



Universidad de Buenos Aires

Facultad de Ciencias Exactas y Naturales

Maestría en Explotación de Datos y Descubrimiento de Conocimiento

Tesis de Maestría

*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, Diciembre de 2021

# Resumen

La presente tesis aborda el problema del análisis bibliométrico dentro de una revista científica local en un periodo de 20 años desde su primera edición, que al momento de escritura del trabajo no figuraba en las principales bases de datos (como Scopus o Web of Science). El propósito de este trabajo fue crear una aplicación de código abierto para explorar redes de coautoría, y analizar específicamente el caso de la "Revista Argentina de Antropología Biológica" (RAAB). En este estudio se utilizaron métodos como minado web para la creación de la base de datos, y la librería Shiny para la creación de la aplicación en lenguaje R. La aplicación desarrollada fue luego utilizada para explorar los datos y obtener resultados sobre la red de coautoría en el contexto de la RAAB. De esta forma, se constataron hipótesis acerca del crecimiento de la red en el tiempo, los roles de los autores fundadores, la absorción de componentes a lo largo del tiempo y la presencia de características comunes en redes de coautoría científica como Mundo Pequeño y Libre Escala. Se espera que este aporte pueda expandirse a otras revistas dentro de la disciplina, como a otras disciplinas.

Palabras clave: redes de coautoría; minado web; aplicación web ; análisis temporal; herramientas código abierto;

# Abstract

This thesis addresses the problem of bibliometric analysis within a local scientific journal over a period of 20 years since its first edition, which at the time of writing was not included in the mainstream databases (such as Scopus or Web of Science). The purpose of this work was to create an open source application to explore co-authorship networks, and to specifically analyze the case of the "Revista Argentina de Antropología Biológica" (RAAB). In this study we used a variety of methods such as web mining for database building, and Shiny for the development of the application in R language. The developed application was then used to explore the data and obtain results about the co-authorship network in the context of the RAAB. In this way, hypotheses about the growth of the network over time, the roles of the founding authors, the absorption of components over time and the presence of common features in scientific co-authorship networks such as Small World and Free Scale were ascertained. It is hoped that this contribution can be expanded to other journals within the discipline, as well as to other disciplines.

Keywords: coauthorship networks; web mining ; web application ; temporal analysis; open source tools;

# Agradecimientos

A Nancy, Máximo, Emma, familia, directores, maestros, amigos, compañeros y muchas otras personas que me dieron apoyo, soporte, ayuda y compañía en este largo trayecto.

# Índice

<b>Resumen</b>	<b>2</b>
<b>Abstract</b>	<b>2</b>
<b>Agradecimientos</b>	<b>3</b>
<b>Índice</b>	<b>4</b>
<b>Capítulo 1: Introducción</b>	<b>9</b>
Conceptos clave en el análisis de redes de coautoría	10
Redes y Grafos	10
Redes de Coautoría	10
Dirección de un grafo	12
Grado	12
Fuerza de colaboración	12
Intermediación	12
Red conexa	13
Componente	13
Comparación de modelos	14
Mundo Pequeño	14
Libre Escala	15
Comunidades	15
Características del caso de aplicación	16
Objetivos	18
Objetivo 1: Desarrollar una herramienta para el análisis de redes de coautoría	18
Objetivo 2: Analizar redes de coautorías de la RAAB	18
Hipótesis	19
<b>Capítulo 2: Materiales y Métodos</b>	<b>20</b>
Generación de la base de datos	20
Análisis de la estructura del sitio	20
Modelo de Bases de datos	20
Obtención de los datos: Web scraping	21
Resultados intermedios	21
Generación de la base relacional	21
Generación de los grafos	22
Desarrollo de la aplicación para análisis de redes de coautoría	22
Estructura de la interfaz gráfica	22
Sección 1: Consideraciones	23
Sección 2: Análisis Estático	23
Visualización de coautoría	23
Artículos asociados	23
Estructura de la red	23

Comparación con modelos	24
Validación de Mundo Pequeño	24
Validación de Libre Escala	24
Comunidades	25
Sección 3: Análisis temporal	25
Análisis de la red	26
Hipótesis 1: La cantidad de autores y artículos aumenta en el tiempo	26
Pruebas estadísticas empleadas	26
Hipótesis 2: Los autores fundadores se refuerzan en el tiempo, lo cual se relaciona con un aumento en alguna de las métricas (Grado, Intermediación o Fuerza de colaboración)	26
Pruebas estadísticas empleadas	27
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	27
Análisis de Componentes	27
Relación con actividades de investigación	28
Búsqueda de comunidades	28
Métrica de evaluación de algoritmos: Modularidad	29
Optimización de modularidad multi-nivel	29
Búsqueda de sentido en los grupos obtenidos	30
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	31
Mundo Pequeño (small world)	31
Libre Escala (free scale)	31
Comparación con otras redes de coautoría	31
Disponibilidad de la herramienta: local, en la red, en internet	32
Código Fuente	32
<b>Capítulo 3: Resultados</b>	<b>33</b>
Herramienta desarrollada para el análisis de redes de coautorías	33
Sección 1: Consideraciones	33
Datos Crudos	34
Composición	34
Resumen	34
Preguntas	34
Sección 2: Análisis Estático	34
Configuración	35
Visualización	35
Artículos asociados	35
Estructura de la red	35
Métricas de la red	36
Métricas de los nodos	36
Detalles de los componentes	36
Comparación con modelos	36
Validación de Mundo Pequeño	37
Validación de Libre Escala	37

Comunidades	37
Sección 3: Análisis temporal	38
Animación del grafo de manera acumulada	38
Medidas de cada periodo	38
Medidas Acumuladas a cada periodo	38
Ranking de autores en el tiempo	39
Ranking de autores en el tiempo para medidas acumuladas	40
Análisis de la red de coautorías de la RAAB	40
Hipótesis 1: La cantidad de autores y artículos aumenta en el tiempo	40
Cantidad de autores en artículos por año	44
Hipótesis 2: Los autores fundadores se refuerzan en el tiempo, lo cual se relaciona con un aumento en alguna de las métricas (Grado, Intermediación o Fuerza de colaboración).	45
Ranking de 5 autores por año y su evolución temporal	47
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	51
Verificación de absorción de componentes en la red	53
Relación con actividades de investigación	55
Búsqueda de comunidades	58
Búsqueda de sentido en los grupos obtenidos	60
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	62
Comparación con modelos	62
Comparación con otras redes de coautoría	64
<b>Capítulo 4: Discusión y Conclusiones</b>	<b>66</b>
Desarrollo de una herramienta para el análisis de redes de coautoría	66
Análisis del caso aplicado: redes de coautorías de la RAAB	67
Consideraciones metodológicas y perspectivas futuras	69
Perspectivas futuras	69
Actualización de la base de datos	69
Aspectos técnicos	69
Proyección sobre artículos	70
Análisis de textos y tópicos en el tiempo	70
Análisis de conectividad / puentes	70
Otros Análisis	71
Comparación con otras revistas	71
Consideraciones finales	71
<b>Bibliografía</b>	<b>73</b>
<b>Anexos</b>	<b>78</b>
Anexo: resumen proceso obtención y generación de la base de datos	78
Anexo: web scraping	79
Adquisición de los datos	80
Interpretación de los archivos	82
Interpretación de los archivos HTML	82

