

Sistema científico argentino :

Interdisciplina y tópicos de investigación

Maximilian Rey, Mirko Yves Bahoz Torrico, Nicolas Gonzalez Villagra y
Matias Ramdan Ferressini

Datasets

Estos son datos públicos generados, guardados y publicados por organismos de gobierno de la República Argentina.

1122
DATASETS

Temas



Organizaciones

Ministerio de Ciencia, Tecnología e Innovación (6) x

Etiquetas

Ciencia (3)

Innovación (3)

Tecnología (3)

GÉNERO (2)

Instituciones (2)

INSTITUCIONES CIENTÍFICO-TECNOLÓGICAS (2)

Becarios (1)

Becas (1)

¿Qué datasets buscás?



Ordenar por:

Última modificación



Personal de ciencia y tecnología

Sistema de Información de Ciencia y Tecnología Argentino (SICYTAR)

El personal de CyT en una unidad estadística incluye a todas las personas involucradas directamente en I+D así como a aquellas que...



CSV JSON

Proyectos de ciencia, tecnología e innovación

Sistema de Información de Ciencia y Tecnología Argentino (SICYTAR)

Son unidades estadísticas que incluyen los proyectos de I+D y/o de innovación (conjunto de actividades que se llevan a cabo para crear...



CSV JSON

Producciones científico-tecnológicas

Sistema de Información de Ciencia y Tecnología Argentino (SICYTAR)

Son unidades estadísticas que incluyen producciones científico-tecnológicas descriptas por el tipo de producción, sus autores, camp...



CSV JSON

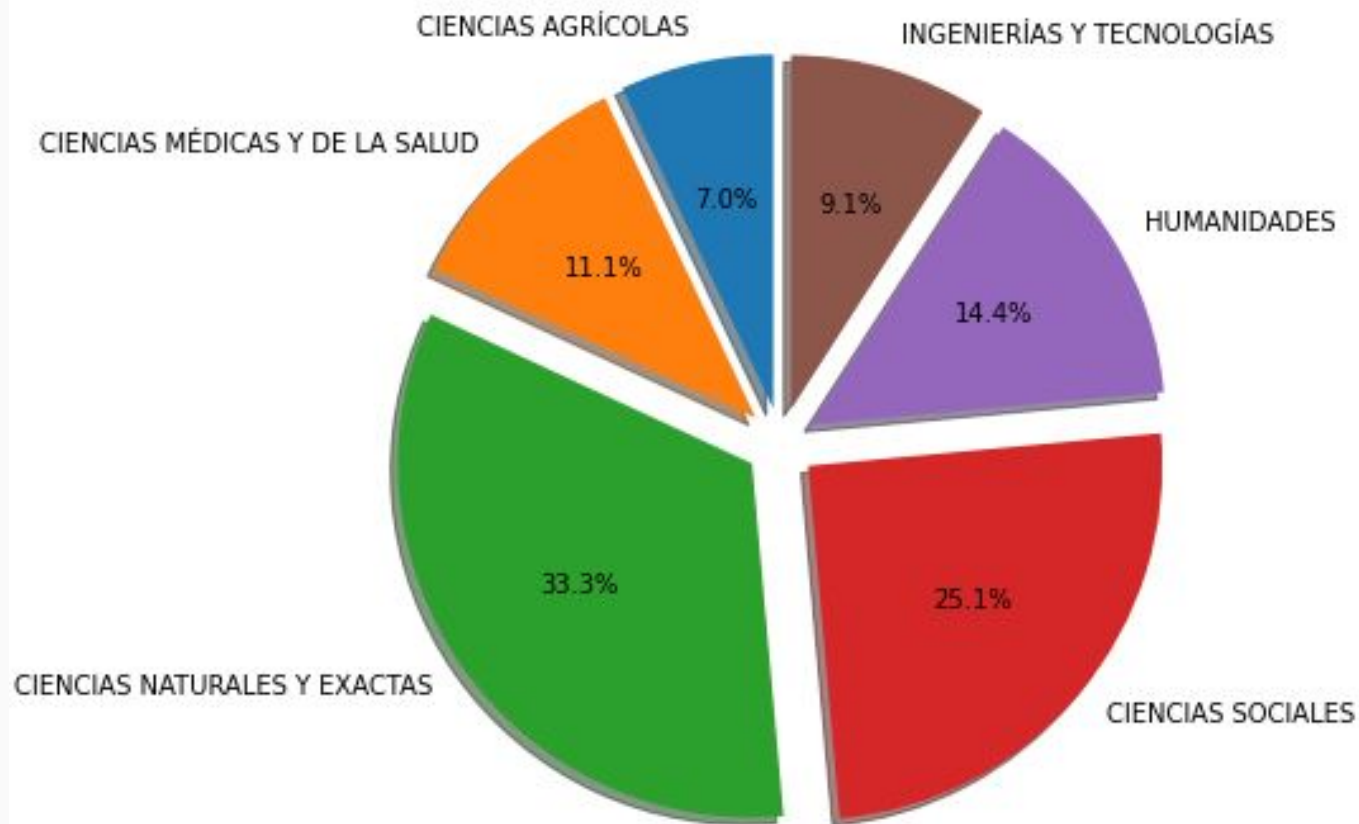
SICYTAR

Producciones científico-tecnológicas

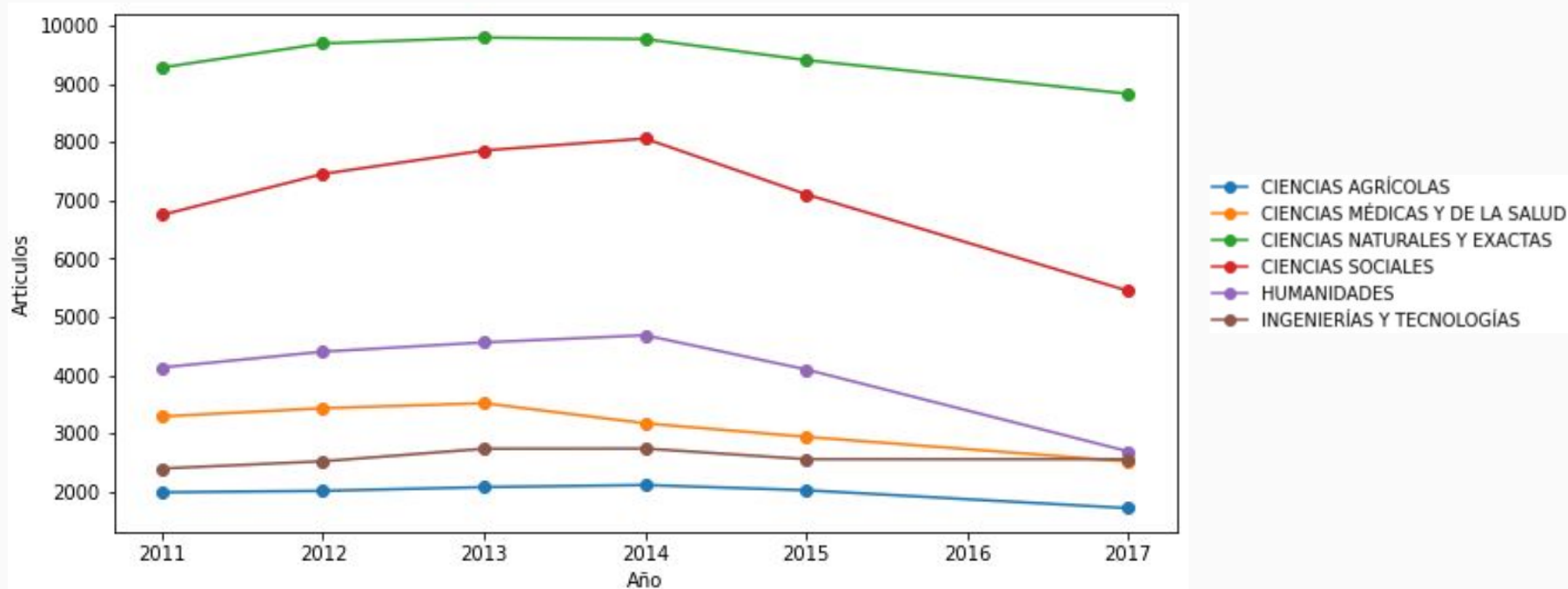
La base de datos comprende más de **155.000** artículos científicos producidos por más de **43.000** investigadores.

