



Maestría en Explotación de Datos
y Descubrimiento del Conocimiento
Universidad de Buenos Aires

Facultad de Ciencias Exactas y Naturales
Facultad de Ingeniería



*Diseño de una herramienta para minería de redes de coautoría.
Estudio del caso de la Revista Argentina de Antropología Biológica*

Tesista: Ing. Julio Augusto Spairani

Directora de Tesis: Dra. María Gabriela Russo

Co-directora: Dra. María Elena Buemi

Buenos Aires, 21 de Diciembre de 2021

Introducción

El Problema

- Circuitos científicos en Latinoamérica
- Mainstream en inglés
- Métricas sesgadas por mayor muestra

El Enfoque: Redes de Coautoría

- Colaboración entre autores.

Herramientas Existentes

- Mayoritariamente basadas en WoS y Scopus

Antecedentes: Casos de Aplicación Existentes

Caso de Aplicación

Revista Argentina de Antropología Biológica (RAAB)

- Autores de diversas nacionalidades
- Artículos en Español Inglés y portugués
- Acceso Abierto (disponible en internet)
- En 2016:
 - Cumplio 20 años desde su primera edición en 1996
 - En circuitos científicos abiertos y latinoamericanos
 - No en mainstream internacional
- Desde junio 2021, es accesible desde Scopus.

Objetivos

General:

Realizar un aporte al estudio bibliométrico de revistas científicas a través del desarrollo e implementación de una herramienta para análisis de redes de coautoría.

Específicos:

1. Desarrollar una herramienta para el análisis de redes de coautoría
2. Analizar redes de coautorías de la RAAB

Hipótesis

1. La cantidad de autores y artículos aumenta en el tiempo.
2. Los autores fundadores se refuerzan en el tiempo, lo cual se relaciona con un aumento en alguna de las métricas (Grado, Fuerza de colaboración o Intermediación)
3. Los componentes más grandes absorben a los más pequeños a lo largo del tiempo, dado que autores periféricos pasan a trabajar en conjunto con autores principales.
4. La red agregada total (desde 1996 hasta 2016) presenta características comunes a la mayoría de las redes de coautorías científicas como Mundo Pequeño y Libre Escala.

Material es y métodos

Herramientas (R + librerías , Open Refine)

Objetivo 1

- Obtención de los datos
 - Web Scraping
 - Generación de una DB estructurada, adaptación al caso.
- Creación de la APP
 - Sección general
 - Sección análisis períodos
 - Sección temporal

