = 1, 2, \dots, n$ )

$A$ : conjunto de palabras de la ecuación de búsqueda

$B$ : condición del período de publicación (2000 – 2018)

$art_i$ : conjunto de artículos preseleccionados que cumplen con al menos una de las condiciones  $A$  o  $B$ .

Se consideró siguiendo a Justicia de la Torre (2017), una colección de documentos  $\Omega$  sobre la que se pretende realizar los análisis. En este caso  $\Omega = \{D_1, \dots, D_n\}$  es una colección de documentos, donde cada  $D_1$  es un documento. Por su parte, el texto se divide en unidades mínimas documento, palabra, concepto, frase y párrafo.

Las palabras que forman un texto, pueden proporcionar un primer nivel del corpus. En este contexto, una palabra se identificará como la identificamos con el símbolo  $m_j$ , donde  $j = 1, \dots, k$ , siendo  $k$  el total de palabras de un documento. Cada  $m_j$  es una secuencia de caracteres dentro de una frase. Por ello,  $VD = \{m_1, \dots, m_k\}$ , será el conjunto de palabras extraídas del documento.

Ahora bien, una bolsa de palabras podrá tener una representación mediante números reales cuando la presencia y la ausencia de un término en el documento, corresponda a un valor real. Por lo anterior, la frecuencia relativa, estaría dada por el número de veces que aparece un determinado término en un documento, calculado sobre el número de términos de dicho documento, se calculará como:

$$D_i(t_j) = \frac{f_{ij}}{\sum_{i=1}^k f_{ij}}$$

Otro de los elementos empleados para el análisis, fue tomado de los trabajos de Westergaard (2018), Franceschini *et al.* (2012) y Mørk *et al.* (2013), y está relacionado con las co-ocurrencias, que corresponden a un recuento ponderado para cada par de entidades usando la siguiente fórmula

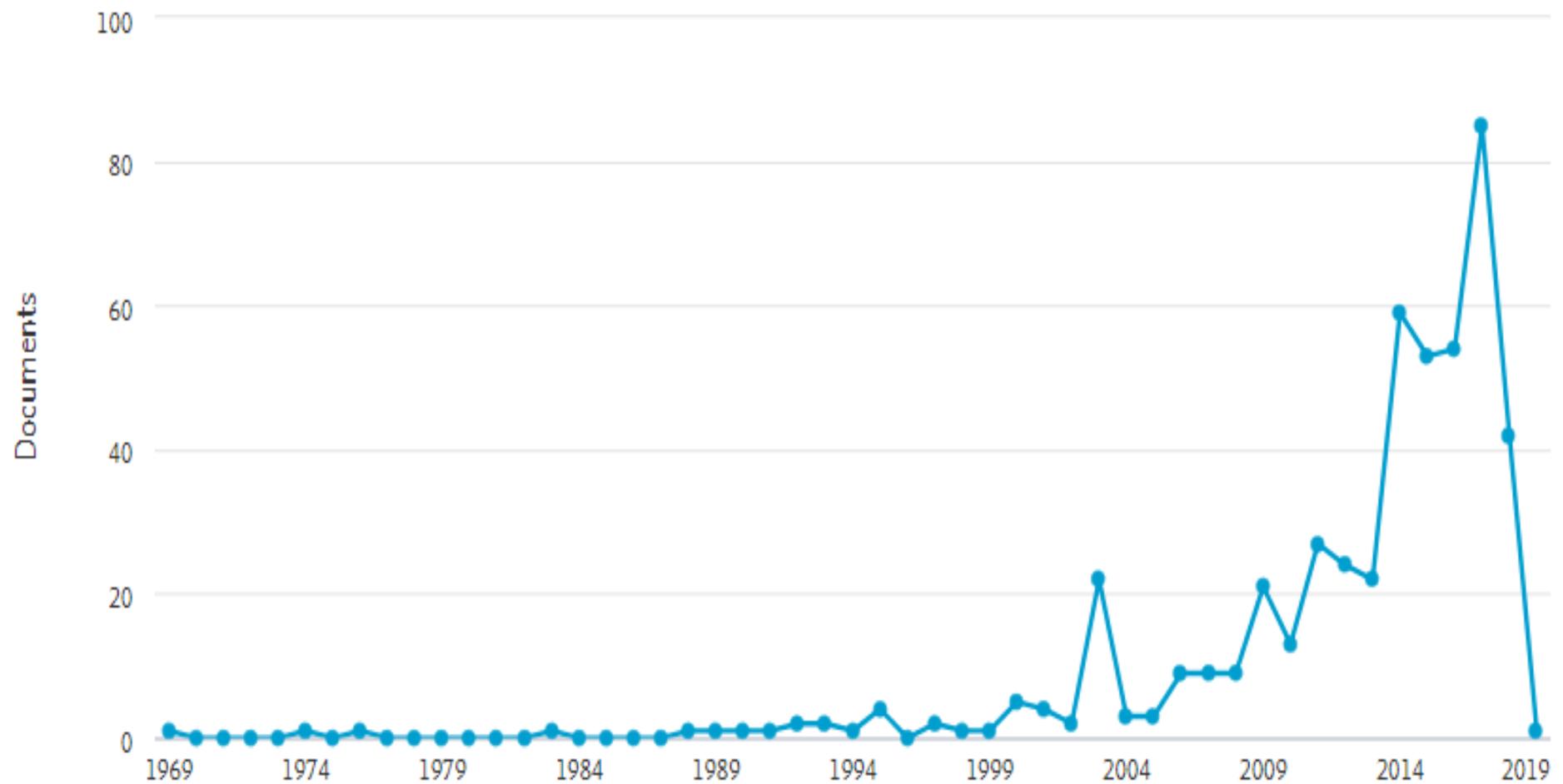
$$C(i, j) = \sum_{k=1}^n w_d \delta_{dk}(i, j) + w_p \delta_{pk}(i, j) + w_s \delta_{sk}(i, j)$$

Donde  $\delta$  es una función de indicador que toma en cuenta si los términos  $i, j$  co-ocurren dentro del mismo documento ( $d$ ), párrafo ( $p$ ) o frase ( $s$ ).  $w$  es el peso de concurrencia aquí establecido en 1.0, 2.0 y 0.2, respectivamente. Con base en el recuento ponderado, el puntaje  $S(i, j)$  se calculó usando un alfa de 0.6 ( $\alpha=0.6$ ).

$$S(i, j) = C_{ij}^{\alpha} \left( \frac{C_{ij} C_{..}}{C_{.i} C_{.j}} \right)^{1-\alpha}$$