Los artículos se dividen en **233** disciplinas agrupadas en **42** disciplinas de área y **6** disciplinas de gran área comprendidas en los años **2011-2015** y **2017**.

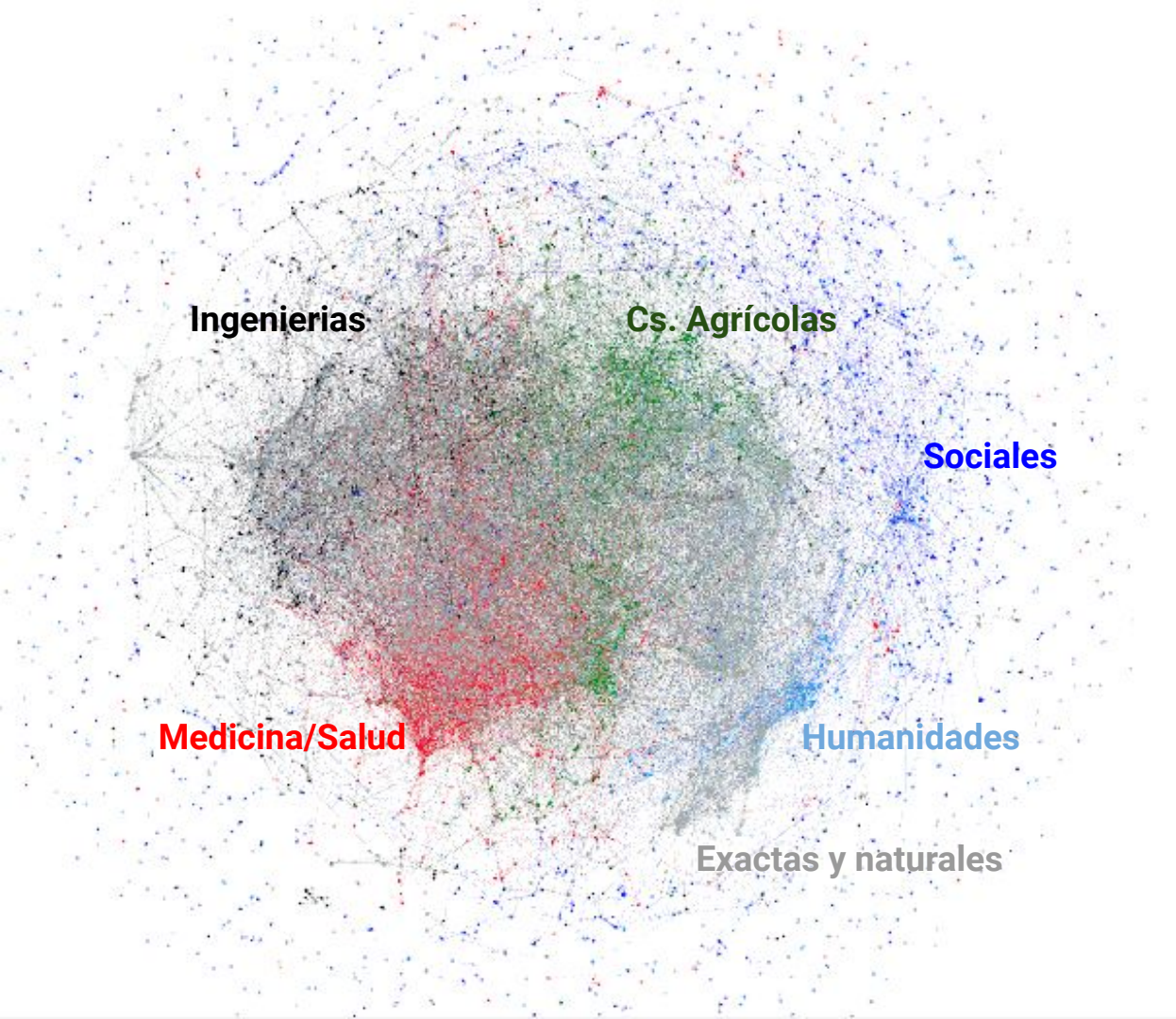
Disciplinas de gran área



Producción periodo 2011-2017



Construcción del grafo



-**Nodos:** autores de trabajos (31337)

-**Aristas:** +1 por coautoría (118014)

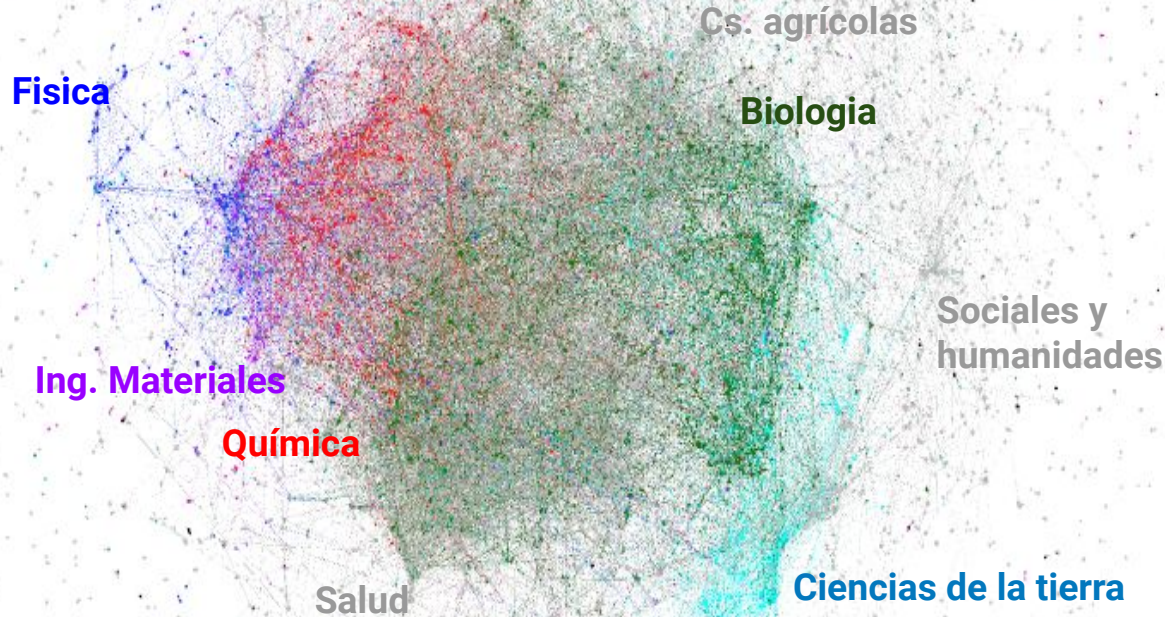
-**Curado:**

- nodos aislados
- nodos de grado 1

-**Procesamiento:**

- NetworkX -> Gephi
- Algoritmo OpenOrd

Exactas y naturales



-**Nodos:** autores de trabajos (31337)

-**Aristas:** +1 por coautoría (118014)