Objetivo 2: Análisis utilizando del caso específico

Resultados

Objetivo 1: Aplicación web

App en vivo: <https://jaspa.shinyapps.io/raab-coaut-tesis/>

Fuente: https://github.com/jas1/raab_coaut_tesis/tree/freeze_entrega_tesis

Objetivo 2: Análisis aplicado con uso de la app

Resultados - Hipótesis 1

La cantidad de autores y artículos aumenta en el tiempo



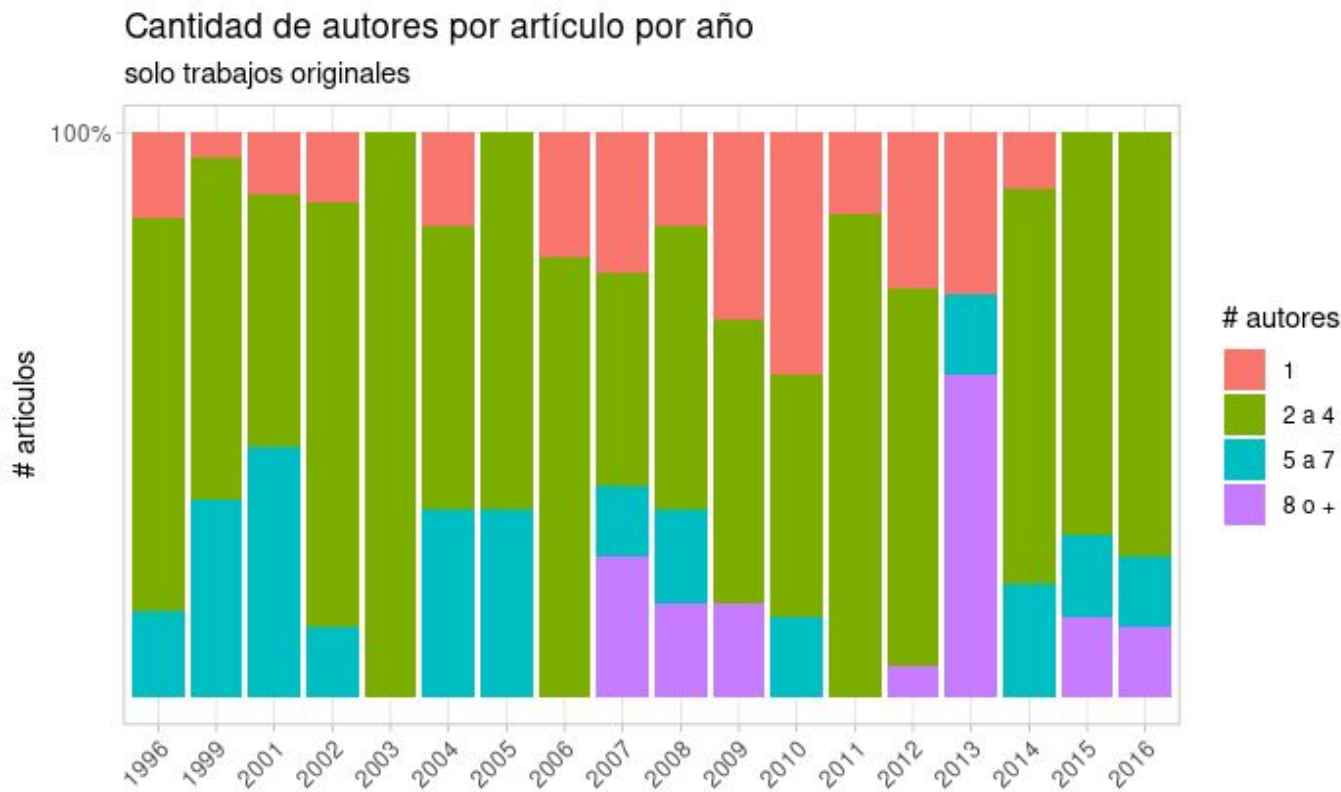
Resultados - Hipótesis 1

El resultado de la prueba no paramétrica de Mann-Kendall mostró que:

- No existe evidencia de una tendencia significativa para la cantidad total de autores.
($K=-0.013$, $p=0.93$)