Interpretación de los archivos bibTex	82
Interpretación de los archivos PDF	82
Anexo: Base de datos: Estructura relacional	83
Anexo: Generación de grafos	85
Descripción de las propiedades de los nodos y aristas	85
Propiedades de los Nodos	86
Propiedades de las Aristas	86
Anexo: sección preguntas	88
Anexo: métricas de red disponibles	89
Anexo: métricas de nodos	90
Anexo: Propiedades modelos	91
Anexo: Mapa de la Interfaz de usuario	92
Anexo: Tablas numéricas complementarias para la Hipótesis 2	94
Anexo: Detalle periodos inferiores para Mundo pequeño y Libre Escala	98
Mundo pequeño para el periodo 1996 - 2001	98
Libre escala 1996 - 2001	99
Mundo pequeño para el periodo 2002 - 2006	101
Libre escala 2002 - 2006	103
Mundo pequeño para el periodo 2007 - 2011	104
Libre escala 2007 - 2011	105
Mundo pequeño para el periodo 2012 - 2016	107
Libre escala 2012 - 2016	108
Anexo: Cantidad de autores en artículos por año	110
Anexo: software y librerías utilizadas	111
Anexo: Detalle de algoritmos para búsqueda de comunidades	115
Anexo: Detalle procesamiento bibTEX	116
Problemas de transformación y encoding bibtex	116
Anexo: Detalle consideraciones PDF	120
Obtención de los pdf	120
Interpretación de los pdf	120
Anexo: Detalle estandarización autores	121
Revisión y validación con expertos	121
Iteraciones de clustering	122
Resultado de Revisión y validación con expertos	127

# Capítulo 1: Introducción

El área académica es una de las áreas de mayor desigualdad en cuanto al acceso a la información. El impacto de los estándares forjados en el tiempo afectó considerablemente el desarrollo de los circuitos de conocimiento de la periferia afectando la participación e interacciones entre quienes pertenecen a diferentes circuitos científicos. Los circuitos de publicación fueron diferenciados por un principio de jerarquía que se basa en tres factores, los recursos institucionales, disciplinas seleccionadas y su habilidad con el idioma Inglés. Existen por lo menos cuatro tipos de circuitos que permiten a los científicos acumular prestigio de diferentes maneras: el denominado “*mainstream* internacional” (*Thomson Reuters -Web of Science-, Elsevier -Scopus-, Google*), “entornos de acceso abierto” (*DOAJ, Dial-net, INASP*), “regionales de latinoamérica” (*LATINDEX, SCIELO, CLACSO, REDALYC, AJOL*) y “nacionales” (*publicaciones locales que no figuran en los otros circuitos*) ([Beigel 2014](#)).

[Santa y Herrero-Solana \(2010\)](#) realizaron una comparación de trabajos de América Latina y Caribe dentro de Scopus y Web Of Science, resaltando que estas bases son cuestionadas por varios autores en cuanto a si tienen validez para los trabajos realizados en dichas regiones. Sin embargo, la mayoría de los trabajos bibliométricos se realizan sobre estas bases.

Dentro de las técnicas de análisis bibliométrico se encuentran las redes de coautoría que nos acercan a un enfoque social de las interacciones entre los autores. Estas redes son un tipo de redes sociales, y teniendo en cuenta lo mencionado por [Freeman \(2004\)](#), en el enfoque moderno de análisis de redes sociales se utilizan cuatro factores definen el campo de análisis de redes sociales, que son:

1. El análisis de redes sociales es motivado por una intuición basada en las conexiones entre los actores sociales
2. Está basado en datos sistemáticos y empíricos
3. Se sustenta mucho en gráficas
4. Confía en el uso de matemática y modelos computacionales.

*“La Coautoría es una de las formas mejor documentadas y más tangibles de colaboración científica. Casi todos los aspectos de redes de colaboración científica pueden ser confiablemente seguidos a través de analizar redes de coautoría con métodos bibliométricos.”* ([Glänzel-Schubert 2005: 1](#)).

Las redes de coautoría resultan entonces una herramienta útil para analizar patrones de colaboración científica, tanto dentro de una disciplina, región geográfica o revista científica, como en forma comparativa ([Kumar 2015](#)).

Al analizar redes de coautoría en una disciplina particular, las características relacionadas al grado de cohesión o fragmentación, el tamaño de los nodos y componentes o el patrón de conexión entre nodos, entre otras, brindan información sobre el diálogo colaborativo entre los distintos grupos de investigación que la conforman y la “importancia” relativa de cada uno en el contexto particular. Siguiendo lo propuesto por [Morel et al. \(2009\)](#), es de particular interés analizar la red a fin de evaluar el incremento en el número de publicaciones, grupos de investigación y colaboraciones entre los grupos, así como el cambio en estas dinámicas a lo largo del tiempo.



Existen numerosos estudios de análisis bibliométricos aplicados a revistas científicas. Entre los más relevantes para esta tesis puede nombrarse el trabajo realizado por [Souza et al. \(2012\)](#) analizando una revista de latinoamérica en la que también publican autores internacionales. Otro estudio más abarcativo desde un punto de vista bibliométrico y también en el contexto latinoamericano es el análisis realizado por [Ordoñez et al. \(2009\)](#). Finalmente, [Koseoglu \(2016\)](#) revisa temas como la evolución temporal de una revista específica, evaluando métricas, visualizaciones y comparaciones.

Un proyecto más abarcativo es Bibliometrix, en lenguaje R y de código abierto, sin embargo este proyecto tiene como orígenes de datos sólo aquellos provenientes de Web Of Science y Scopus, o archivos en esos formatos, y la sección de análisis de coautoría es limitada. En el artículo de Bibliometrix, [Aria-Cuccurullo \(2017\)](#), además del propio trabajo, también es importante la mención de otras herramientas para realizar diferentes tareas de bibliometría, y también paquetes de R que incluyen tareas de bibliometría. Como complemento de estas herramientas se puede mencionar la revisión realizada por [Cobo et al. \(2011\)](#), entre las que se encuentran VosViewer de [Van Eck-Waltman \(2010\)](#) y SciMat de [Cobo et al. \(2012\)](#).

En esta tesis se desarrolló una herramienta para el análisis de redes de coautoría y se aplicó a la Revista Argentina de Antropología Biológica (RAAB), como ejemplo dentro de una disciplina particular y a nivel local (Argentina).

En este trabajo se utilizaron varios de los análisis mencionados en la revisión de [Kumar \(2015\)](#); como por ejemplo: la desambiguación de nombres de autor, el estudio de mundo pequeño, la búsqueda de comunidades, la visualización de la red, tamaño del componente más grande, entre otros. Algunos de estos se profundizan en la sección siguiente.

## Conceptos clave en el análisis de redes de coautoría

### Redes y Grafos

Una Red o Grafo es una colección de objetos, llamados **Vértices** y **Aristas**. Un par de vértices está conectado a través de una Arista.

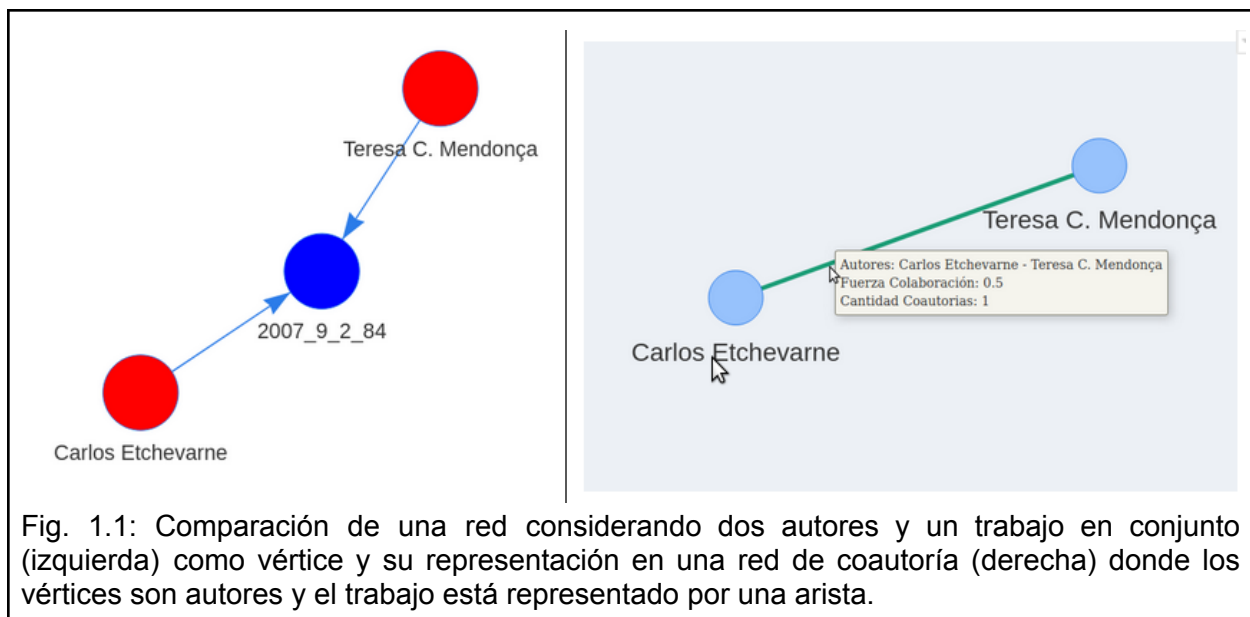
$$G = (V, A)$$

En esta tesis utilizamos Nodos o Vértices, y Aristas o Conexiones.