-**Curado:**

- nodos aislados
- nodos de grado 1

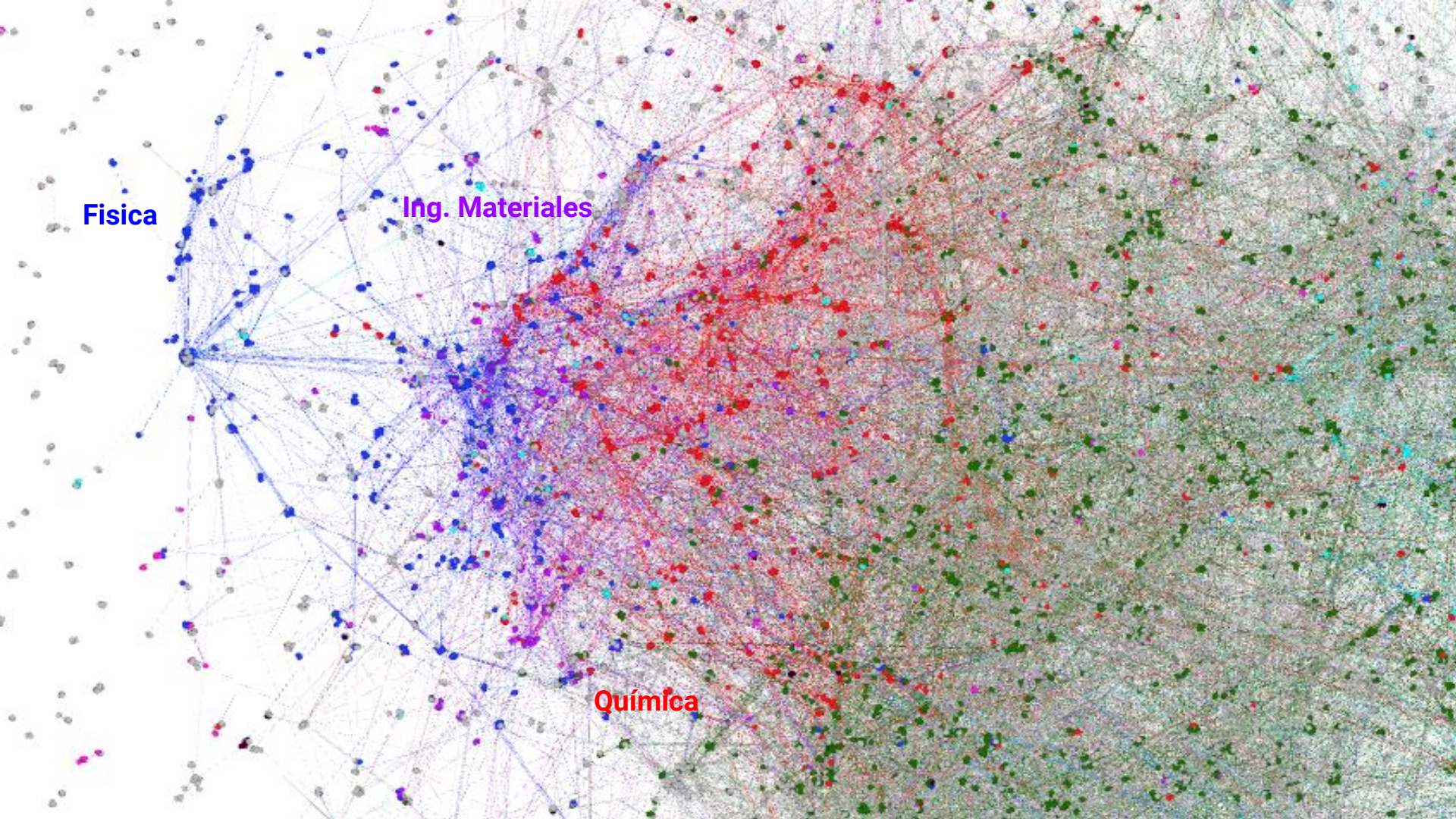
-**Procesamiento:**

- NetworkX -> Gephi
- Algoritmo OpenOrd

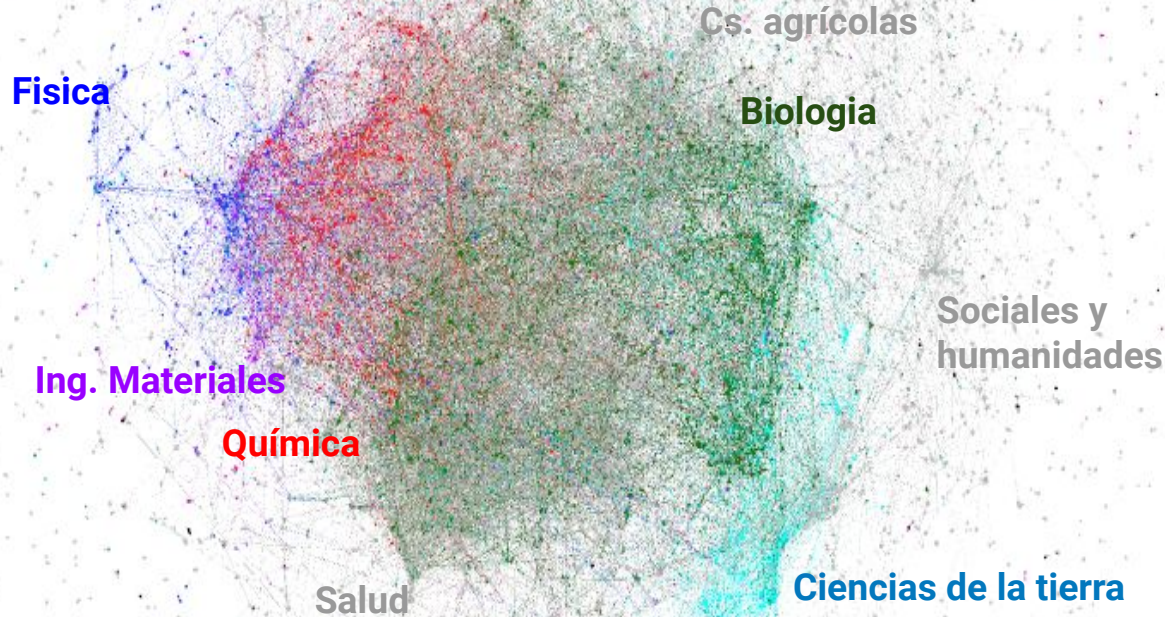
Física

Ing. Materiales

Química



Exactas y naturales



-**Nodos:** autores de trabajos (31337)