- Tampoco para para la cantidad de publicaciones originales ($K=-0.013$, $p=0.69$).

Los resultados no son concordantes con la hipótesis 1.

Resultados - Hipótesis 1 - Análisis complementario



Resultados - Hipótesis 2

Los autores fundadores se refuerzan en el tiempo, lo cual se relaciona con un aumento en alguna de las métricas (Grado, Fuerza de colaboración o Intermediación)

Prueba de Rangos con Signo de Wilcoxon para muestras pareadas comparando el año 1996 con el 2016

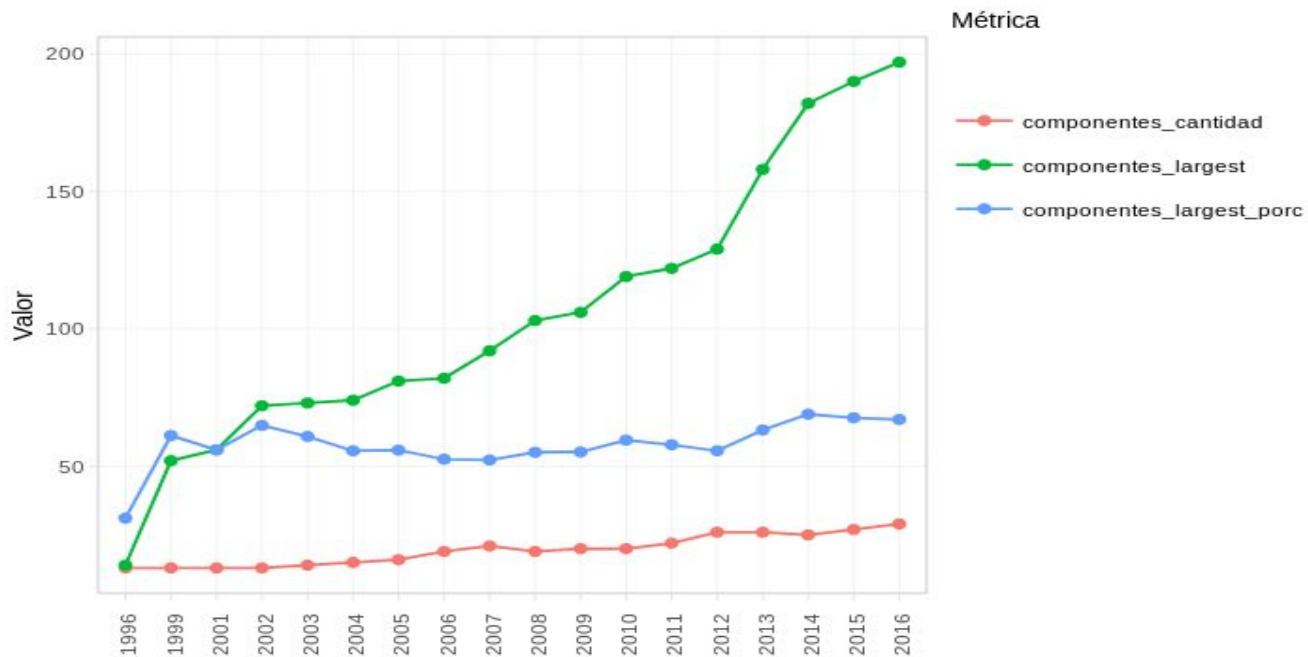
Grado: se observó un aumento significativo ($V=0$, $p < 0.001$, $n = 45$)