Gráfica 1. Producción académica indexada en Scopus sobre la materia

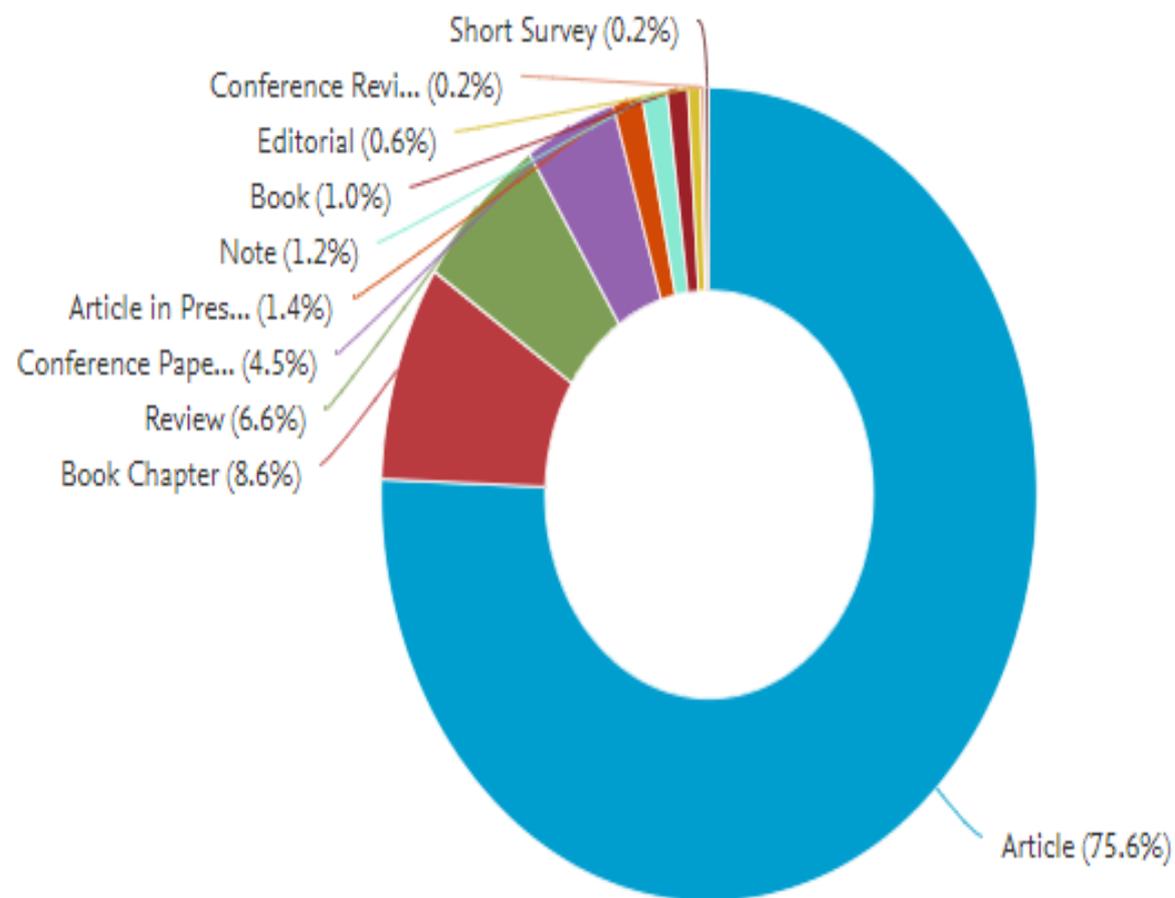
Documents by year



Fuente: elaboración propia usando Scopus.

Gráfica 2. Tipo de documento

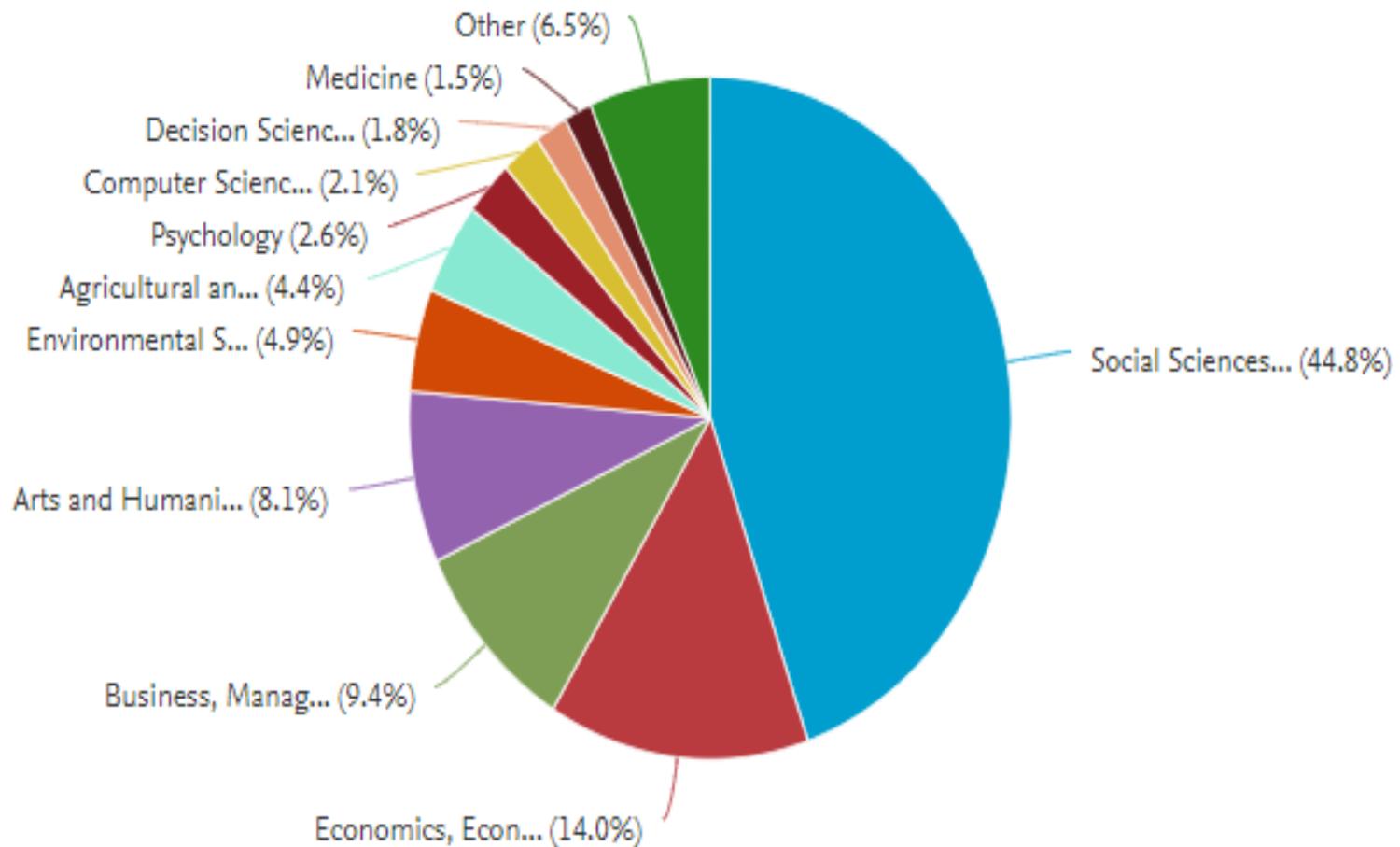
Documents by type



Fuente: elaboración propia usando Scopus.