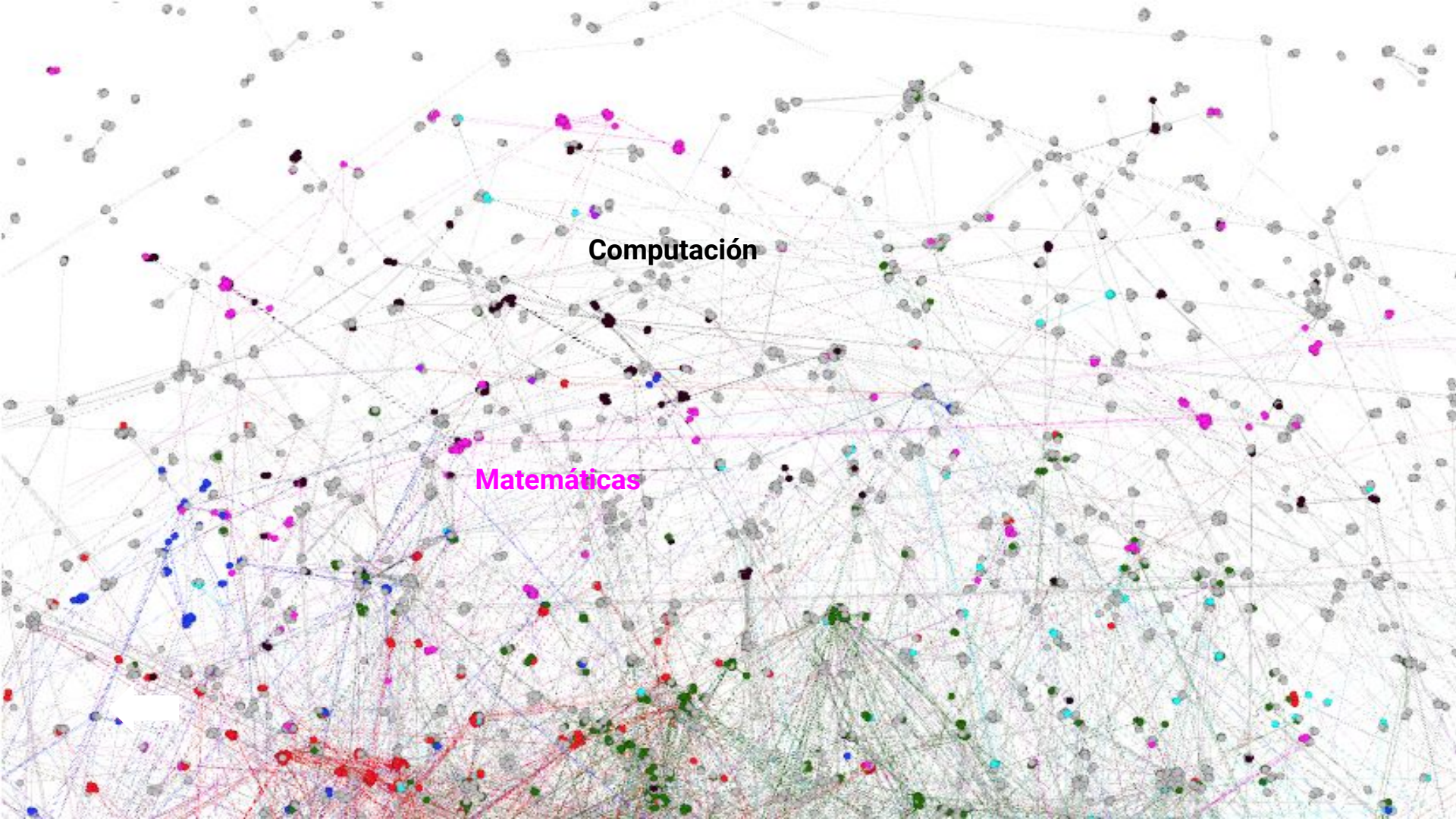
-**Aristas:** +1 por coautoría (118014)

-**Curado:**

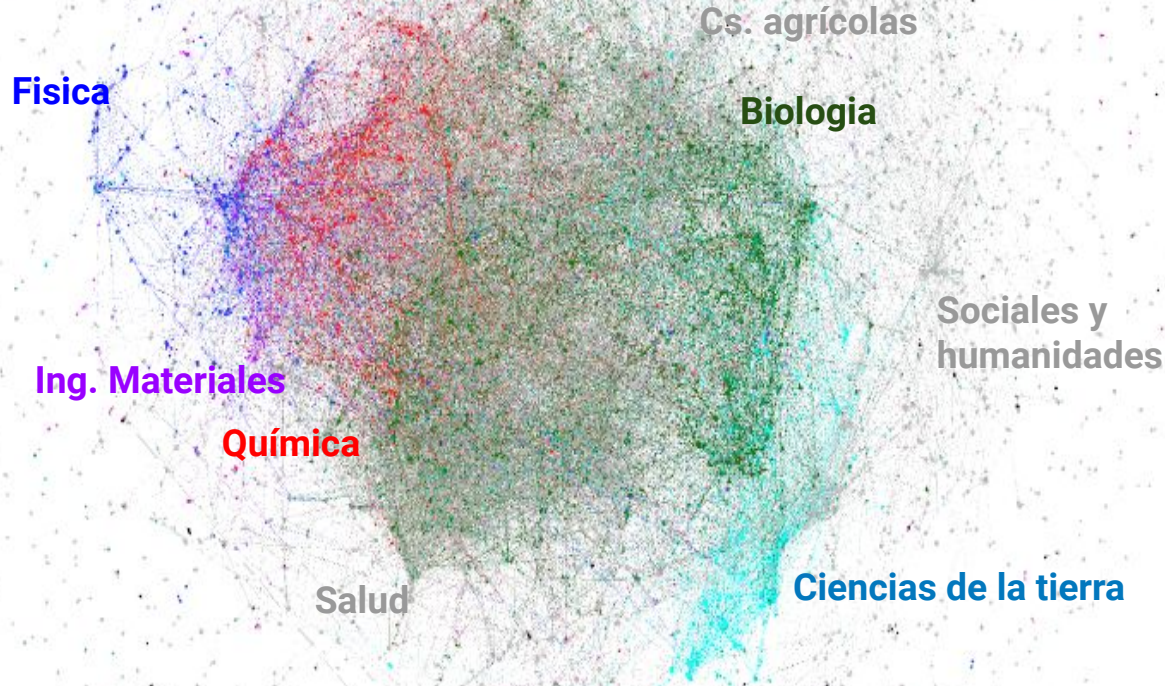
- nodos aislados
- nodos de grado 1

-**Procesamiento:**

- NetworkX -> Gephi
- Algoritmo OpenOrd



Exactas y naturales



-**Nodos:** autores de trabajos (31337)

-**Aristas:** +1 por coautoría (118014)

-**Curado:**

- nodos aislados
- nodos de grado 1

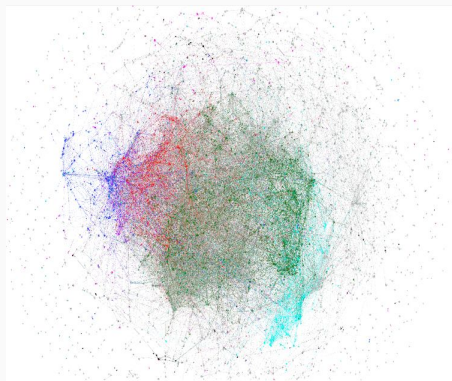
-**Procesamiento:**

- NetworkX -> Gephi
- Algoritmo OpenOrd



Computación

¿Cómo encontrar la interdisciplina?



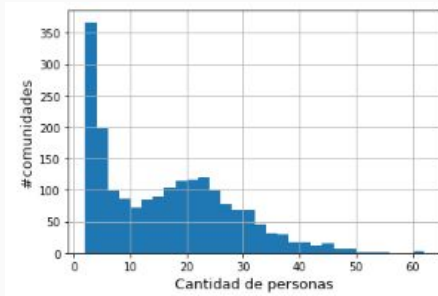
Detección de comunidades por Modularidad

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

Randomize : on
Use edge weights : on
Resolution: 0.0105
Total comunidades : 1961



persona_id	modularity_class
2665	486
3403	486
3903	486



Búsqueda de grupos interdisciplinarios

- La información de las disciplinas está estructurada de forma jerárquica en el dataset original:



Grandes áreas: Naturales y Exactas, Ingenierías y tecnologías, Ciencias Médicas/Salud, Ciencias Agrícolas, Humanidades, Sociales.

Áreas: Ciencias Físicas, Ciencias Químicas, Matemáticas...

Disciplinas : Física de los Materiales Condensados, Física de partículas y campos, Física de los fluidos y plasma...

¿Cómo encontrar la interdisciplina?

persona_id	modularity_class
2665	486
3403	486
3903	486

- Cada persona pertenece a una comunidad.
- Tiene asociada la información de su disciplina.
- Podemos sacar el ratio de personas que pertenecen a una dada disciplina en una comunidad
- Buscamos comunidades que simultáneamente tengan ratios altos de la disciplina X y la disciplina Y.
- Exploramos la comunidad

¿Cómo encontrar la interdisciplina?