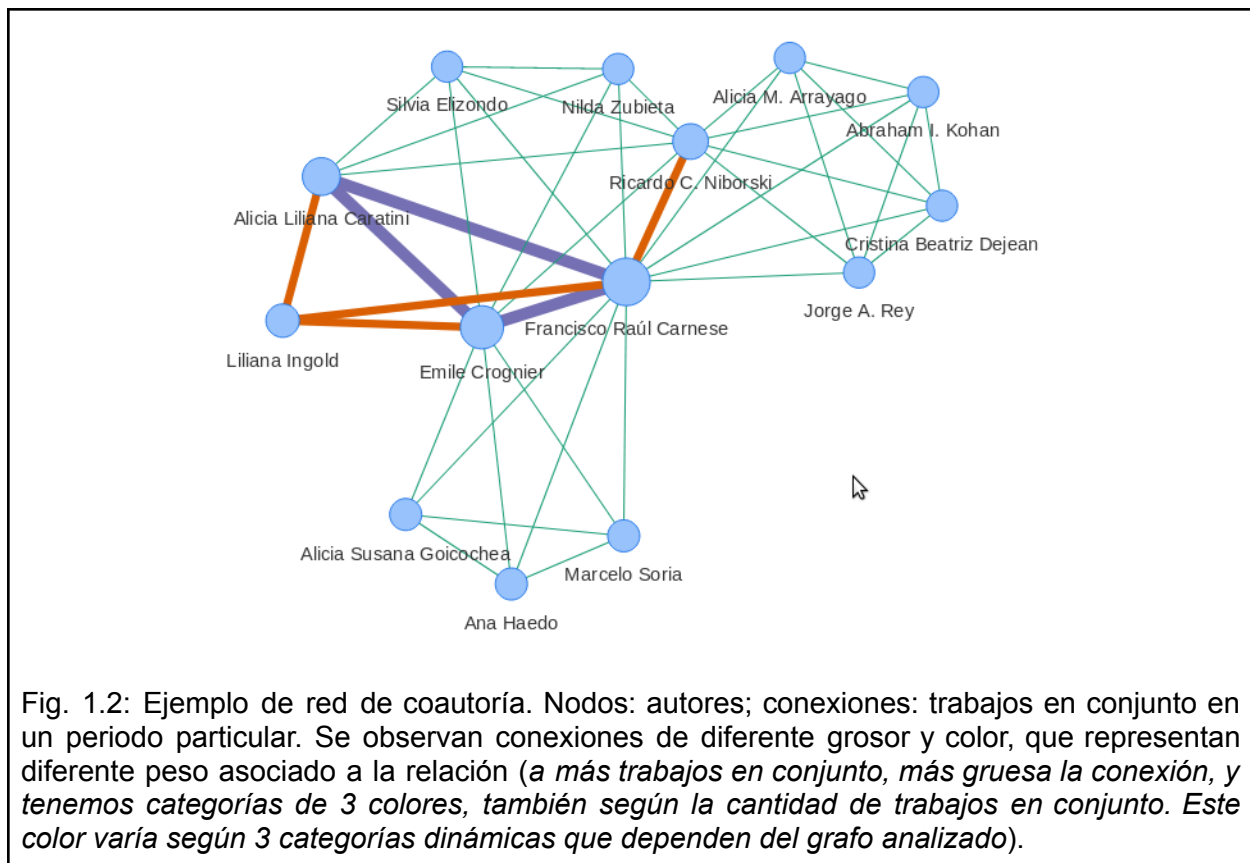
Respecto de la utilidad de las redes, se menciona: “*Si queremos entender sistemas complejos, primero necesitamos saber cómo sus componentes interactúan entre ellos. En otras palabras, necesitamos un mapa de cómo están conectados. Una red es un catálogo de los componentes de un sistema, comúnmente llamados nodos o vértices, y las interacciones directas entre ellos, llamadas conexiones o aristas. Este tipo de representación de redes ofrece un lenguaje común para estudiar sistemas que pueden diferir enormemente en naturaleza, apariencia o alcance.*” (Barabási 2016: Section 2.2 Networks and Graphs).

### Redes de Coautoría

En una red de coautoría los Vértices/Nodos representan los Autores de los artículos, y las Aristas/Conexiones representan la relación coautoría, lo que significa: *estos autores generaron por lo menos un artículo en conjunto*, como se observa en la [Figura 1.1](#)



Asimismo, la conexión entre nodos puede tener distintas características que pueden representarse gráficamente, en el ejemplo de la [Figura 1.2](#) dos autores que publicaron más de un artículo en conjunto se unen mediante una arista más gruesa o de diferente color.



## Dirección de un grafo

Una red puede ser Dirigida (*directed*) o No Dirigida (*undirected*). Cuando una red es Dirigida significa que las relaciones van **desde** un nodo origen **hasta** el nodo destino, marcando un sentido en la relación. Por ejemplo, en una red de nodos Autores y Documentos, los Autores “crean” Documentos y esa relación “crean” tiene dirección desde autores hacia documentos.

Por otro lado, las redes No Dirigidas son aquellas en las que no hay una dirección específica entre los nodos. Las redes de coautoría son de naturaleza no dirigida, puesto que representa que ambos autores trabajan en conjunto.

## Grado

El grado de un nodo es la cantidad de conexiones que posee. Para redes no dirigidas si lo vemos desde el punto de vista de la matriz de adyacencia  $A$ , el grado  $k$  del nodo  $i$ , es la suma de la fila  $i$  de la matriz.

$$k_i = \sum_j a_{ij}$$

En el caso de redes sin peso, el grado toma valores 1 o 0, mientras que cuando son con peso, le es asignado algún valor específico.

## Fuerza de colaboración

La fuerza de colaboración, traducida del inglés *collaboration strength*, es una métrica descrita por [Newman \(2004\)](#), que plasma cómo se reparte el esfuerzo de la generación de un artículo. Su fórmula es la siguiente:

$$FC = 1 / (n - 1)$$

siendo  $FC$ : fuerza de colaboración,  $1$  representa el artículo creado, y  $n-1$  representa que cada autor divide su tiempo entre los  $n-1$  otros autores con los que trabaja en el artículo.

En este trabajo las relaciones entre nodos presentan más de un atributo. Por un lado se toma el Grado que usa la definición sin peso, luego se utiliza la Fuerza de colaboración como una métrica de peso de la relación y, finalmente, se incluye el conteo de la cantidad de artículos realizados en conjunto como otro atributo de interés.

## Intermediación

La intermediación, del inglés *betweenness*, es una métrica descrita por [Freeman \(1977\)](#) y representa la importancia de un nodo según su participación en caminos para llegar de un nodo a otro. Intermediación  $I$ , del vértice  $v$  es la sumatoria del cociente entre  $\sigma_{st}(v)$ , la cantidad de caminos más cortos desde  $s$  hasta  $t$  pasando por  $v$ , y  $\sigma_{st}$ , la cantidad de caminos más cortos desde  $s$  hasta  $t$ , siendo  $s, t$  y  $v$  tres nodos diferentes:

$$I(v) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Vemos un ejemplo de esta métrica en la [Fig. 1.3](#).