Gráfica 3. Documento por área

Documents by subject area

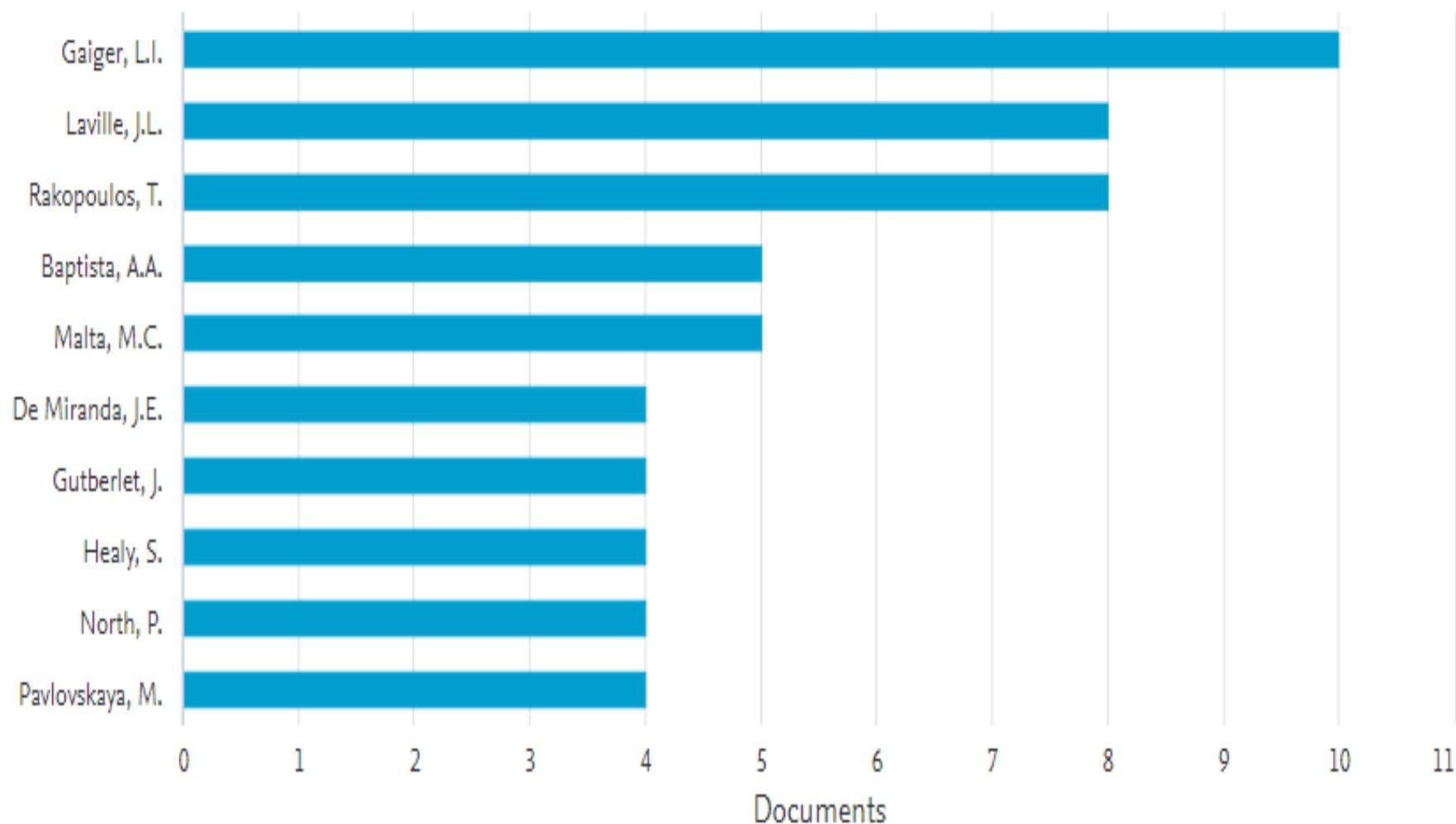


Fuente: elaboración propia usando Scopus.

## Gráfica 4. Autores con mayores publicaciones

## Documents by author

Compare the document counts for up to 15 authors

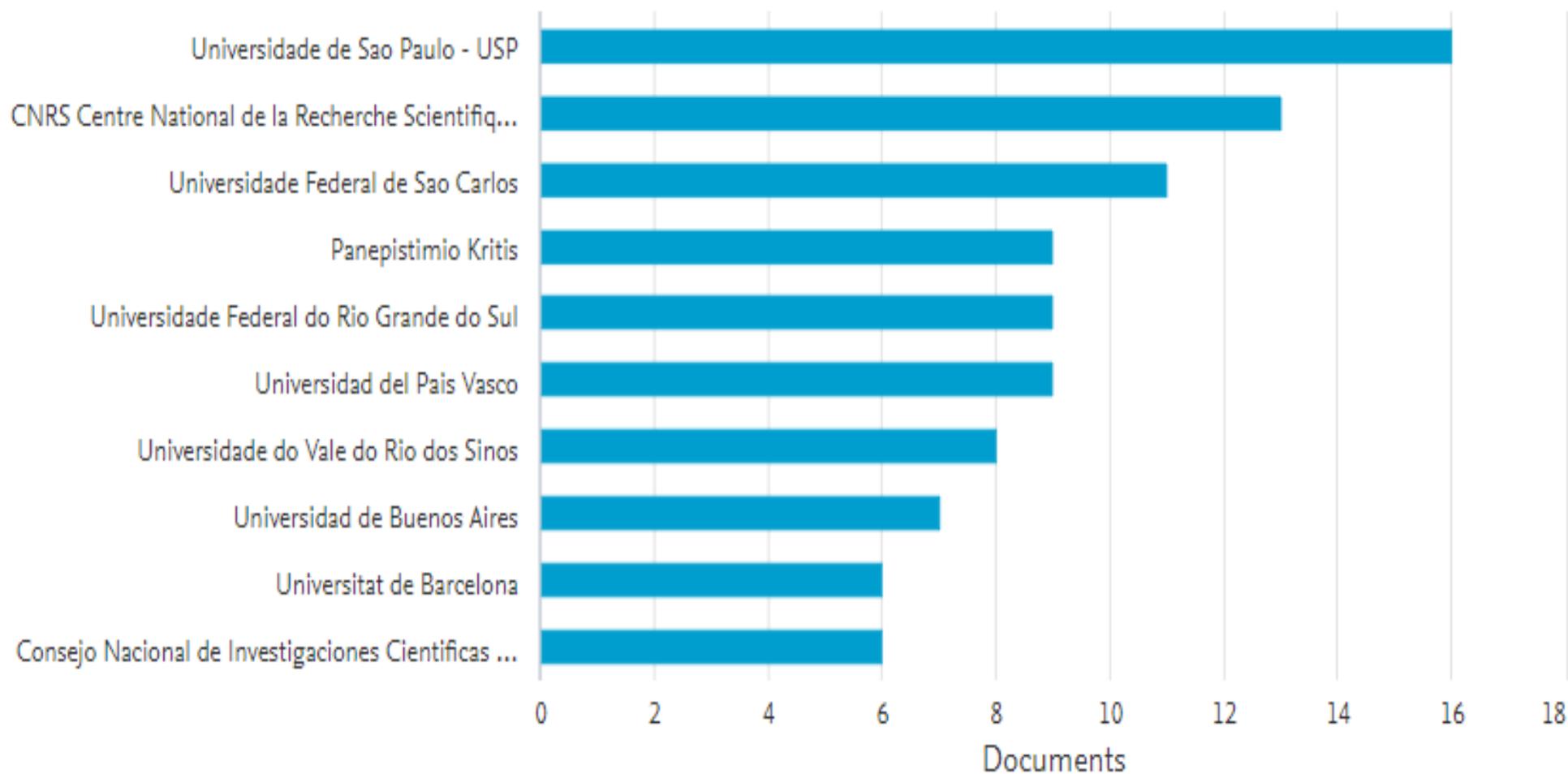


Fuente: elaboración propia usando Scopus.

## Gráfica 5. Organización con mayores documentos

### Documents by affiliation

Compare the document counts for up to 15 affiliations



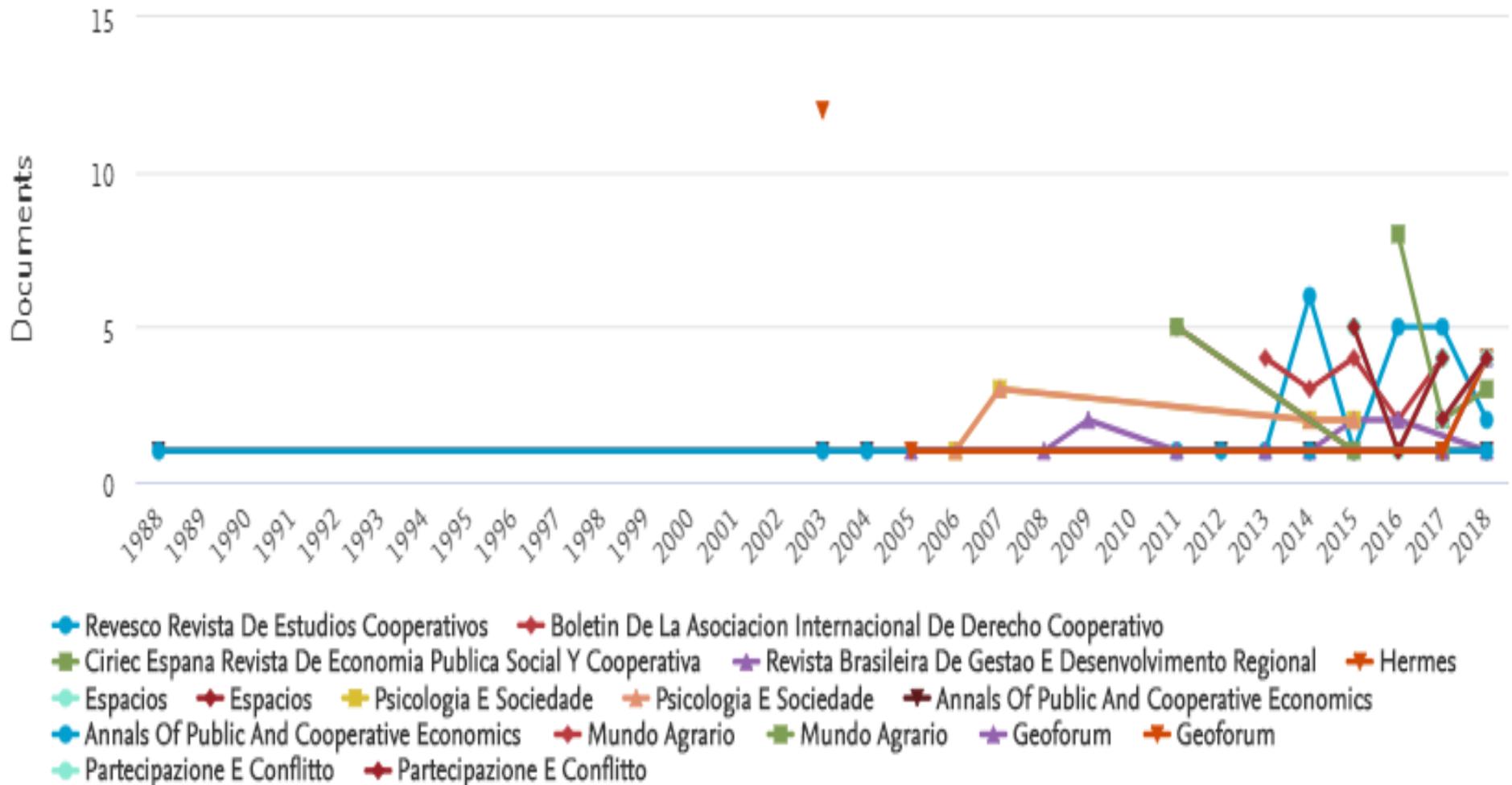
Fuente: elaboración propia usando Scopus.