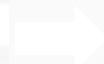
- Ejemplo : Busco al nivel mayor de Gran Area, comunidades con más de 10 personas, con Exactas >30 % y Humanidades > 30%.

	modularity_class	gran_area_descripcion_x	ratio_x	gran_area_descripcion_y	ratio_y	personas_por_disciplina_x	personas_por_disciplina_y
0	20	CIENCIAS NATURALES Y EXACTAS	0.375000	HUMANIDADES	0.416667	9	10
1	222	CIENCIAS NATURALES Y EXACTAS	0.428571	HUMANIDADES	0.428571	6	6
2	293	CIENCIAS NATURALES Y EXACTAS	0.400000	HUMANIDADES	0.400000	6	6
3	844	CIENCIAS NATURALES Y EXACTAS	0.583333	HUMANIDADES	0.333333	14	8
4	1006	CIENCIAS NATURALES Y EXACTAS	0.486486	HUMANIDADES	0.378378	18	14
5	1267	CIENCIAS NATURALES Y EXACTAS	0.307692	HUMANIDADES	0.307692	4	4
6	1855	CIENCIAS NATURALES Y EXACTAS	0.368421	HUMANIDADES	0.421053	7	8

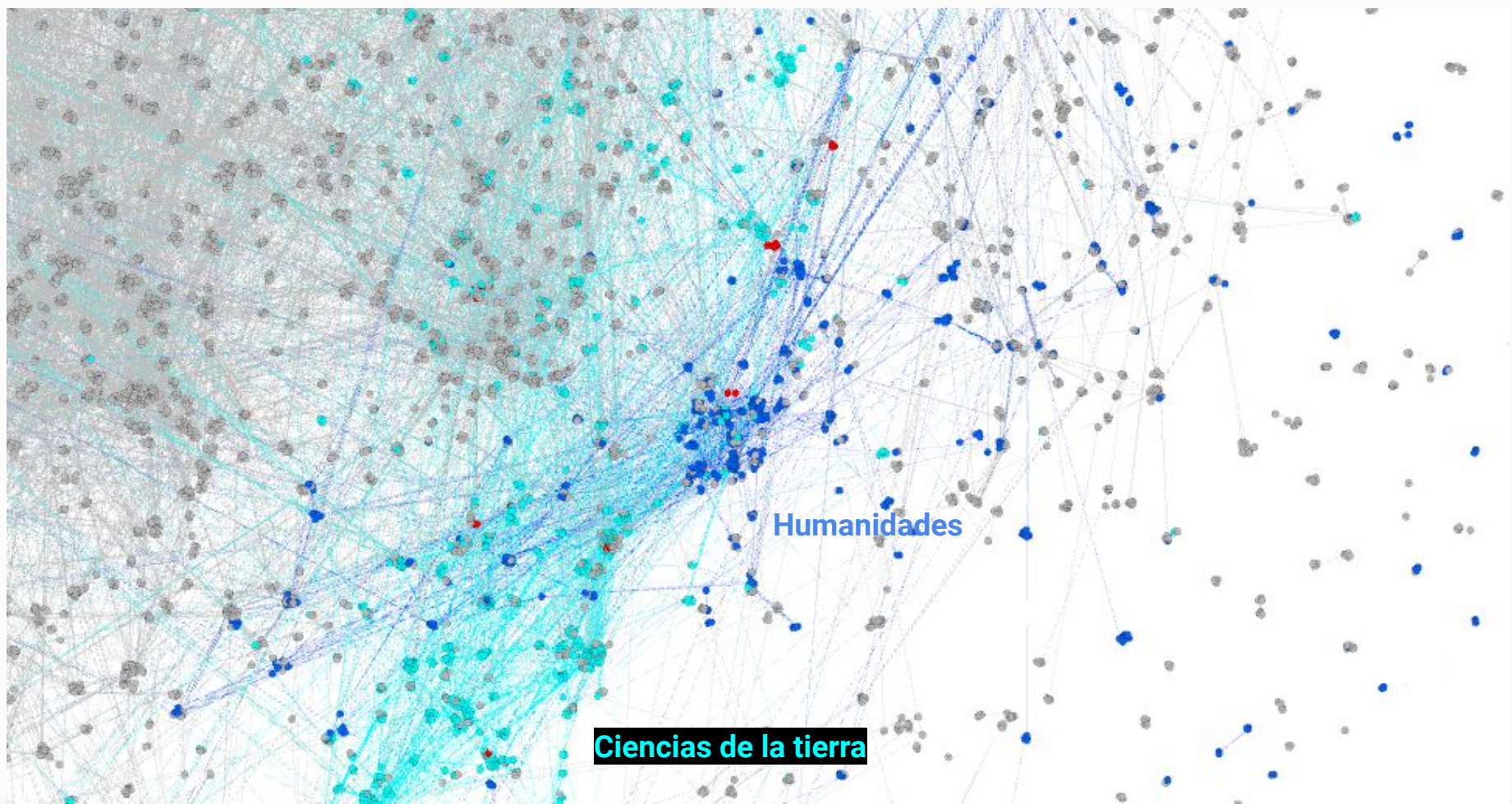
	nombre	apellido	gran_area_descripcion	disciplina_descripcion
298	MARIA MARTA	SAMPIETRO VATTUONE	CIENCIAS NATURALES Y EXACTAS	Ciencias Medioambientales (los aspectos sociales van en 5.7 "Geogr...
1216	EDUARDO ALBERTO	GÓMEZ	CIENCIAS NATURALES Y EXACTAS	Geociencias multidisciplinaria
1443	JUAN PABLO	PÉREZ PANERA	CIENCIAS NATURALES Y EXACTAS	Geología
2088	GABRIELA CATALINA	CUSMINSKY	CIENCIAS NATURALES Y EXACTAS	Geociencias multidisciplinaria
2784	MARCELO	CARDILLO	SIN DATOS	SIN DATOS
2805	FEDERICO	SCARTASCINI	HUMANIDADES	Arqueología
2849	MARCELO SEBASTIÁN	MOYANO	CIENCIAS NATURALES Y EXACTAS	Geología
3674	DIEGO SEBASTIÁN	FERNÁNDEZ	CIENCIAS NATURALES Y EXACTAS	Otras Ciencias de la Tierra y relacionadas con el Medio Ambiente
3711	MARÍA LYDIA VALENTINA	CALVO MARCILESE	CIENCIAS NATURALES Y EXACTAS	Paleontología
12043	FERNANDO	GOMEZ	CIENCIAS NATURALES Y EXACTAS	Geociencias multidisciplinaria
13282	JIMENA	ROLDAN	HUMANIDADES	Arqueología
13764	MARIO GABRIEL	MALDONADO	HUMANIDADES	Arqueología
13971	GRACIELA DE FÁTIMA	SÁNCHEZ	SIN DATOS	SIN DATOS
14199	LILIANA DEL VALLE	NEDER	CIENCIAS NATURALES Y EXACTAS	Geociencias multidisciplinaria
14891	ROSANA DEL VALLE	RIOS	CIENCIAS NATURALES Y EXACTAS	Geología
15543	ALVARO JOSÉ	CORDOMÍ	HUMANIDADES	Arqueología
16202	MARIA GISELA	LEFEBVRE	CIENCIAS NATURALES Y EXACTAS	Zoología, Ornitología, Entomología, Etología
20595	JACQUELINE	SARMIENTO	HUMANIDADES	Otras Historia y Arqueología
15999	MARÍA CRISTINA	PEREA	CIENCIAS NATURALES Y EXACTAS	Ecología