Fuerza de colaboración: la fuerza de colaboración aumentó ($V=0$, $p < 0.001$, $n = 45$)

Intermediación: se observó que existe evidencia que aumentó ($V=0$, $p < 0.001$, $n = 45$)

Resultados - Hipótesis 3

Los componentes más grandes absorben a los más pequeños a lo largo del tiempo, dado que autores periféricos pasan a trabajar en conjunto con autores principales

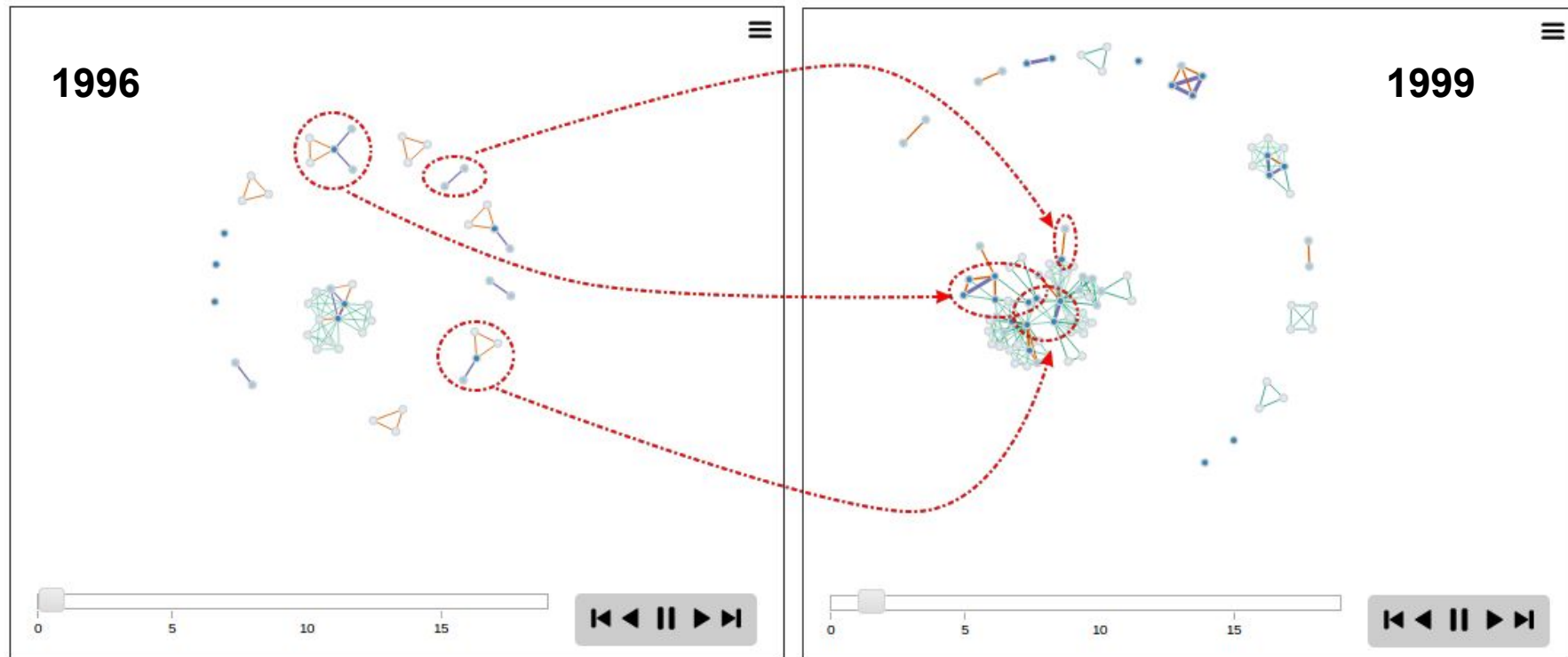


Resultados - Hipótesis 3

período (t0 a t1)		Cantidad total t1 (A)	Cantidad componente mayor t0 (B)	Cantidad componente mayor t1 (C)	Representa % Red (t1) (D=C/A)	Nuevos nodos (E)	Nodos absorbidos (F)	Total nuevos nodos en t1 (G = C - B) (G = E+G)
1996 1999	a	85	14	52	61.18	27	11	38
2012 2013	a	250	129	158	63.2	17	12	29
2013 2014	a	264	158	182	68.9	13	11	24

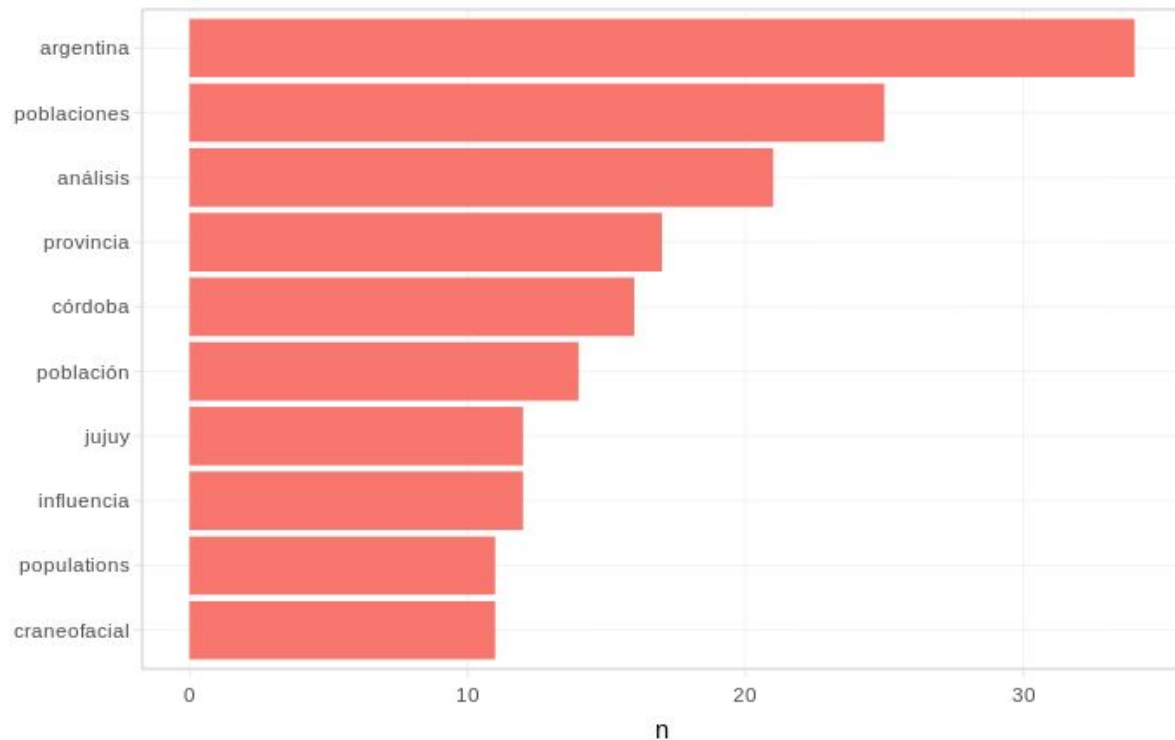
Se observa que los resultados son concordantes con la hipótesis de que los componentes periféricos pasan a ser parte del componente mayor a lo largo del tiempo dado que se generan nuevos trabajos entre los autores de los componentes.