Gráfica 6. Documentos por revista

Documents per year by source

Compare the document counts for up to 10 sources

Compare sources and view CiteScore, SJR, and SNIP data

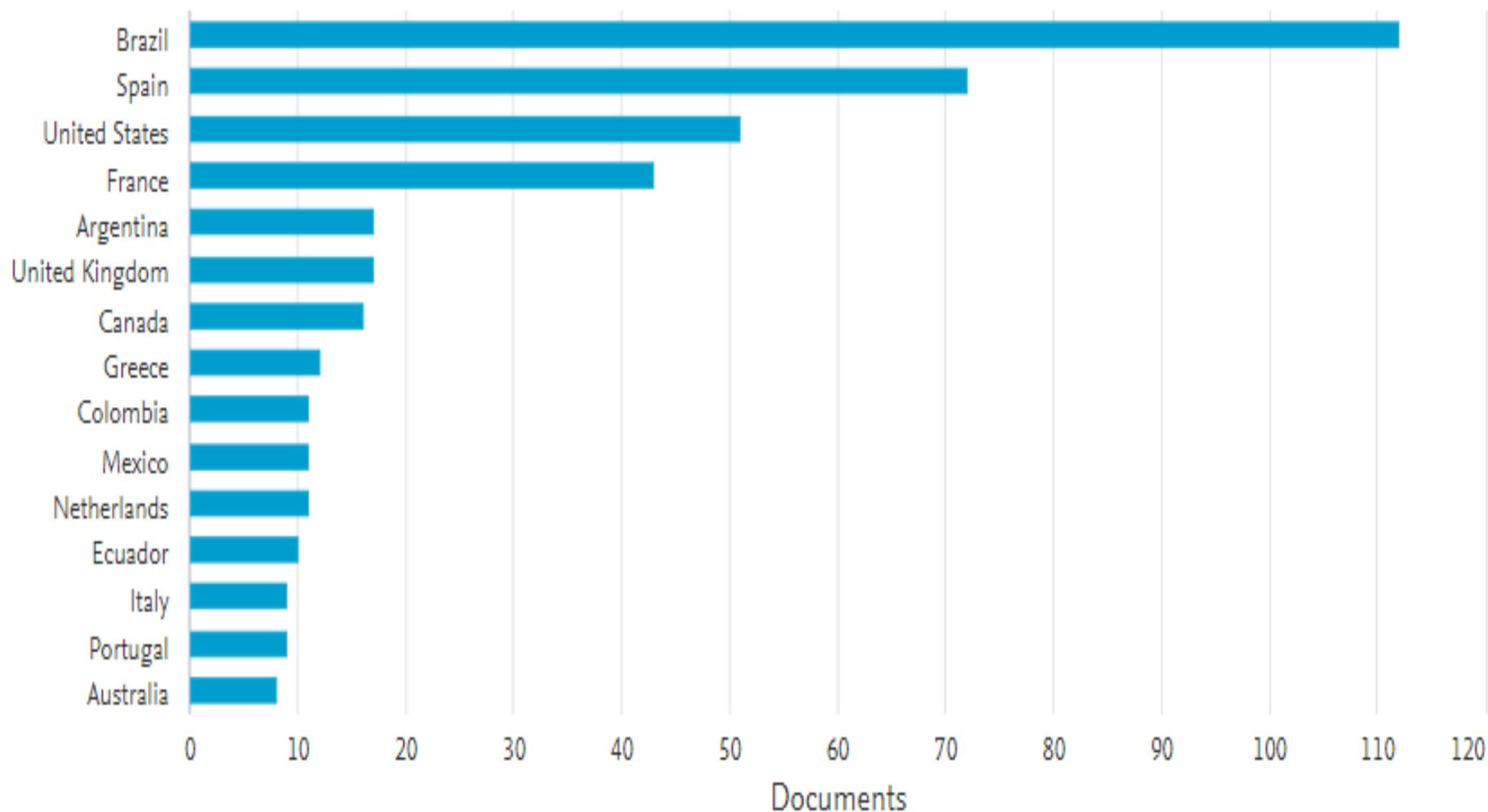


Fuente: elaboración propia usando Scopus.