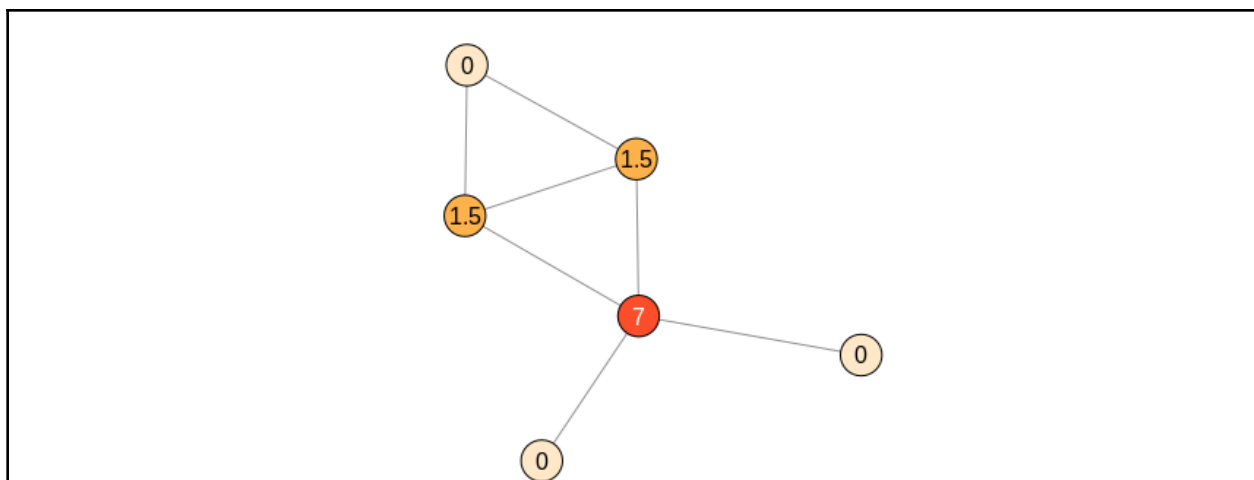


Fig. 1.3: Ejemplo de Intermediación. El nodo rojo y los nodos naranjas tienen más intermediación dado que para llegar a otros nodos es necesario pasar por ellos. Por otro lado, la intermediación 0 en el resto de los nodos implica que no es necesario pasar por ellos para llegar a otro nodo.

## Red conexas

Una red conexas implica que se puede recorrer el camino entre cualquiera de todos sus nodos. En contraposición, una red inconexas es una red en que alguno de sus nodos no puede conectarse con los demás.

En el caso de una red de coautoría se espera que sea una red inconexas dado que alguno de sus nodos participantes pueden ser autores independientes, que solo han publicado trabajos por su cuenta y no han participado con otros autores.

## Componente

Según [Morel et al. \(2009\)](#), un componente es una porción de la red en donde todos los nodos están conectados por lo menos por una arista. El análisis de componentes da un vistazo general de la estructura de la red ([Fig. 1.4](#)) y provee información valiosa en relación a su nivel de fragmentación. También debe considerarse el caso de autores individuales, que son componentes por sí mismos.

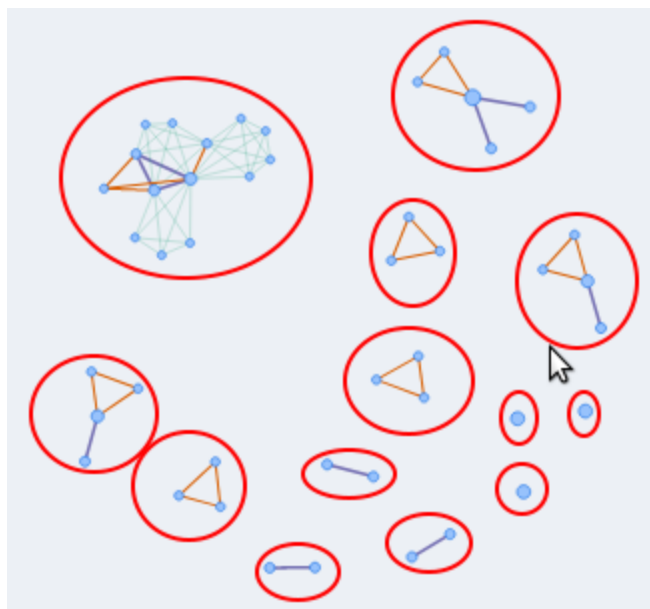


Fig. 1.4: Ejemplo de gráfico de componentes en una red de coautorías. Los componentes se resaltan en rojo.

## Comparación de modelos

En los análisis de redes es de utilidad realizar comparaciones con modelos.

*“El estudio de mecanismos propuestos para generar ciertas propiedades que se observan comúnmente en redes del mundo real (como distribuciones de libre escala o efectos de mundo pequeño)”* (Kolaczyk-Csárdi 2014: 69).

## Mundo Pequeño

El estudio de “mundo pequeño” (*Small world*) está asociado a evaluar ciertas propiedades de la red que permiten que se pueda afirmar que las personas o nodos se conectan con otras a través de una cantidad pequeña de saltos, más allá de que se hallen muy separadas entre sí en la red.

Las propiedades clave son un alto coeficiente de *clustering*, también llamado transitividad, y un promedio bajo de camino más corto. El coeficiente de *clustering* de un nodo mide hasta qué punto los vecinos de un nodo también están interconectados, mientras que el coeficiente de clustering de la red es el promedio del coeficiente de clustering de todos los nodos. El promedio de camino más corto entre nodos, mide el camino más corto de un nodo A a un nodo B. Para esta métrica también se trabaja con el promedio para todos los nodos en la red. Para un análisis más detallado de estas métricas se puede ver [Humphries-Gurney \(2008\)](#) que analizan dos versiones del coeficiente de clustering una propuesta por [Watts-Strogatz \(1998\)](#) y otra presentada por [Newman-Moore-Watts \(2000\)](#).

Luego de obtener estos valores se comparan contra el promedio de las mismas métricas obtenidas en grafos aleatorios utilizando el modelo de [Erdős-Rényi \(1959\)](#); con los mismos parámetros  $m$  y  $n$ , siendo estos la cantidad de vértices y la cantidad de aristas.

Para comprender un poco de sus orígenes, el estudio de *Small world*, surge de la frase en inglés ‘*My it’s a small world*’ / ‘*Que mundo pequeño*’ cuando dos personas conocen a otra y, por lo tanto, estarían interconectados a través de una tercera. Se relaciona además al estudio de [Milgram \(1967\)](#) que consistió en enviar mensajes desde una persona en un punto A hasta otra en un punto B bajo ciertas reglas, como por ejemplo que solo podían mandar el mensaje a personas conocidas inmediatas.

Para ganar una intuición gráfica puede verse la [Fig. 1.5](#) donde se presenta como ejemplo el caso de que el autor en el círculo rojo desea contactarse con el autor en el círculo azul. Cabe recordar que las aristas que unen nodos son trabajos realizados en conjunto.

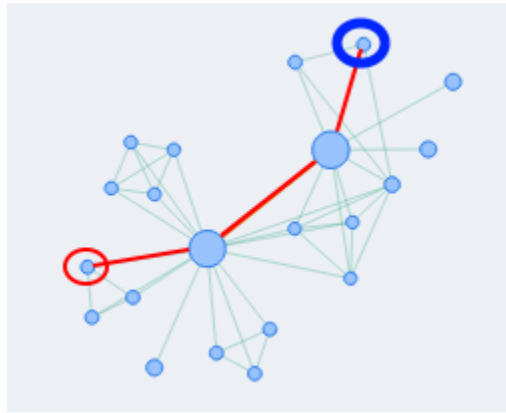


Fig. 1.5: Si hablase el nodo del círculo rojo que desea contactarse con el nodo azul “*El coautor de mi trabajo original, tiene otro coautor que trabajó con el nodo en círculo azul que nos podría presentar*”

## Libre Escala

La propiedad de libre escala se basa en analizar la distribución de grado de la red. Cabe recordar que el grado de cada nodo representa con cuántos otros nodos se conecta directamente el nodo analizado. Una red Libre escala es una red en la cual su distribución de grado sigue la ley de potencia ([Barabási 2016](#)), lo cual significa que la red tiene muchos nodos con pocas conexiones y pocos nodos con muchas conexiones. La fórmula correspondiente es:

$$Pk \sim k^{-\gamma}$$

Siendo la P la probabilidad del grado k es equivalente a k elevado a la  $-\gamma$  siendo un exponente a averiguar.