Resultados - Hipótesis 3 - Visualización



Resultados - Hipótesis 3 - Frecuencia palabras

Frecuencia de palabras período 1996 - 2016 para el componente mayor



Resultados - Hipótesis 3 - Análisis complementario

Búsqueda de comunidades

Comunidad	Cantidad autores	Cantidad artículos	Primeras 10 palabras por frecuencia en el grupo
10	41	191	craneofacial (8), crecimiento (8), influencia (7), poblaciones (5), desarrollo (4), prenatal (4), argentina (3), contexto (3), estudio (3), growth (3)
37	37	183	jujuy (6), provincia (5), poblaciones (4), regional (4), argentina (3), distribución (3), peso (3), san (3), adultos (2), adults (2)
23	21	189	influencia (7), argentina (6), crecimiento (6), nutricional (5), población (5), socio (5), infante (4), aires (3), ambientes (3), craneofacial (3)
33	26	63	craneofacial (8), crecimiento (8), influencia (7), poblaciones (5), desarrollo (4), prenatal (4), argentina (3), contexto (3), estudio (3), growth (3)
18	21	45	fuego (7), tierra (7), grande (6), fagnano (4), regional (4), río (4), argentina (3), estables (3), isótopos (3), resultados (3)

Resultados - Hipótesis 4

La red agregada total (desde 1996 hasta 2016) presenta características comunes a la mayoría de las redes de coautorías científicas como Mundo Pequeño y Libre Escala

propiedad	total	1996-2001	2002-2006	2007-2011	2012-2016
Mundo pequeño criterio 1	SI	SI	NO	SI	SI
Mundo pequeño criterio 2: laxo	NO	NO	SI	NO	SI
Mundo pequeño criterio 2: estricto	NO	NO	NO	NO	NO
Libre escala	SI	SI	SI	NO	SI

Resultados - Hipótesis 4 - Análisis complementario

Comparación con otras redes de coautoría

métricas	Biomedical (Newman, 2001)	Computer science (Newman, 2001)	Management and organization (Acedo et al.,2006)	Tourism and hospitality (Yeet al., 2013)	Strategic Management (Koseoglu 2016)	RAAB (este estudio)
artículos por autor	6.4	2.6	2.04	1.1	0.88	2.14
autores por artículo	3.75	2.22	1.88	1.87	1.13	3.68
coeficiente de clustering (transitividad)	0.066	0.496	0.68	0.748	0.13	0.543
componente principal representado en porcentaje	92.6	57.2	45.4	59.3	69	67.01
distancia media	4.6	9.7	-	7.2	5.05	3.41
tamaño del componente principal	1,395,693	6,396	4,625	1,376	296	197

Discusión y Conclusiones

- Base de datos
- Aplicación web para análisis de redes de coautoría
- Código abierto
- Proceso de armado de la Base de datos
- Proceso de armado de la aplicación
- Análisis aplicado de caso de Estudio de la RAAB utilizando la Aplicación

Discusión y Conclusiones - Trabajos futuros

- Actualización de la base de datos
- Aspectos técnicos
- Análisis de textos y tópicos en el tiempo
- Análisis de conectividad / puentes
- Comparación con otras revistas
- Otros Análisis
 - Mejor comprensión y expansión de componentes y comunidades (filiación, geolocalización, etc)
 - Agregar otras métricas para el análisis

Discusión y Conclusiones - Consideraciones finales

- **Aporte al estudio bibliométrico** de una revista que, al momento de realización del trabajo, **no** se encontraba indexada en las bases de datos pertenecientes al denominado ***mainstream*** internacional.
- Las metodologías descritas y los análisis realizados podrían resultar de utilidad para otras revistas de alcance local y regional, tanto dentro de la disciplina estudiada como en otras.
- Multidisciplinar, con sus desafíos y oportunidades.
- Múltiples posibilidades que se abren a futuro.

Gracias !