## Gráfica 7. Documento por país

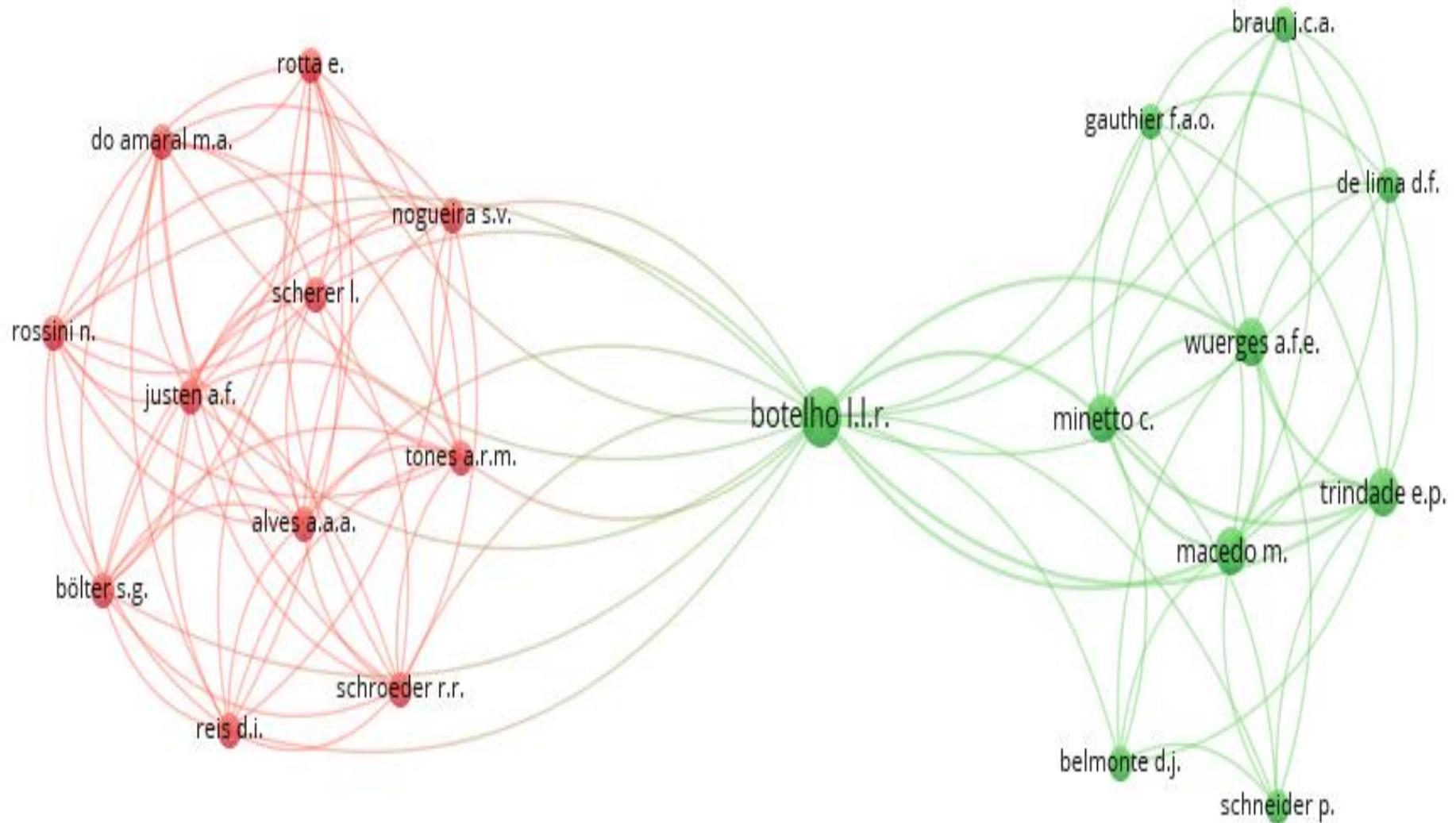
## Documents by country or territory

Compare the document counts for up to 15 countries/territories



Fuente: elaboración propia usando Scopus.

## Gráfica 8. Co-autoría autores (789)



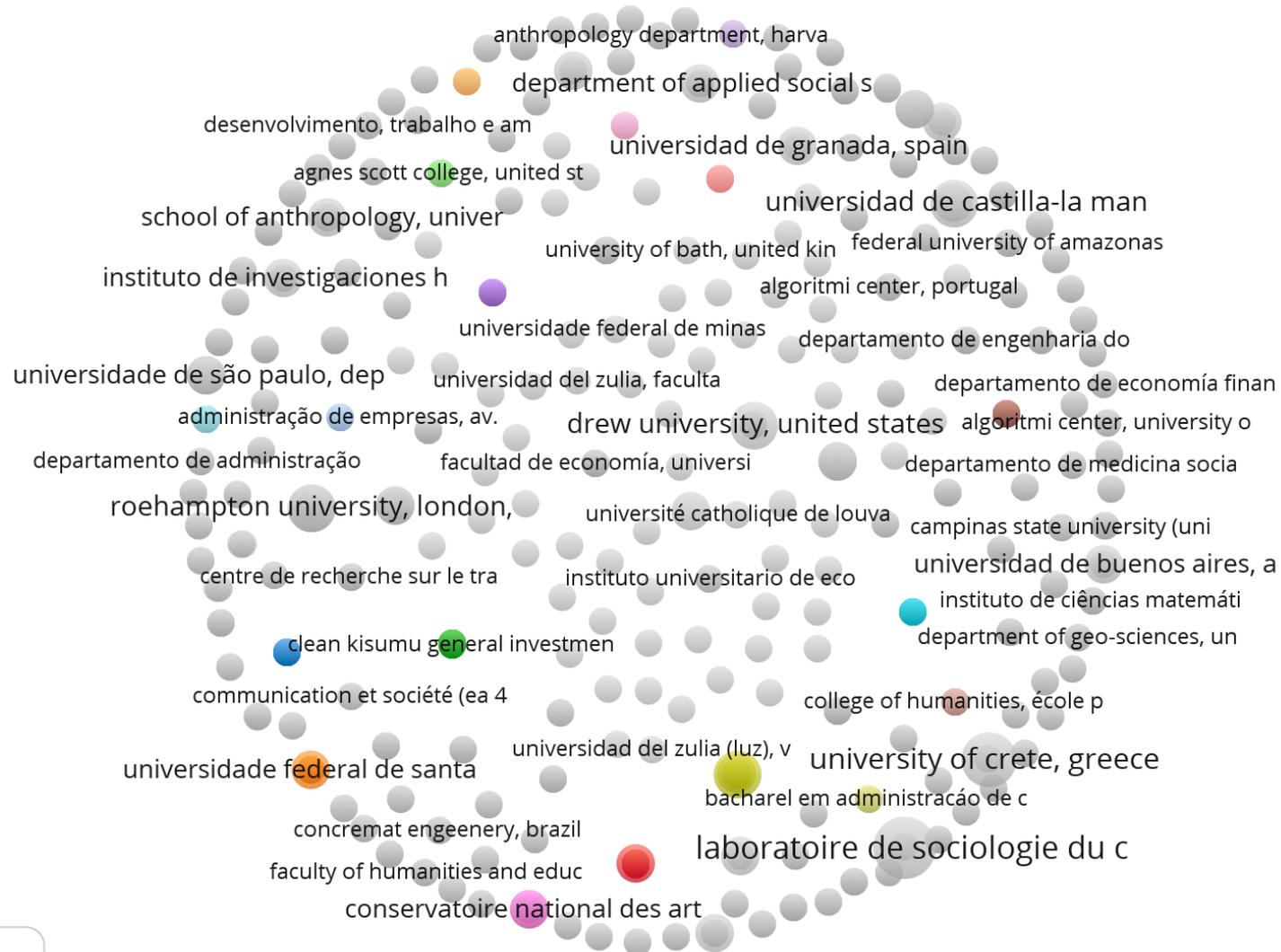
Fuente: elaboración propia usando Scopus.

Gráfica 9. Co-autoría autores (21)

id	label	x	y	cluster	weight<Links>	weight<Total link strength>	weight<Documents>
55	belmonte d.j.	0.6391	-0.371	2	6	6	1
71	botelho l.l.r.	0.1684	-0.0084	2	20	24	3
80	braun j.c.a.	0.8956	0.3987	2	7	7	1
186	de lima d.f.	10.615	0.2313	2	7	7	1
304	gauthier f.a.o.	0.6867	0.299	2	7	7	1
439	macedo m.	0.8155	-0.15	2	9	13	2
478	minetto c.	0.6118	-0.0114	2	9	13	2
659	schneider p.	0.8855	-0.4153	2	6	6	1
715	trindade e.p.	10.545	-0.0905	2	9	13	2
770	wuerges a.f.e.	0.8474	0.067	2	9	13	2
12	alves a.a.a.	-0.6453	-0.1201	1	11	11	1
91	bölter s.g.	-0.9634	-0.1896	1	11	11	1
220	do amaral m.a.	-0.8698	0.2783	1	11	11	1
377	justen a.f.	-0.825	0.0126	1	11	11	1
515	nogueira s.v.	-0.4107	0.201	1	11	11	1
600	reis d.i.	-0.7617	-0.3356	1	11	11	1
618	rossini n.	-10.392	0.0785	1	11	11	1
622	rotta e.	-0.634	0.357	1	11	11	1
658	scherer l.	-0.6265	0.1167	1	11	11	1
660	schroeder r.r.	-0.4931	-0.2952	1	11	11	1
712	tones a.r.m.	-0.3973	-0.053	1	11	11	1