## Comunidades

Explicado en palabras de [Savic et al. \(2015\)](#): “*La estructura de comunidad es una propiedad típica de las redes sociales. Una comunidad (agrupamiento o módulo) es una parte de una red (grupo de nodos) donde las conexiones internas son más densas que las externas. Descubrir comunidades nos ayuda a entender la estructura de la red, a identificar subgrupos cohesivos y a dibujar un mapa entendible de la red.*” (Savic et al. 2015: 2).

La necesidad de búsqueda de comunidades está asociada a buscar patrones según ciertas propiedades de la red. Esta búsqueda de patrones se puede realizar de diferentes maneras,

para lo cual existen distintos algoritmos. La búsqueda de comunidades se puede realizar sobre el componente principal o sobre el total de la red. En casos como el del presente trabajo resulta de interés realizar la búsqueda sobre el total de la red dado que existen componentes grandes, más allá del componente principal.

## Características del caso de aplicación

La “Revista Argentina de Antropología Biológica” (RAAB) surge luego del éxito alcanzado en las Primeras Jornadas Nacionales de Antropología Biológica, realizadas en septiembre de 1993. El objetivo de la RAAB, como se mencionaba en su primera edición, fue documentar la producción científica volcada en sucesivas Jornadas Nacionales ([Cesani et al. 2013](#); [Sardi et al. 2016](#)). La RAAB es el órgano oficial de comunicación científica de la Asociación de Antropología Biológica Argentina (AABA), por lo que es elegida como uno de los medios de publicación por parte de las personas que integran la asociación.

En 2016 se cumplieron 20 años desde la primera publicación de la RAAB en 1996. La revista publica por lo menos un volumen anual sin interrupciones desde 2001 de uno o más números, y desde 2014 cada volumen anual presenta 2 números. A estos números a lo largo de la tesis también se los llama “ediciones”. Cada número tiene artículos publicados que pueden corresponder a diferentes categorías incluyendo: artículos científicos originales, de revisión, comentarios bibliográficos, reportes técnicos y comunicaciones breves.

En este trabajo se utilizaron sólo los Trabajos originales para el análisis de la red de coautoría de la revista, siguiendo a [Ordoñez et al. \(2009\)](#) quienes mencionan que uno de los criterios de calidad que hacen a una revista especializada es el componente investigativo y original de sus artículos. Las temáticas de las publicaciones abarcan numerosos campos vinculados a la antropología biológica, como por ejemplo: evolución y variación humana, adaptación, genética, crecimiento y desarrollo, antropología forense, morfología y primatología, entre otros. Los idiomas de publicación incluyen español, portugués e inglés. Por estas características la RAAB recibe colaboraciones tanto de autores locales como del ámbito internacional, principalmente latinoamericanos. En este sentido, el análisis de coautorías brinda una oportunidad para estudiar las colaboraciones dentro de la disciplina tanto a nivel local como regional.

La comunidad científica que realiza investigaciones en antropología biológica en el país es relativamente pequeña y si bien los trabajos son frecuentemente publicados en diversas revistas, muchas veces internacionales, el análisis de la RAAB en particular puede pensarse como una primera aproximación al estudio de algunos aspectos de la dinámica disciplinar. Teniendo en cuenta que la RAAB acompañó la consolidación de la disciplina en el país y todas las características anteriormente mencionadas, puede plantearse que el crecimiento y evolución de los grupos de investigación tendrían un correlato en la dinámica de nodos, componentes y artículos a lo largo del tiempo en la red de coautorías.

Finalmente, cabe destacar que esta revista es de acceso abierto y sus artículos se encuentran disponibles en internet (<https://revistas.unlp.edu.ar/raab>) y desde Junio 2021 esta disponible en Scopus, una de las bases de datos incluidas en el denominado *mainstream* internacional

([Beigel 2014](#)). Sin embargo, al inicio de este trabajo la RAAB figuraba solamente en contextos latinoamericanos (como por ejemplo SciELO), lo cual constituyó uno de los motivos de su elección como caso de aplicación de esta tesis y resultó determinante en relación a la metodología empleada para el armado de la base de datos y su análisis.



## Objetivos

El **objetivo general** de la presente tesis es realizar un aporte al estudio bibliométrico de revistas científicas a través del desarrollo e implementación de una herramienta para análisis de redes de coautoría.

Para ello, se plantearon dos objetivos específicos.

### Objetivo 1: Desarrollar una herramienta para el análisis de redes de coautoría

La creación de una herramienta fue planteada para facilitar el análisis de redes de coautoría científica mediante la simplificación de tareas repetitivas, como por ejemplo: variación de parámetros para aplicar filtros y/o algoritmos, selección de diferentes nodos/aristas/grupos, aplicación de filtros a los datos, visualización general de la red desde diferentes puntos de vista, obtención de detalles de nodos y aristas, obtención de detalles de los datos, entre otras.

Dentro de este objetivo se incluyó facilitar la exploración de los datos mediante interacciones intuitivas brindadas a través de una interfaz *web* de acceso abierto para diferentes personas interesadas en realizar los análisis. Se plantearon tres finalidades principales de la herramienta: un breve análisis de los datos que dan origen a la red, el análisis de la red para ciertos periodos seleccionados, y finalmente el análisis temporal de la red.

### Objetivo 2: Analizar redes de coautorías de la RAAB

Utilizando la herramienta desarrollada, se propuso analizar el caso particular de la Revista Argentina de Antropología Biológica (RAAB), principalmente en relación a cómo fue el comportamiento de la red en el tiempo, considerando el período de 20 años desde su creación (1996 a 2016). Dentro de este objetivo, se plantearon cuatro análisis principales:

- 1) Verificar si la disciplina Antropología Biológica creció en el país analizando la cantidad de autores y artículos a lo largo del tiempo en la Revista Argentina de Antropología Biológica.
- 2) Analizar si el avance de la revista depende de los autores fundadores o si está impulsado por nuevos autores a través de tres métricas: Fuerza de colaboración, Grado e Intermediación.
- 3) Entender en mayor profundidad el comportamiento interno de la red analizando la evolución temporal del tamaño del componente principal e identificando momentos específicos de absorción de componentes.
- 4) Realizar validación de propiedades comúnmente presentes en redes del mundo real como Mundo Pequeño y Libre Escala, y contrastarlas con lo obtenido en redes de coautoría realizadas en otros trabajos.

## Hipótesis

En congruencia con lo planteado en el Objetivo 2, se plantearon las siguientes hipótesis:

- Hipótesis 1: La cantidad de autores y artículos aumenta en el tiempo.
- Hipótesis 2: Los autores fundadores se refuerzan en el tiempo, lo cual se relaciona con un aumento en alguna de las métricas (Grado, Intermediación o Fuerza de colaboración).
- 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.
- 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.

## Capítulo 2: Materiales y Métodos

El proceso realizado durante esta tesis puede dividirse en tres grandes etapas: 1) generación de la base de datos, 2) desarrollo de la aplicación para análisis de la red de coautoría, y 3) análisis de la red. En esta sección se mencionan de manera general las principales tareas realizadas y algunas herramientas utilizadas, mientras que la información más detallada sobre los procedimientos empleados se encuentra en los anexos correspondientes.

### Generación de la base de datos

La generación del conjunto de datos fue realizada casi en su totalidad en *scripts* en Lenguaje de programación R ([R Core Team, 2021](#)) con múltiples librerías. Para la sección del procesamiento de autores se utilizó la herramienta Openrefine.