producto_id	título	
438	2671	Caracterización geoambiental y cultural del Período Formativo en Selvas Occidentales Meridionales: sitio ?Horco Molle? (Dpto. de Yerba Buena, Tucumán).
1213	112784	ASOCIACIONES DE FORAMINÍFEROS BENTÓNICOS EN SECCIONES HOLOCENAS DEL ESTUARIO DE BAHÍA BLANCA (BUENOS AIRES, ARGENTINA): CARACTERIZACIÓN PALEOAMBIENTAL
4975	116590	Formative Pre-Hispanic Agricultural Soils in Northwest Argentina
5906	117554	Los cráneos trofeo (Condorhuasi-Alamito) dentro del pensamiento religioso andino
8155	119837	Análisis de morfometría geométrica en puntas Cola de Pescado del Uruguay
13985	125738	Degradación de suelos e incremento de erosión por acciones antropogénicas y cambios climáticos. Tucumán, Argentina.
25968	142466	Diversidad artefactual y explotación de materias primas en la costa norpatagónica. Golfo San Matías, Río Negro
23	1187	Agricultura prehispánica en Yasyamayo (valle de Santa María ? Tucumán ? Argentina)
5608	92368	Tiempo y espacio: el sitio 'Talapazo' (Valle de Yocavil, Provincia de Tucumán)
7946	94727	Exploring morphometric variations in Fishtail projectile points from Uruguay, Pampa and Patagonia
568	2801	Micropaleontological record of Holocene estuarine stages in the Bahía Blanca estuary, Argentina
2900	64517	Geoarchaeological review of the agricultural Pre-Hispanic period knowledge at the lowlands of Northwest Argentina
478	1214	Agricultural suitability and fertility in occidental piedmont of Calchaquies Summits (Tucuman, Argentina)
7862	44931	Reconstruction of agrarian practice and land impact in the drylands: A geoarchaeological approach
9741	46835	Arqueoestratigrafía de Referencia para el Registro de la Ocupación Humana Durante los Siglos XVI y XVII en la Localidad Esteco I, Departamento de Anta, Salta, Argentina
9765	46869	Distribución espacial de los asentamientos de los periodos de Desarrollos Regionales e Inca en el tercio central de la sierra de Quilmes (valle de Yocavil-Tucumán)



Productos publicados por al menos dos personas de la comunidad 844 (recorte)



Ubicación de la comunidad 844 en el grafo, en la zona entre Humanidades y Ciencias de la Tierra.

Buscamos a qué comunidad pertenece algún investigador que conozcamos

- Comunidad 149 : Enzo (aparece junto a gente que trabaja con neurociencias y relacionadas)

VIKTORIYA	SEMESHENKO	CIENCIAS SOCIALES	Economía, Econometría
VALERIA	DELLA MAGGIORE	CIENCIAS MEDICAS Y DE LA SALUD	Neurociencias (incluye Psicofisiología)
ARIEL	HAIMOVICI	CIENCIAS NATURALES Y EXACTAS	Otras Ciencias Físicas
CECILIA	SAMAME	CIENCIAS SOCIALES	Otras Psicología
NICOLAS	VON ELLENRIEDER	INGENIERIAS Y TECNOLOGIAS	Otras Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de...
DANIELA	SZENKMAN	HUMANIDADES	Lingüística
PABLO	BARTTFELD	CIENCIAS NATURALES Y EXACTAS	Otras Ciencias Biológicas
DIEGO JAVIER	MARTINO	CIENCIAS MEDICAS Y DE LA SALUD	Psiquiatría
CARLOS	GELORMINI	CIENCIAS SOCIALES	Otras Psicología
ESTEBAN FABIAN	VAUCHERET PAZ	CIENCIAS MEDICAS Y DE LA SALUD	Pediatría
FABRICIO HUGO	BAGLIVO	SIN DATOS	SIN DATOS
MARCOS LUIS	PIETTO	CIENCIAS SOCIALES	Otras Psicología
AGUSTIN MARIANO	IBÁÑEZ	CIENCIAS MEDICAS Y DE LA SALUD	Neurociencias (incluye Psicofisiología)
PAULA CELESTE	SALAMONE	CIENCIAS MEDICAS Y DE LA SALUD	Neurociencias (incluye Psicofisiología)
HECTOR	BERRA	CIENCIAS MEDICAS Y DE LA SALUD	Otras Ciencias de la Salud
JESICA CELESTE	FERRARI	SIN DATOS	SIN DATOS
MARIA	ROCA	CIENCIAS SOCIALES	Psicología especial (incluye terapia para el aprendizaje, habla, a...
ENZO RODOLFO	TAGLIAZUCCHI	CIENCIAS NATURALES Y EXACTAS	Otras Ciencias Físicas

Buscamos a qué comunidad pertenece algún investigador que conozcamos

- Comunidad 149 : Enzo (aparece junto a gente que trabaja con neurociencias y relacionadas)

titulo
Disrupted functional connectivity of the pain network in Fibromyalgia
Decision-Making in Frontotemporal Dementia: Clinical, Theoretical and Legal Implications
The cortical processing of facial emotional expression is associated with social cognition skills and executive functioning: A preliminary study
Functional imaging reveals movement preparatory activity in the vegetative state
Facial and semantic emotional interference: A pilot study on the behavioral and cortical responses to the dual valence association task
Cortical deficits of emotional face processing in adults with ADHD: its relation to social cognition and executive function
Sea slugs, subliminal pictures and vegetative state patients: Boundaries of consciousness in classical conditioning
Spontaneous BOLD event triggered averages for estimating functional connectivity at resting state
The human Turing machine: a neural framework for mental programs
Parsing a perceptual decision into a sequence of moments of thought
Impact of head models in N170 component source imaging: results in control subjects and ADHD patients
A big-world network in ASD: Dynamical connectivity analysis reflects a deficit in long-range connections and an excess of short-range connections
Effects of practice on task architecture: Combined evidence from interference experiments and random walk models of decision-making
Subliminal presentation of other faces (but not own face) primes behavioral and evoked cortical processing of empathy for pain
Fluid Intelligence and Psychosocial Outcome: From Logical Problem Solving to Social Adaptation
Performance errors of ingroup/outgroup stimuli and valence association in the implicit association task: brain bias of ingroup favoritism
The Environmental Pollutant Endosulfan Disrupts Cerebral Cortical Functions at Low Doses

¿De qué hablan las disciplinas?

- Detección de tópicos a través de los resúmenes
- Problemas: errores de transcripción, varios idiomas, un mismo resumen con palabras en español, datos con unidades
- Soluciones: detector de idiomas, patrones en datos mal cargados, stopwords más restrictivas, lemmatization

Detección de tópicos

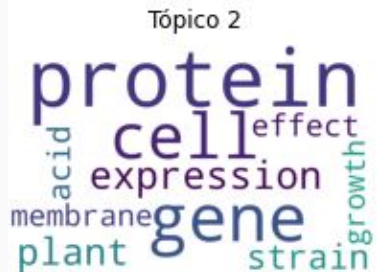
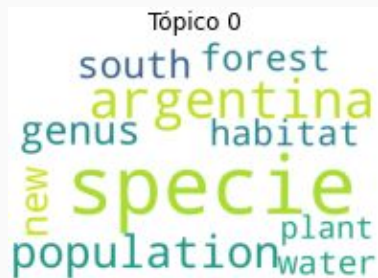
NMF

```
sklearn.feature_extraction.text.CountVectorizer
```

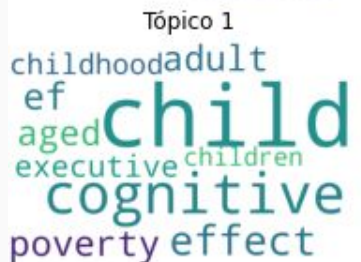
```
sklearn.feature_extraction.text.TfidfTransformer
```

```
sklearn.decomposition.NMF
```

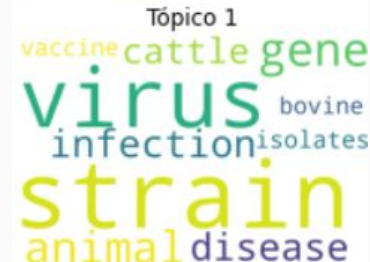
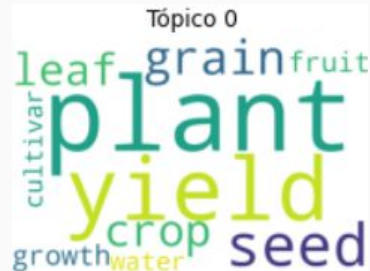