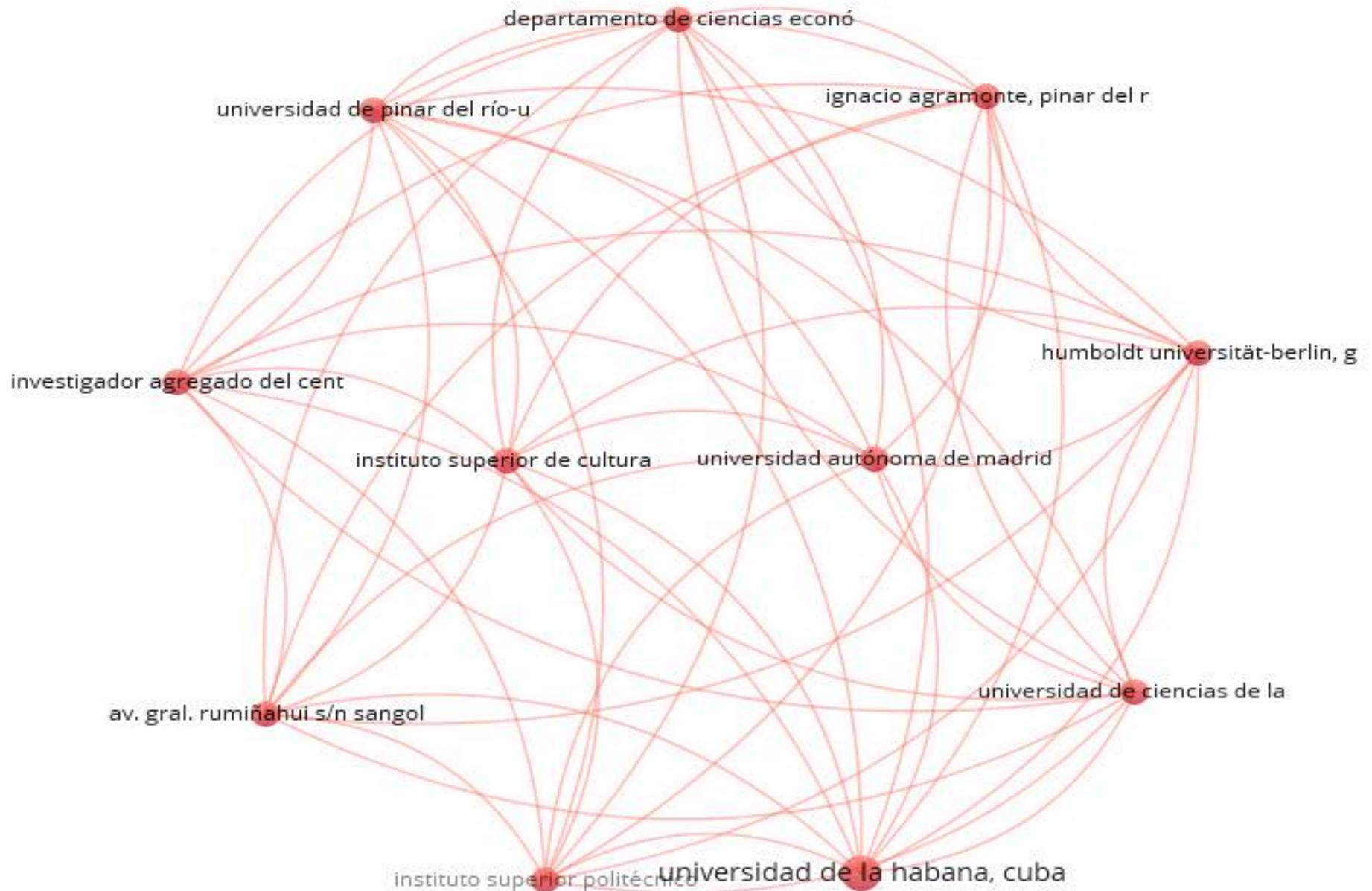
Fuente: elaboración propia usando Scopus.

# Gráfica 10. Co-autoría organizaciones (678)



Fuente: elaboración propia usando Scopus.

# Gráfica 11. Co-autoría organizaciones (11)



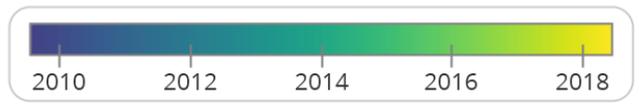
Fuente: elaboración propia usando Scopus.

# Gráfica 12. Co-autoría organizaciones (11)

id	label	x	y	cluster	weight<Links>	weight<Total link strength>	weight<Documents>	weight<Citations>	weight<Norm. citations>	score<Avg. pub. year>	score<Avg. citations>	score<Avg. norm.>
2	acadêmico de engenharia elétrica, universidade do planalto catarinense (uniplac), brazil	-0.702	-0.5965	7	4	4	1	0	0	2017	0	0
3	administration course department, federal institute of paran� (ifpr), rodovia pr 562, s/n, flor da serra, coronel vivida, paran�, brazil	0.6189	0.6735	59	1	1	1	0	0	2017	0	0
4	administra�o de empresas, av. t-15 no 800, apto. 101, setor bueno, cep 74230-010, goi�nia (go), brazil	-0.6429	0.127	12	3	3	1	0	0	2008	0	0
6	agnes scott college, united states	-0.4365	0.6262	11	3	3	1	2	64.615	2018	2	64.615
7	agroneg�cio pela ufg - goi�s, brasil, av. t-15 no 800, apto. 101, setor bueno, cep 74230-010, goi�nia (go), brazil	-0.6428	0.127	12	3	3	1	0	0	2008	0	0
9	aix-marseille university, france	0.4779	0.8018	28	2	2	1	2	0.7162	2015	2	0.7162
10	algoritmi center, portugal	0.4836	0.3931	26	2	2	1	2	0.7162	2015	2	0.7162
11	algoritmi center, universidade do minho, portugal	-0.6289	0.4236	119	1	1	1	0	0	2017	0	0

Fuente: elaboraci n propia usando Scopus.

# Gráfica 13. Co-autoría país (52)



Fuente: elaboración propia usando Scopus.



Gráfica 15. Co-autoría país (27)

id	label	x	y	cluster	weight<Links>	weight<Total link strength>	weight<Documents>	weight<Citations>	weight<Norm. citations>	score<Avg. pub. year>	score<Avg. citations>	score<Avg. norm. citations>
5	belgium	-0.859	0.0852	1	2	3	5	33	110.439	2014.4	6.6	22.088
8	brazil	-0.4211	-0.0285	1	10	16	111	102	418.776	20.139.009	0.9189	0.3773
13	croatia	-0.5341	-0.23	1	5	5	4	1	0.3224	2015.75	0.25	0.0806
17	estonia	-10.781	0.0455	1	1	1	1	3	2.931	2017	3	2.931
18	france	-0.5272	0.08	1	5	6	44	94	262.597	20.096.818	21.364	0.5968
36	netherlands	-0.8943	0.0264	1	7	9	11	56	317.638	20.156.364	50.909	28.876
37	new zealand	-0.7369	-0.142	1	3	3	1	0	0	2016	0	0
2	argentina	-0.2756	-0.1446	2	7	9	17	27	138.361	20.122.353	15.882	0.8139
20	greece	0.4653	-0.1709	2	4	5	13	51	496.877	20.172.308	39.231	38.221
26	ireland	0.2056	-0.0898	2	5	6	4	4	61.714	2017.25	1	15.429
33	mexico	-0.069	-0.1496	2	2	3	12	5	92.154	20.109.167	0.4167	0.7679
4	austria	13.296	-0.1337	3	1	1	1	0	0	2009	0	0
19	germany	11.644	-0.1055	3	8	8	8	46	199.358	2.013.625	5.75	2.492
47	south korea	13.129	-0.1227	3	1	1	1	1	0.3117	2012	1	0.3117
58	venezuela	13.287	-0.1077	3	1	1	5	2	0.6769	2013	0.4	0.1354
11	colombia	0.1491	0.1113	4	4	5	11	2	12.673	20.156.364	0.1818	0.1152
14	cuba	0.6275	0.2519	4	4	5	6	1	15.429	20.161.667	0.1667	0.2571
16	ecuador	0.3815	0.163	4	7	8	10	3	2.878	2016.6	0.3	0.2878
48	spain	0.4381	0.1498	4	7	10	73	142	507.601	20.120.411	19.452	0.6953
3	australia	-0.1714	0.9594	5	1	3	8	94	239.899	2.010.625	11.75	29.987
31	malaysia	-0.1458	0.9617	5	1	1	1	0	0	2015	0	0
55	united states	-0.1329	0.7412	5	9	12	50	332	798.263	2012.9	6.64	15.965
9	canada	-0.5695	-0.3462	6	7	10	16	74	269.815	20.141.875	4.625	16.863
29	kenya	-0.7417	-0.5078	6	2	2	1	6	92.571	2016	6	92.571
49	sweden	-0.6895	-0.4569	6	4	5	2	6	92.571	2016.5	3	46.286
39	peru	0.2366	-0.4514	7	1	1	1	6	0.8387	2010	6	0.8387
54	united kingdon	0.2067	-0.3879	7	5	5	17	115	429.101	20.155.294	67.647	25.241

Fuente: elaboración propia usando Scopus.