Las herramientas de software utilizadas se detallan en el [Anexo: software y librerías utilizadas](#), donde puede verse qué librerías se usaron para cuáles secciones y su correspondiente cita. Sin embargo, a lo largo del presente capítulo se mencionan algunos detalles sobre el uso de herramientas específicas.

### Análisis de la estructura del sitio

La Revista Argentina de Antropología Biológica se encuentra en formato digital y consta de una página web (<https://revistas.unlp.edu.ar/raab>) en la que están disponibles todos los artículos publicados durante los 20 años considerados para este estudio. Es importante destacar que la utilización de datos públicos, como los artículos publicados en la revista, es un enfoque altamente favorable para la replicación de los resultados obtenidos, así como para la aplicación de la metodología en casos similares.

En primer lugar se estudió la estructura del sitio web de la RAAB para identificar la información que pudiera ser de utilidad en el presente estudio. Se identificó que existía una página “archivo” que consta de un listado de todas las ediciones de la revista. Para cada edición de la revista, existía una página en la que se publican datos de la edición y las secciones con los *links* a los artículos correspondientes, así como sus archivos PDF relacionados.

Una vez determinados los datos a extraer, se procedió al armado de los *scripts* para realizar *web-scraping*, luego para el procesamiento de los mismos, y finalmente para el almacenamiento en un formato más conveniente.

### Modelo de Bases de datos

En esta tesis se utilizó una versión simplificada del modelo propuesto por [Mallig \(2010\)](#). Luego de explorar la estructura de la página web se decidió solo conservar las entidades Persona / Autoría / Artículo; conservando los datos completos de la web como son nombre de autor, nombre del artículo, y la relación de autoría que existe entre ambos. Con respecto a los datos de publicación de la entidad Journal del modelo citado sólo se incluyeron ediciones y años dado que esta tesis involucró un análisis de redes solamente de un journal en particular. Finalmente, no se incluyeron Orden, Referencias, Citaciones, Afiliaciones, ni Organizaciones,

dado que en el año en que se realizó la extracción inicial, la página de la revista no contaba con estos datos para todos los artículos y autores.

## Obtención de los datos: Web scraping

Para realizar la extracción de datos se usó *web-scraping* que es una técnica de extracción de información de la *web*. Mediante *scripts* se automatizó la extracción de la información necesaria y su posterior transformación a formatos deseados para su explotación. La explicación de este proceso con mayor detalle, incluyendo los datos extraídos, descarga e interpretación pueden ser vistos en el [Anexo: Web Scraping](#).

El foco del *web-scraping* de la página fue extraer los datos de los artículos y las ediciones de la revista (a veces referidos como volúmenes). Los mismos se encontraban en diferentes formatos, principalmente HTML, pero también se encontraron archivos PDF asociados a los artículos y anexos.

Una vez extraídos los archivos de Internet, se creó en primer lugar una estructura de base de datos relacional para almacenar el resultado de la extracción y, de esta forma, facilitar la consulta a los datos obtenidos.

Para este caso específico además de la estructura de base de datos relacionales se generó otra en formato de grafos. El resultado de la base relacional fue transformado a un formato de grafos a fin de poder utilizarlo posteriormente en la herramienta de análisis de coautoría, facilitando la exploración de los datos.

A continuación se presentan brevemente algunos de los procedimientos empleados. Más detalles acerca de todas las metodologías de esta sección pueden encontrarse en el [Anexo: resumen proceso obtención y generación de la base de datos](#).

## Resultados intermedios

Luego de la extracción inicial, los datos se guardaron en tablas intermedias en *data.frames* de R y a su vez en archivos intermedios con formato de texto separado por tabulaciones. Generar archivos intermedios forma parte fundamental del proceso a fin de no repetir tareas y ahorrar tiempos de procesamiento en caso de querer ejecutar todo el procedimiento nuevamente. Cada función intermedia fue configurada con parámetros por defecto, que pueden ser sobreescritos en caso de desearlo.

## Generación de la base relacional

El armado de las tablas finales requirió de un extenso proceso de estandarización de las referencias a los autores (nombres y apellidos) y sus filiaciones institucionales. En el caso de los autores se logró de manera exitosa, mientras que en el de las filiaciones surgieron dificultades, como la falta de datos disponibles en ese momento en la revista, que imposibilitaron la estandarización y su posterior uso. Para más detalle de este proceso puede consultarse el [Anexo: Detalle estandarización autores](#).

Una vez finalizado el proceso de estandarización de autores, las tablas finales se guardaron en una base de datos relacional *SQLite*. Se optó por esta herramienta dado que es fácilmente instalable a través de la [librería RSQLite](#) y brinda la posibilidad de tener un formato de base relacional sin la necesidad de instalar programas externos a R. Asimismo, permite el

almacenamiento como base de datos relacional en un formato conveniente para su reproducibilidad.

La estructura final fue una N:M, en donde N autores pueden participar en la creación de M artículos. Para representar esta estructura se contó con una tabla de artículos, una tabla de autores, y una tabla intermedia uniendo autores y artículos, en este caso denominada “autores\_articulos”. Un mayor detalle de estos procedimientos se encuentra en el [Anexo: Base de datos: Estructura relacional](#).

### Generación de los grafos

Los grafos fueron generados de manera tal que respondan a la aplicación de ciertos filtros de búsqueda a la base relacional final. Además, se incorporó el filtrado de estos datos para poder acceder al resultado de forma más eficiente según los criterios escogidos. En particular, estos filtros se relacionaron con distintos periodos (años) a considerar en el análisis de la revista. Para más detalles acerca de la generación de grafos y las propiedades de sus vértices y aristas se recomienda consultar el [Anexo: Generación de Grafos](#).

## Desarrollo de la aplicación para análisis de redes de coautoría

El desarrollo de una herramienta para análisis de redes de coautoría (objetivo 1 de esta tesis), tuvo como finalidad centralizar tareas en una aplicación *web*. De esta forma, se buscó facilitar la disponibilidad de herramientas rutinarias de análisis de redes tanto para su utilización en el presente trabajo como para potenciales futuros usuarios. En este sentido, se consideró clave brindar la posibilidad de exploración de los datos mediante interacciones intuitivas a través de una interfaz *web* para facilitar el acceso a diferentes personas.

La aplicación fue creada enteramente en Lenguaje R con múltiples librerías, que se detallan en el [Anexo: software y librerías utilizadas](#). Entre ellas se destaca *Shiny*, que permite crear una aplicación *web* en el entorno de R, facilitando la reproducibilidad al reducir las herramientas utilizadas dado que no se utilizan otros lenguajes para poder crear la aplicación.

Para el uso de las herramientas de trabajo y buenas prácticas se siguieron las recomendaciones de [Wilson et al. \(2017\)](#).

Para crear la aplicación se utilizó principalmente el entorno de desarrollo RStudio ([RStudio Team, 2015](#)), dado que facilita considerablemente el desarrollo de aplicaciones y *scripts* con el lenguaje R.

Se armó la aplicación en forma de proyecto como un Rproject, puesto que ayuda a gestionar los archivos de código. Esta forma de trabajo resultó útil principalmente para gestionar referencias entre archivos y facilitar el uso de algunas librerías.

### Estructura de la interfaz gráfica

Para poblar las pantallas de la herramienta de análisis de red de coautoría, se utilizaron las tablas finales de la base de datos relacional.

Para armar el esqueleto de la interfaz gráfica se utilizó la librería [Shiny](#) que es la librería de referencia en lenguaje R para realizar aplicaciones *web* y, como complementos,

[shinyDashboard](#) que se utilizó para segmentar la aplicación en diferentes secciones; y [shinyDashboardPlus](#), a fin de tener disponible el menú lateral derecho que brinda la sección de configuración.

A continuación se mencionan algunas metodologías empleadas durante el desarrollo de diferentes secciones de la aplicación: 1) Consideraciones; 2) Análisis Estático; y 3) Análisis Temporal.

## Sección 1: Consideraciones

Esta sección fue desarrollada con el fin de proveer un análisis exploratorio de los datos. Este análisis se focalizó en el acceso a los datos “crudos”, la composición de los datos, un resumen estadístico y algunas preguntas básicas de interés.