Ciencias Naturales y exactas



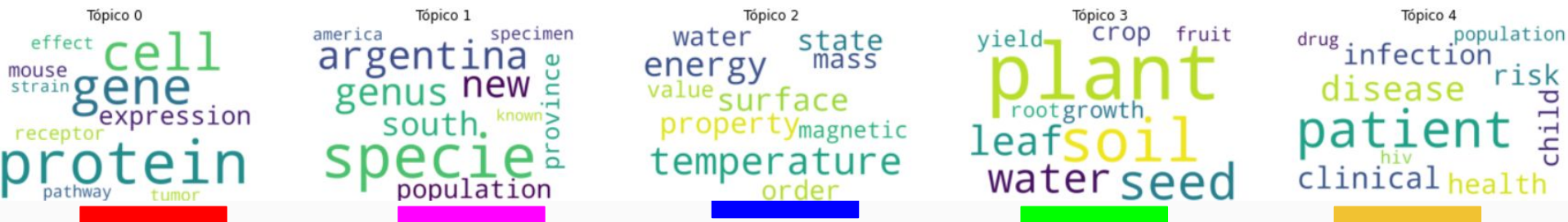
Ciencias Sociales



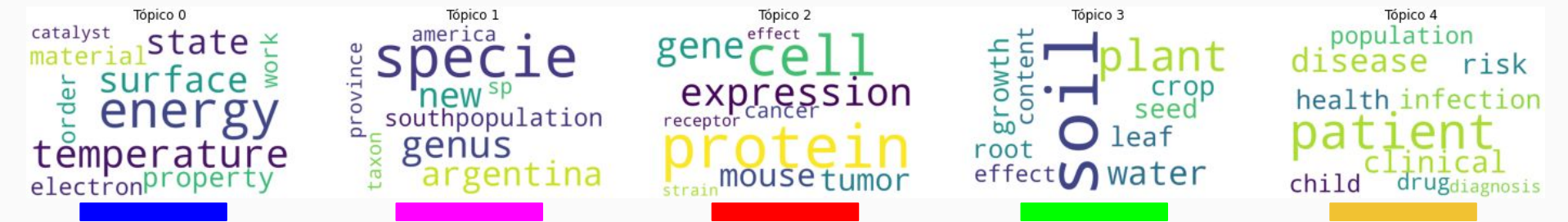
Ciencias Agrícolas



2011



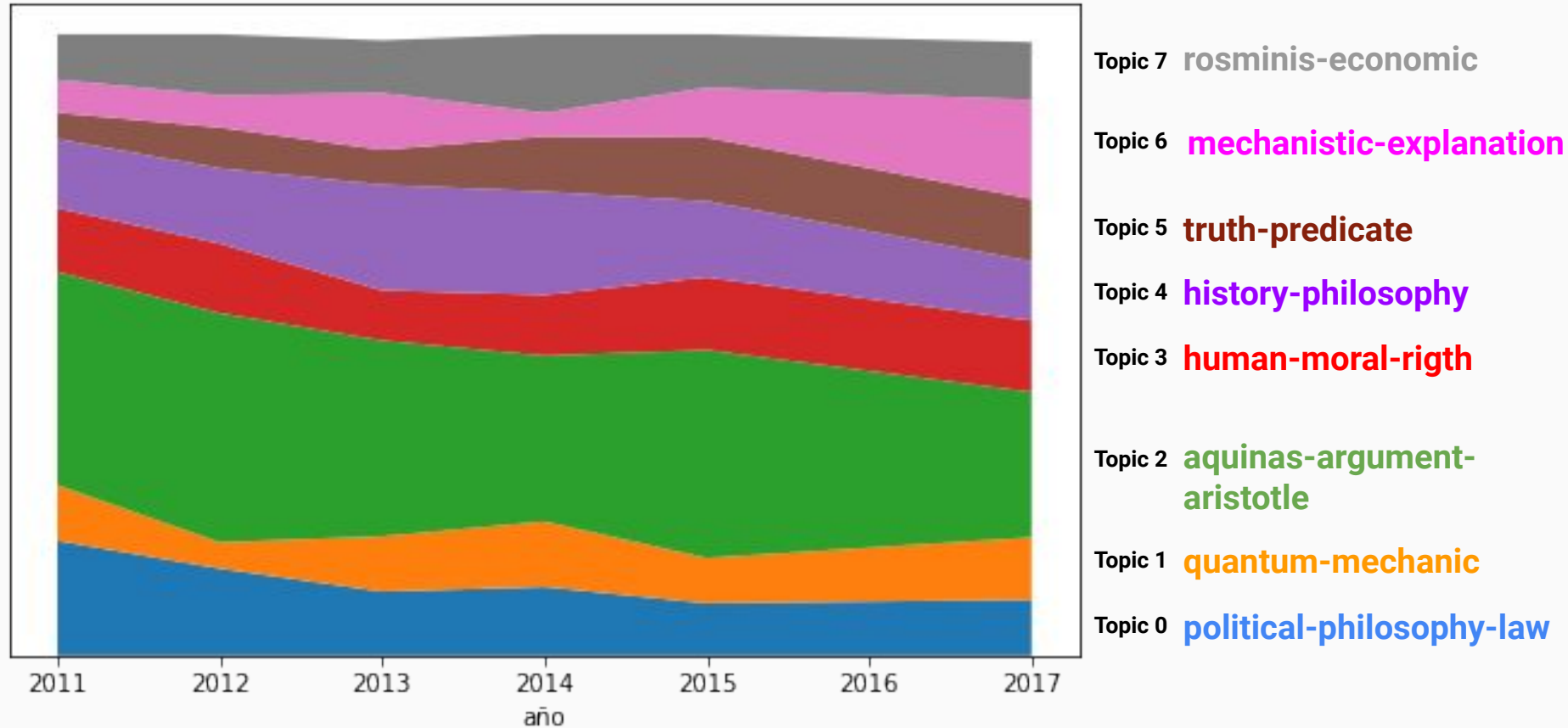
2014



2017



Filosofía, Ética y Religión



Ciencias de la Tierra y Medio Ambiente

Tópico 0

cretaceous fauna assemblage
specie
nov
sp south formation
specimen miocene taxons
sp nov america late genus
fossil south america phylogenetic

Tópico 1

water quality sediment
lake coastal tidal
groundwater stream basin
wetland discharge
aquifer surface water
ecosystem environment
river surface hydrological
environmental

Tópico 2

subduction fault arc andean
crustal deposit
rock age
tectonic belt zircon late magma
zone volcanic sierra
basin formation deformation

Tópico 3

southern value event
summer south solar
mean climate wind
ice south america
rainfall precipitation
wave temperature
anomaly glacier
america observed extreme