Gráfica 18. Co-ocurrencia palabras clave (1708, 5 KW, 62)



Fuente: elaboración propia usando Scopus.

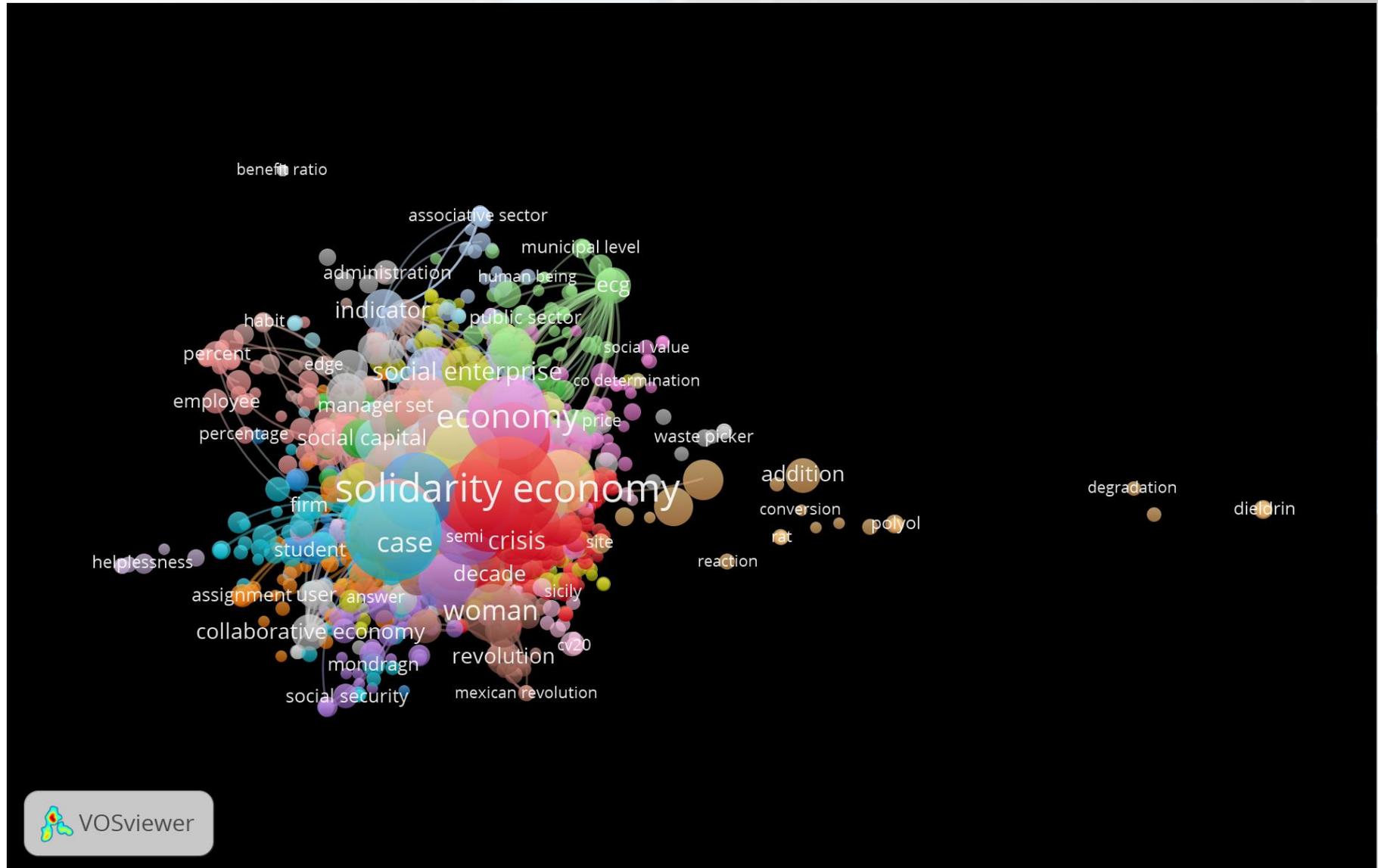
## Co-ocurrencia palabras clave (1708, 5 KW, 62)

### DISCUSIÓN Y RESULTADOS

	Palabras	Descripción
1	agricultural cooperative, agricultural cooperatives, Brazil, capitalism, cooperative sector, development, Ecuador, education, historical perspective, latin america, neoliberalism, political economy, social movement, sustainability, sustainable development, United States	Agricultura y medio ambiente
2	Governance, local development, management, public policy, social and solidarity economy, social and solidarity-based economy, social capital, social entrepreneurship, social innovation, social movements, third sector	Gobernanza y movimientos sociales
3	Article, civil and solidarity-based economy, civil society, democracy, human, human engineering, participation, policy, public sphere, recycling, solidarity-based economy	Políticas públicas y economía social
4	Cooperative, crisis, economic development, entrepreneurship, gender, Greece, solidarity	Desarrollo
5	Argentina, citizenship, cooperativism, poverty, self-management, solidary economy, work	Ciudadanía
6	community economy, economics, ethics, social economy, social enterprise, solidarity economy	Economía social
7	Cooperatives, public policies, rural development, Spain	Cooperativas y desarrollo rural

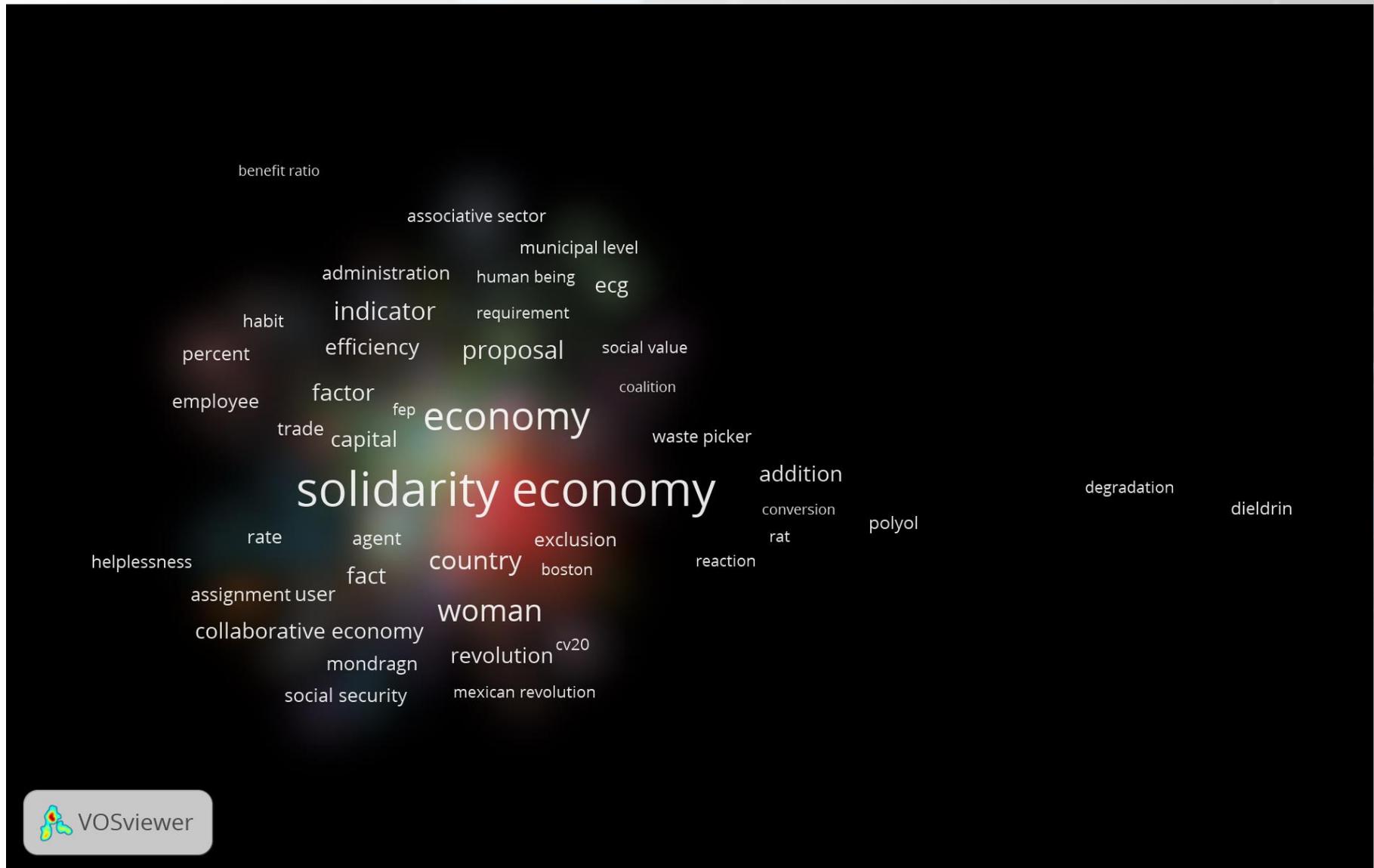
Fuente: elaboración propia usando Scopus.