El acceso a los datos “crudos” se implementó mediante una tabla interactiva creada utilizando la [librería DT](#) (por *data tables*). Para visualizar la composición de las tablas en relación a los tipos de datos y la cantidad de observaciones se utilizó la librería [visdat](#). Utilizando la librería [skimr](#) se agregó como funcionalidad un resumen estadístico de la distribución de los datos en la tabla. Finalmente, se incorporaron algunas preguntas básicas que pueden utilizarse para familiarizarse con los datos mediante gráficos interactivos realizados utilizando la librería [plotly](#).

## Sección 2: Análisis Estático

La finalidad de esta sección fue poder configurar, filtrar, visualizar, agrupar y estudiar métricas de la red en detalle. Para realizar estas tareas, esta sección se dividió en varias subsecciones: Visualización de coautoría, Artículos asociados, Estructura de la red, Comparación con modelos y Comunidades.

### Visualización de coautoría

La sección de visualización incluyó diferentes tipos de visualización del grafo: vista de red, Biofabric y Mapa de calor (heatmap). Esta sección fue armada con controles de la [librería shiny](#). La visualización en forma de red fue realizada con la librería [visnetwork](#). La Visualización Biofabric ([Longabaugh, 2012](#)) fue armada con la adaptación de R disponible en la librería [RBioFabric](#). Finalmente, la visualización de tipo mapa de calor (heatmap) fue realizada con las librerías [ggplot2](#) y [plotly](#).

### Artículos asociados

Esta sección se desarrolló de manera asociada a la visualización en formato de red, siguiendo a [Newman \(2004\)](#). Se incorporaron las funcionalidades tanto de selección de nodo Autor como de arista, y se incluyeron en los datos asociados visualizaciones de subgrafo y *heatmap*, y un resumen de diferentes métricas.

### Estructura de la red

En esta sección se incluyeron métricas asociadas a la red, los nodos y los componentes, que resultan útiles para extraer conclusiones a partir de las visualizaciones.

Las métricas de la red fueron seleccionadas en base a las propuestas por [Newman \(2004\)](#). La gran mayoría de las métricas se obtuvieron con la librería [igraph](#), exceptuando a la fuerza de colaboración, que fue calculada por el autor de esta tesis en base a los cálculos especificados por [Newman \(2004\)](#).

El Promedio de colaboradores no fue utilizado en esta tesis dado que en este caso la revista no contaba con información explícita para una correcta distinción entre el autor principal y los colaboradores sino que solamente figuraba el orden en la página de internet. Además, se

agregaron otras métricas como: el porcentaje que representa el *cliqué* más grande, la densidad de la red, la cantidad de autores en el componente más grande, la cantidad de componentes, la cantidad de contribuciones y la cantidad de relaciones. Más detalles de las métricas incluidas se encuentran en el [Anexo: métricas de red disponibles](#).

En relación a los nodos, teniendo en cuenta la revisión de las medidas dentro de las redes de [Newman \(2016\)](#) y [Freeman \(1978\)](#), se incluyeron en la visualización distintas métricas (que se describen en el [Anexo: métricas de nodos](#)), la distribución de grado, el histograma de variables y la densidad de variables.

Finalmente, la extracción de componentes se realizó utilizando la librería [igraph](#) y se incorporó la funcionalidad de obtención de métricas para cada componente de manera similar a las utilizadas para la red total ([Anexo: métricas de red disponibles](#)). En esta sección se incorporó además un análisis de texto realizado a partir de los títulos de los artículos asociados al componente, utilizando la librería [tidytext](#). Este análisis incluyó la identificación de las palabras más frecuentes en ese componente, la cantidad de palabras frecuentes y los filtros aplicados de *stop words* en español e inglés (dado que existen algunos títulos en dicho idioma).

### Comparación con modelos

Los modelos que se incluyeron en la aplicación son: el de Red Aleatoria según el modelo Érdos-Rényi configurado a través de nodos y aristas, el modelo de Mundo Pequeño de Watts-Strogatz, y el de Libre Escala de Barabasi-Albert ([Kolaczyk-Csárdi 2014](#)).

Las comparaciones de la red con los modelos de Mundo Pequeño (*Small World*) y Libre escala (*Scale-Free*) se realizaron teniendo en cuenta ciertas propiedades de la red que se describen en el [Anexo: Propiedades modelos](#).

Para la creación de los grafos a partir de modelos se utilizó la librería [igraph](#). La evaluación de los modelos de forma visual se realizó a través de histogramas utilizando la librería [ggplot](#).

### Validación de Mundo Pequeño

Para la validación de Mundo Pequeño se utilizaron dos métodos de generación de redes aleatorias: la validación 1 descrita en [Bialonski et al. \(2010\)](#) y la validación 2 se basa en lo planteado por [Humphries-Gurney \(2008\)](#). Ambos utilizan el método de generación de grafos aleatorios de [Erdős-Rényi \(1959\)](#).

En la validación 1 se utilizó el método implementado a través de la función `sample_gnm` de la librería [igraph](#). Por otro lado, para la validación 2, se usó la función `smallworldness` de la librería [ggraph](#) que utiliza la función `degree.sequence.game` de la librería [igraph](#) para generar los aleatorios, pasando el parámetro de la función `method` como `"simple.no.multiple"`. Esto hace que los resultados obtenidos varíen entre validación 1 y 2.

Los parámetros de configuración de la ejecución de `degree.sequence.game` son diferentes, el único configurado es la cantidad de iteraciones: se usó la misma cantidad que para el modelo configurado, en este caso 1000. La validación 1 se basa en la cantidad de nodos y de aristas del grafo original, y se hacen 1000 iteraciones aleatorias para obtener los resultados. Por otro lado, `degree.sequence.game` se basa en generar un grafo aleatorio dada una lista de grados tomada del grafo original, en la implementación de la función `smallworldness`. Es importante aclarar que en la documentación de la función `degree.sequence.game`, se recomienda ver otras funciones por problemas de bucles que se generan al ejecutarla y, entre ellas, recomienda el modelo de Erdős-Rényi basado en probabilidades. Para más información se recomienda ver la documentación de la librería [igraph](#).

### Validación de Libre Escala

Para verificar que la distribución de grado de la red sigue la ley de potencia se utilizó la función `fit_power_law` de la librería [igraph](#), que se basa en lo planteado por [Clauset, et al. \(2009\)](#) y



[Newman \(2005\)](#). Esta función puede recibir como parámetros los valores a ajustar a la ley de potencia, los utilizados dentro de la herramienta que corresponden a una lista con el grado de todos los nodos de la red, y el tipo de implementación *plfit*. En esta tesis se utilizaron como valores a ajustar los grados de los nodos de la red y la implementación *plfit* que utiliza máxima verosimilitud para determinar el alfa dado un parámetro *xmin*. Cuando no se brinda el parámetro *xmin*, como fue en este caso, el algoritmo busca el valor óptimo para maximizar el p-valor de un test de Kolmogorov-Smirnov entre la distribución ajustada y la muestra original. La función utiliza el método de [Clauset, et al. \(2009\)](#) para calcular los parámetros de la distribución ajustada.

## Comunidades

Para la búsqueda de comunidades se emplearon los algoritmos que se encuentran disponibles en el [Anexo: Detalle de algoritmos para búsqueda de comunidades](#). Esta sección en la aplicación se desarrolló con la finalidad de obtener visualizaciones de los subgrafos resultantes de la búsqueda de comunidades.

La relevancia de los resultados de los algoritmos se analizó a través de la medida de modularidad ([Newman-Girvan 2004](#)) tanto teniendo en cuenta como sin tener en cuenta el peso de las relaciones. La implementación utilizada es la que se encuentra en el paquete *igraph*, que emplea la versión de [Clauset et al. \(2004\)](#). En este caso, los pesos de las relaciones en este caso están dados por la fuerza de colaboración de cada relación. En esta sección se incluyó además un análisis de la estructura de cada comunidad, la obtención de métricas y un análisis de texto realizado sobre los títulos de los artículos de manera similar a lo descripto anteriormente para los Componentes.