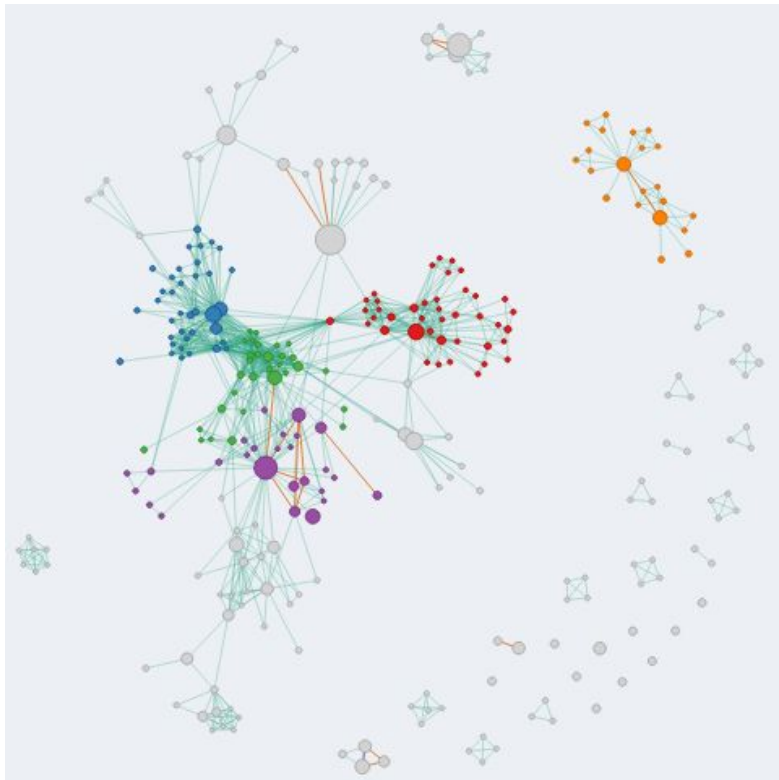
Sean bienvenidos a colaborar con el proyecto

https://github.com/jas1/raab_coaut_tesis

Detalle Modelos - Mundo Pequeño

Cálculo	Método validación 1	Método validación 2
Transitivity (red) (C)	0.5429	0.5429
Promedio Transitivity (simulaciones) (Cr)	0.0250	0.1126
Promedio camino más corto (red) (L)	3.4103	161.3158
Promedio camino más corto (simulaciones) (Lr)	3.0614	22.6893
Gamma = C / Cr	21.716	4.8214
Delta = L / Lr	1.1139	7.1097
Smallwordlness = Gamma / Delta	19.4942	0.6781
Criterio 1 si Delta \sim 1 y Gamma > 1	SI	NO
Criterio 2 Smallwordlness > 1 (laxo)	SI	NO
Criterio 2 Smallwordlness > 3 (estricto)	SI	NO

Detalles Comunidades



Algoritmo Utilizado	Cantidad de comunidades	Valor modularidad	Valor modularidad con peso (dato por fuerza de colaboración)
Optimización de modularidad multi-nivel (Multi-level Optimization of Modularity) (Louvain)	37	0.692	0.7075
Intermediación de Aristas (Edge Betweenness)	43	0.6744	0.6596
Busqueda de comunidades Infomap (Infomap community finding)	46	0.6732	0.6935
Caminatas aleatorias cortas (Short Random Walks)	47	0.6725	0.6822
Autovector lider (Leading Eigenvector)	35	0.6551	0.6385
Etiquetas Propagates (Propagating Labels)	44	0.6459	0.642

Introducción - Bibliografía

El problema

Beigel F. (2014) *Publishing from the periphery*

Santa, S., & Herrero-Solana, V. (2010). *Coverage in scopus vs. web of science off research produced in Latin America and the Caribbean.*

Redes coautoría

Freeman, L.C. (2004) *The Development of Social Network Analysis: A Study in the Sociology of Science*

Barabási, A. L. (2016). *The power of network science, the beauty of network visualization*

Glänzel W., Schubert A. (2004) *Analysing Scientific Networks Through Co-Authorship*

Newman, M. E. J. (2004). "Coauthorship networks and patterns of scientific collaboration"

Herramientas Existentes

Aria, M. and Cuccurullo, C. (2017). *bibliometrix: An R-tool for comprehensive science mapping analysis*

Cobo, M., López-Herrera, A., Herrera-Viedma, E. and Herrera, F. (2012), *SciMAT: A new science mapping analysis software tool*

Van Eck, N.J., Waltman, L. *Software survey: VOSviewer, a computer program for bibliometric mapping.*