# Gráfica 19. Títulos y abstracts (10.304, frecuencia=2, 2464 items



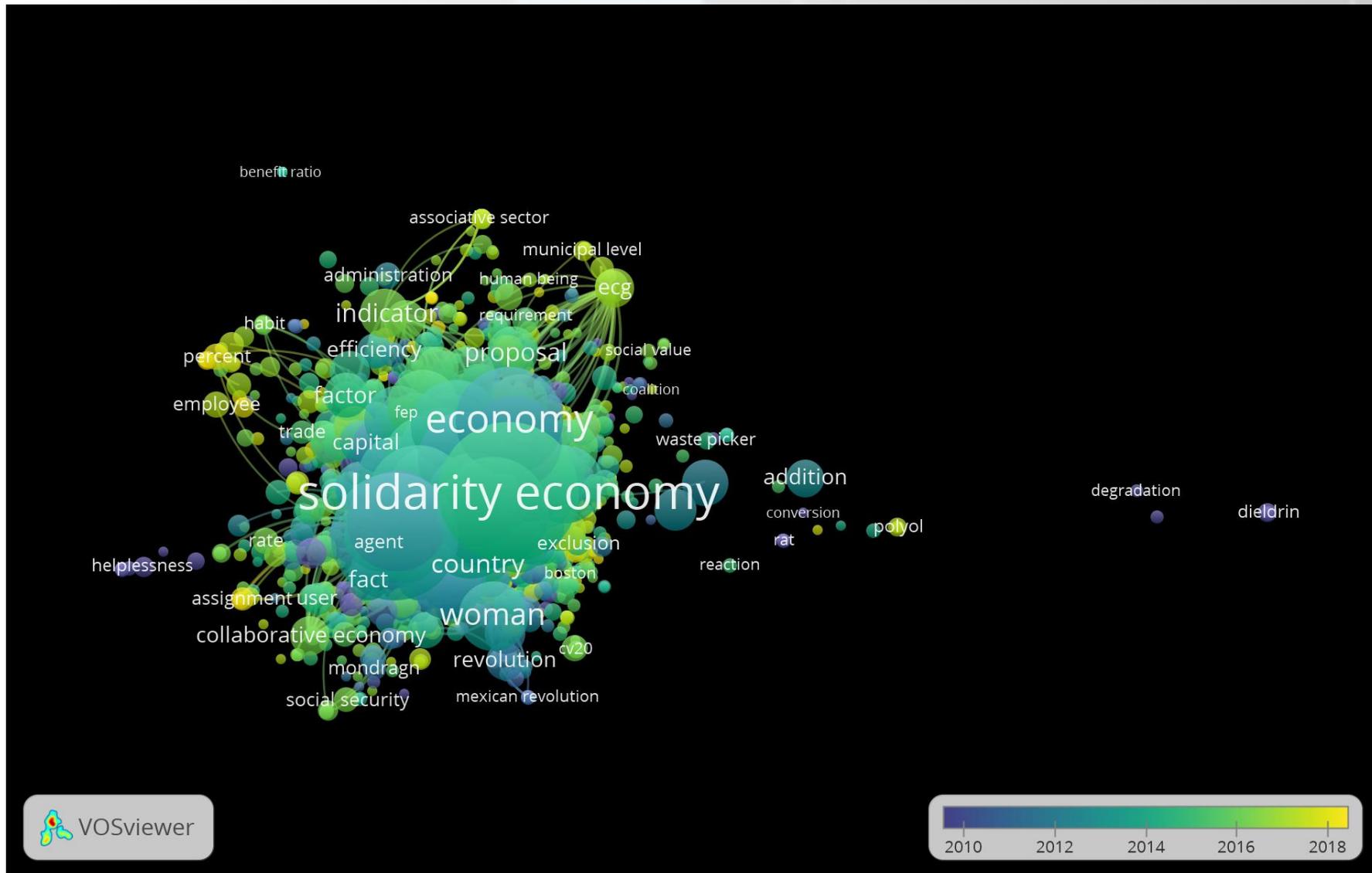
Fuente: elaboración propia usando Scopus.

## Gráfica 20. Títulos y abstracts (10.304, frecuencia=2, 2464 items



Fuente: elaboración propia usando Scopus.

Gráfica 21. Títulos y abstracts (10.304, frecuencia=2, 2464 items)



Fuente: elaboración propia usando Scopus.



En este trabajo se ha realizado un ejercicio para la minería de datos textuales, que extrajo información de uso práctico de una Fuente de Big Data como Scopus

La investigación trató de establecer algunas regularidades en la estructura de la información, de los documentos cubiertos por Scopus, que reflejan las similitudes y diferencias entre las instituciones de investigación y las universidades.

Analizar las publicaciones científicas cualitativa y cuantitativamente, evaluándolas y comparándolas con las investigaciones en el extranjero (en este caso) debe proporcionar una importante fuente de datos para los investigadores que deseen publicar.

En primer lugar, se exploró la variación anual de publicaciones de autores de todos los países y se estimó la visibilidad de la investigación llevada a cabo por autores latinos.

Luego, las redes de coautoría se analizaron estableciendo conexiones de al menos dos publicaciones por país por año. Finalmente, se aplicaron técnicas de minería de texto para identificar semiautomáticamente los temas o temas más recurrentes de la asignatura, en función de la información disponible en los resúmenes de los documentos.

Con respecto a este último punto, utilizamos las técnicas como análisis textual: recuento de palabras (análisis cualitativo), agrupamiento de textos (similitudes de texto), clasificación de textos (diferencias de texto).

Se identificaron 7 tipos de agrupaciones en cuanto a la colaboración internacional, con respecto a los países

Dentro de las opciones de trabajos futuros valdría la pena realizar este análisis estudiando la interacción entre los cluster de países y las aglomeraciones de temáticas identificadas.

Otra aproximación, es realizar un metanálisis de los trabajos para un año específico o una de las subtemáticas identificadas en este documento.

VosViewer

Orange



Jenny Paola Lis-Gutiérrez  
Jenny.lis@konradlorenz.edu.co



SEMINARIO  
INTERNACIONAL



# ECONOMÍA

POPULAR Y SOLIDAR

[WWW.CIDECUADOR.COM](http://WWW.CIDECUADOR.COM)

Una vez finalizado el evento esta presentación  
será publicada en su respectiva página web