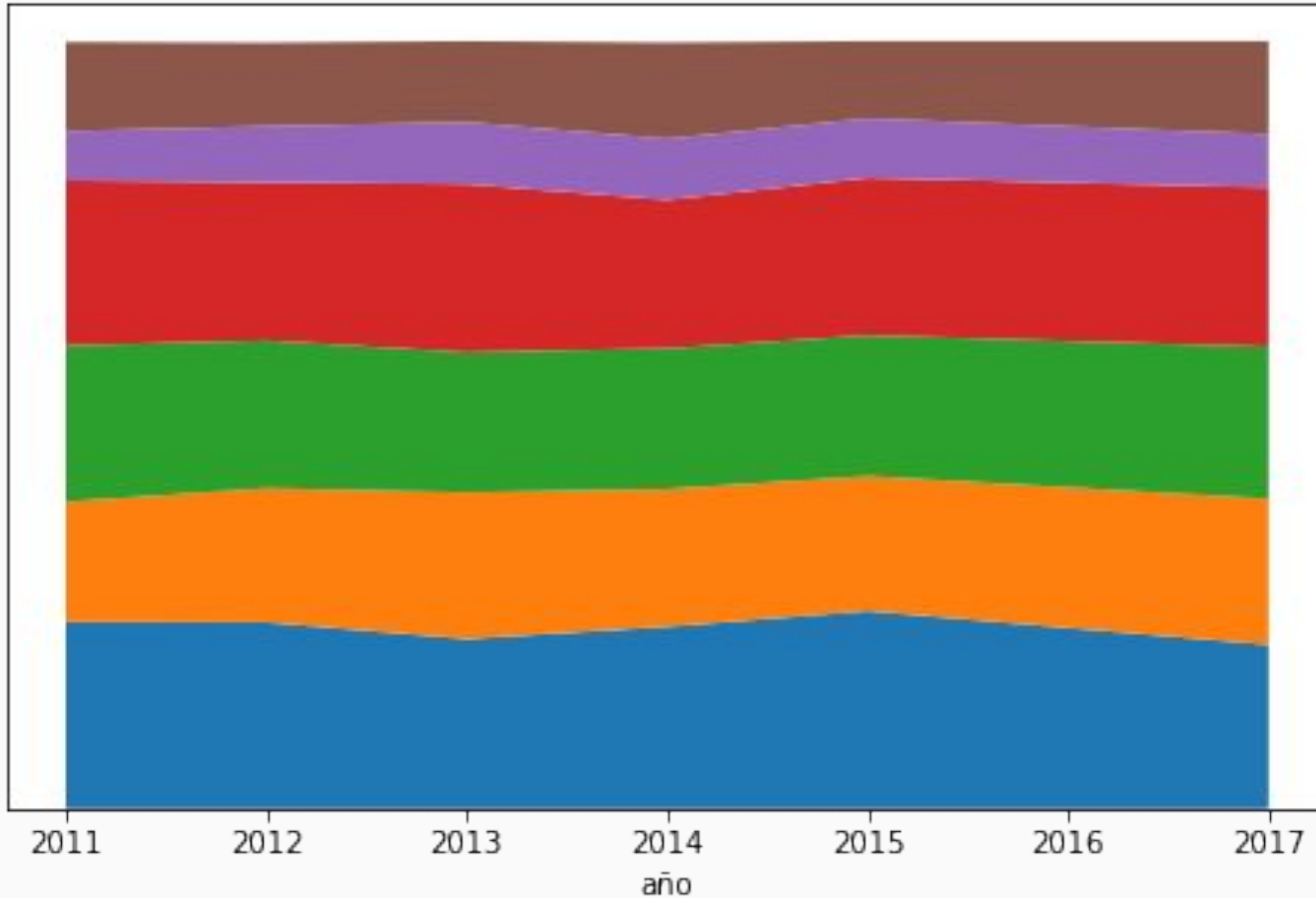
Tópico 4

land glyphosate
agricultural organic
moisture forest
carbon vegetation
crop plant effect
soybean retrieval horizon
soil moisture content
property adsorption

Tópico 5

heavy metal pollution
value cu
pb zn cr fe
plant cd heavy
metal hg effect
sediment mn
mg pb zn ph

Ciencias de la Tierra y Medio Ambiente



Topic 5: **metal-sediment**

Topic 4: **soil-moisture**

Topic 3: **precipitation**

Topic 2: **rock - fault - basin**

Topic 1: **lake - water**

Topic 0: **specie-new-fossil**

- Si bien nos pareció un buen ejercicio observar los diferentes tópicos obtenidos modificando la cantidad de tópicos por parámetros.
- También creemos que esto puede ser algo artesanal y necesitaríamos un conocimiento profundo en el área en el cual se intenta clusterizar.
- Buscamos una métrica que nos ayude a definir la cantidad de tópicos y luego realizar una búsqueda local.

- Por un lado, la librería Sklearn nos permite usar el algoritmo de NMF
- Por otro, no nos brinda (o no encontramos) una métrica que nos ayude a establecer una cantidad de tópicos aproximada para empezar a desarrollar nuestro modelo.
- **GENSIM**, también nos da la opción de usar NMF, y nos brinda un pipeline que nos permite tener una métrica para la cantidad de tópicos.

`models.coherencemodel` – Topic coherence pipeline

Calculate topic coherence for topic models. This is the implementation of the four stage topic coherence pipeline from the paper [Michael Roeder, Andreas Both and Alexander Hinneburg: "Exploring the space of topic coherence measures"](#). Typically, `CoherenceModel` used for evaluation of topic models.

Volviendo al área de “Filosofía, Ética y Religión”

- Se realizaron varias corridas del algoritmo variando la cantidad de tópicos.
- Se consideraron valores desde 4 a 20 tópicos
- Al calcular la coherencia de cada modelo. El resultado óptimo nos dio **4, con una coherencia de 0,8145.**
- Sin embargo los valores no oscilaron lo suficiente.
- Lo mismo realizamos para el área de “**Ciencias de la tierra y medio ambiente**”.
- En este caso se varió entre 4 y 30 tópicos, y se trabajaron con 5800 papers, también nos da un óptimo de **4 tópicos**, con un score de **0,7369.**

¡GRACIAS POR SU ATENCIÓN!

Interdisciplina, más ejemplos

- Comunidad 789 : Economía + Matemática**

	nombre	apellido	gran_area_descripcion	disciplina_descripcion
3380	MARÍA DE LOS ANGELES	CHARA	CIENCIAS NATURALES Y EXACTAS	Matemática Pura
3513	MARCELA	SVARC	CIENCIAS NATURALES Y EXACTAS	Estadística y Probabilidad
3630	RICARDO	FRAIMAN	CIENCIAS NATURALES Y EXACTAS	Estadística y Probabilidad
5941	JIMENA	VICENTIN MASARO	CIENCIAS SOCIALES	Economía, Econometría
6220	GUSTAVO EDUARDO	ROSSINI	CIENCIAS SOCIALES	Economía, Econometría

producto_id	titulo	
135	2151	Características de los Asociados a Cooperativas Lecheras Argentinas y Prioridades Gubernamentales en el Pago de Aportes No Reintegrables
267	2311	Decisiones de Consumo en Carnes: Una Aplicación de un Modelo Probit Multivariado
2000	113582	On local times, density estimation and supervised classification from functional data
3804	115395	Un Modelo de Elección Multinomial de Consumo de Alimentos Fuera del Hogar con Datos de Encuesta de Hogares
5214	116839	LDR: a package for likelihood-based sufficient dimension reduction
25514	142011	Density Estimation for Spatial-Temporal Data
36453	215575	Matemática Discreta
36494	215616	Competitividad del Complejo Lácteo Santafesino

Interdisciplina, más ejemplos

- Comunidad 789 : Economía + Matemática**

	nombre	apellido	gran_area_descripcion	disciplina_descripcion
3380	MARÍA DE LOS ANGELES	CHARA	CIENCIAS NATURALES Y EXACTAS	Matemática Pura
3513	MARCELA	SVARC	CIENCIAS NATURALES Y EXACTAS	Estadística y Probabilidad
3630	RICARDO	FRAIMAN	CIENCIAS NATURALES Y EXACTAS	Estadística y Probabilidad
5941	JIMENA	VICENTIN MASARO	CIENCIAS SOCIALES	Economía, Econometría
6220	GUSTAVO EDUARDO	ROSSINI	CIENCIAS SOCIALES	Economía, Econometría

producto_id	titulo	
135	2151	Características de los Asociados a Cooperativas Lecheras Argentinas y Prioridades Gubernamentales en el Pago de Aportes No Reintegrables
267	2311	Decisiones de Consumo en Carnes: Una Aplicación de un Modelo Probit Multivariado
2000	113582	On local times, density estimation and supervised classification from functional data
3804	115395	Un Modelo de Elección Multinomial de Consumo de Alimentos Fuera del Hogar con Datos de Encuesta de Hogares
5214	116839	LDR: a package for likelihood-based sufficient dimension reduction
25514	142011	Density Estimation for Spatial-Temporal Data
36453	215575	Matemática Discreta
36494	215616	Competitividad del Complejo Lácteo Santafesino