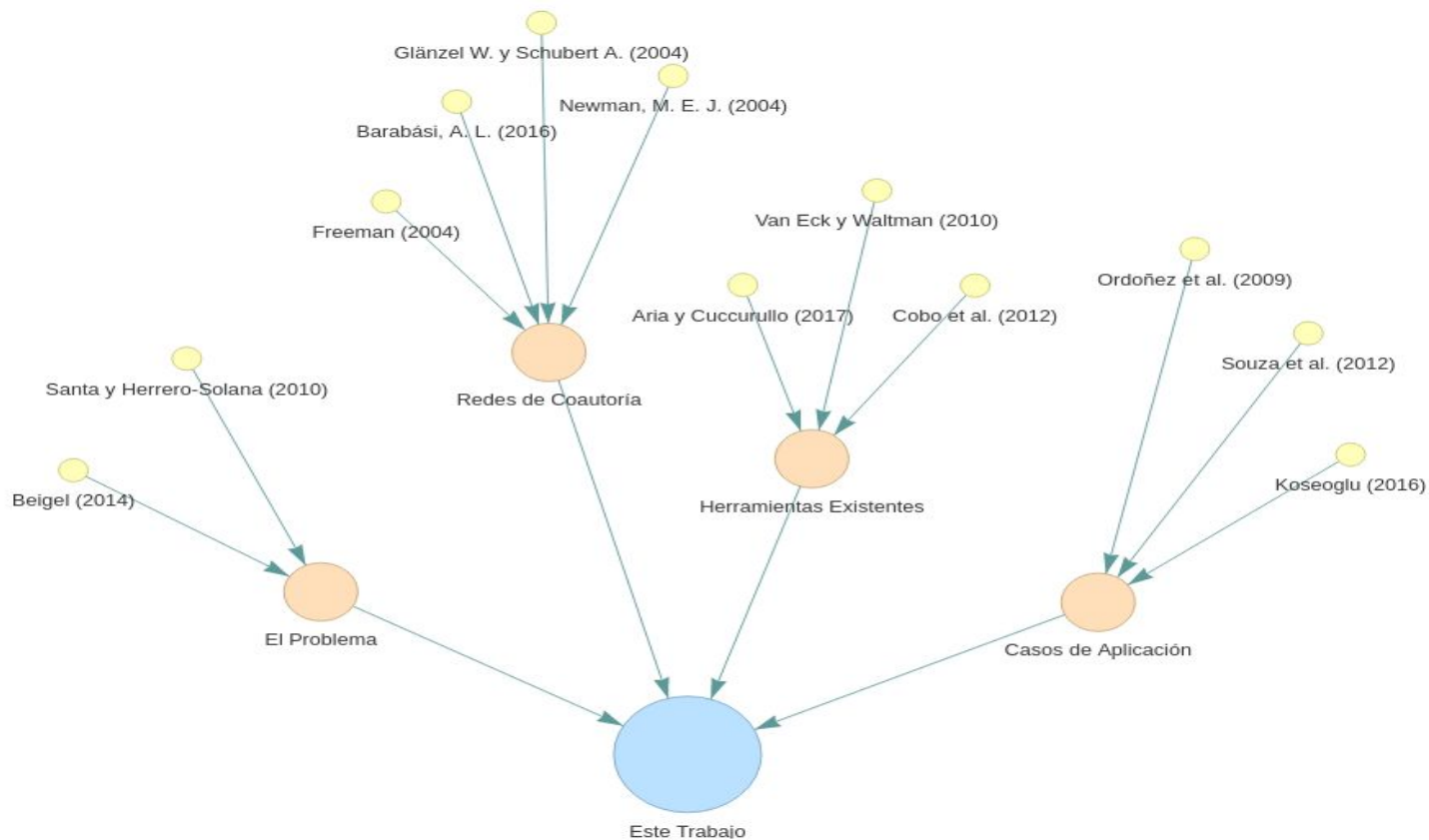
Casos de Aplicación

Ordoñez, G, A. Hernández Badillo, C. Hernández, et al. (2009). "Análisis bibliométrico de la Revista de Economía Institucional en sus primeros diez años"

Souza, C.G. d., Barbastefano, R. G. , Lima, L.S. d. (2012) *Redes de colaboração científica na área de química no Brasil: um estudo baseado nas coautorias dos artigos da revista Química Nova*

Koseoglu, M. A. (2016). "Growth and structure of authorship and co-authorship network in the strategic management realm: Evidence from the Strategic Management Journal"

Introducción - Bibliografía



Resultados - Hipótesis 2 - Análisis complementario

Ranking de 5 autores por año y su evolución temporal