## Sección 3: Análisis temporal

En esta sección se incluyó un análisis temporal de la red desde su primera publicación en 1996 hasta el periodo 2016 puesto como cota por tratarse de los 20 años desde su primera edición. Como vistazo inicial de este análisis se planteó una visualización dinámica. Para su generación se utilizó el grafo acumulado desde el periodo 1996 como  $T=0$  hasta 2016 como  $T=18$ . Para facilitar el análisis de diversos aspectos de la red se incluyeron diferentes medidas como, por ejemplo, qué nodos cumplen distintos roles. Estas medidas se implementaron para análisis de forma individual por periodo o de forma acumulada total, siguiendo el trabajo realizado por [Koseoglu \(2016\)](#).

La animación del grafo de manera acumulada se realizó utilizando los paquetes *igraph*, *ndtv*, *intergraph*, *networkDynamic* y *network*. Para más detalles sobre estas librerías se recomienda ver el [Anexo: software y librerías utilizadas](#).

Para cada período (año) se incluyó además el cálculo de las métricas mencionadas anteriormente en la sección Estructura de la red ([Anexo: métricas de red disponibles](#)). Los gráficos fueron realizados con la librería *ggplot* y se hicieron interactivos utilizando la librería *plotly*. Estas mismas medidas y gráficos se incluyeron además de manera acumulada para cada período.

Finalmente, se realizaron gráficos para analizar el ranking de autores según alguna de las métricas implementadas y su evolución temporal. Las métricas utilizadas son las mismas que las especificadas anteriormente. Los gráficos fueron realizados con la librería *ggplot* y se hicieron interactivos utilizando la librería *plotly*. Asimismo, se incorporó la visualización del ranking para las métricas acumuladas en el tiempo.



## Análisis de la red

El análisis de la red de coautorías de la RAAB se realizó utilizando la herramienta desarrollada a fin de poner a prueba las cuatro hipótesis planteadas.

### Hipótesis 1: La cantidad de autores y artículos aumenta en el tiempo

Se analizó la cantidad de publicaciones por año, discriminado en secciones y por ediciones. Luego se profundizó el análisis en la cantidad de autores y artículos de tipo Trabajo original, en el período de tiempo acotado entre 1996 y 2016. Si bien estos análisis brindaron una primera aproximación, la puesta a prueba de esta hipótesis se realizó mediante pruebas estadísticas.

#### Pruebas estadísticas empleadas

Para evaluar la presencia de tendencias en la serie de tiempo se empleó la prueba no paramétrica de Mann-Kendall con Bootstrap de Bloques. Esta prueba fue seleccionada debido al bajo número de períodos evaluados ( $n < 30$ ), lo que impide asumir normalidad, y adicionalmente, a que es robusta ante la presencia de autocorrelación ([Mann, 1945](#); [Kendall, 1975](#); [Kundzewicz-Robson, 2000](#); [Khaliq et al. 2009](#); [Önöz-Bayazit, 2012](#)).

La hipótesis nula ( $H_0$ ) para esta prueba es la *ausencia de una tendencia en los datos* mientras que la hipótesis alternativa ( $H_a$ ) consiste en afirmar que *hay evidencia de la presencia de una tendencia, positiva o negativa, en los datos*.

La prueba se aplicó a la cantidad total de autores por año y a la cantidad de publicaciones originales por año durante el periodo estudiado. Para ambas pruebas se empleó un nivel alfa crítico de 0,05 y se realizaron con la función *bbsmk* de la librería [modifiedmk](#).

Finalmente, como análisis complementario, se evaluó la cantidad de autores por artículos por año.

### Hipótesis 2: Los autores fundadores se refuerzan en el tiempo, lo cual se relaciona con un aumento en alguna de las métricas (Grado, Intermediación o Fuerza de colaboración)

Para evaluar si el avance de la revista fue dependiente de los autores fundadores (aquellos que publicaron en la primera edición) o si estuvo impulsado por la incorporación de nuevos autores independientes, se analizaron tres métricas: Grado, Intermediación y Fuerza de colaboración. Cada una de estas métricas brinda formas de evaluar los nodos de la red desde diferentes puntos de vista. Por otro lado, evaluar la pregunta de si se refuerzan o no en el tiempo aporta diferentes visiones respecto de los roles que cumplen los autores dentro de la red. Estos roles pueden ser Concentrador (*Hub*), quién trabaja más con los demás y que genera nuevos equipos y es medido por el Grado; Autoridad (*Authorities*), quiénes son generadores de contenido, es decir, quienes más aportes hacen en cantidad de trabajos originales, lo cual está relacionado con la Fuerza de Colaboración; y finalmente quienes generan puentes entre otros autores, quienes tal vez puedan generar nuevas tendencias con mayor facilidad, uniendo equipos con los que trabajaron alguna vez o, también, que su

desaparición de la red puede desconectar equipos, asociados a la distancia de Intermediación ([Kleinberg 1999](#)).

Para poner a prueba esta hipótesis se utilizaron pruebas estadísticas para cada una de las métricas.

### Pruebas estadísticas empleadas

Con el objetivo de comprobar si las métricas empleadas (Grado, Fuerza de colaboración e Intermediación) de los autores fundadores se incrementaron al pasar el tiempo, para cada métrica se seleccionó el período 1996 y el 2016 y se empleó la prueba de Rangos con Signo de Wilcoxon para muestras pareadas. En la hipótesis nula se afirma que no hay diferencia entre la métrica de los autores para el año 1996 y para los mismos autores en el año 2016. Dado que se quiere evaluar si hubo un aumento en la métrica, se analiza la cola derecha de la distribución y en consecuencia la hipótesis alternativa corresponde a afirmar que la fuerza de colaboración de los autores fundadores en el año 2016 es mayor que en el año 1996 para los mismos autores. Como se tiene una muestra mayor a 15 ( $n = 45$ ), lo que puede considerarse una muestra grande ([Siegel y Castellan, 1972](#)), se realizó la corrección por continuidad. La prueba se realizó empleando la función `wilcox.test` de la librería [stats](#) de R y se seleccionó un nivel alfa crítico de 0,05.

Finalmente, de forma complementaria, se realizó un análisis del ranking de cinco autores por año y su evolución temporal para cada una de las métricas.

### 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

Para evaluar esta hipótesis se realizó por un lado un análisis de componentes de la red de coautorías de la RAAB y, por otro lado, una búsqueda de comunidades. En ambos casos se realizó adicionalmente un análisis de minería de texto.

El objetivo asociado a esta hipótesis fue planteado en tercer lugar puesto que es necesario tener un contexto previo para realizar este análisis. Es decir, luego de evaluar la cantidad de autores por artículos por año, resultó de interés ver cómo se fueron generando los componentes, cómo se relacionan entre ellos y su integración en el tiempo.

### Análisis de Componentes

Inicialmente se evaluó la evolución temporal de los componentes por año, teniendo en cuenta las métricas cantidad de componentes, cantidad de autores en componente más grande, y porcentaje de autores en el componente más grande. Este punto de vista brinda una noción de cuándo la red tiene incorporaciones de autores y cómo se comportan respecto de realizar trabajos en conjunto de manera anual. Por otro lado, se realizó un análisis acumulado para observar cómo queda la red total agregada en su evolución temporal, y brinda información acerca de cuándo se realizan absorciones importantes respecto del componente mayor.

Luego de identificar los puntos de incremento de nodos en el componente mayor observando mayor pendiente positiva entre años, se analizaron las transiciones entre periodos evaluando

cómo varía la red estructuralmente de manera visual y corroborando con números según el informe de estructura de la red.

La forma visual fue analizada a través de la animación disponible en la aplicación y comparando cada año inicial de pendiente (T0) con el año final de la pendiente (T1). Esta animación permite observar qué nodos están conectados entre un año T y otro.

Luego, los valores tenidos en cuenta para la comparación numérica se presentan en la aplicación a través de una tabla. Utilizando esta funcionalidad se evaluó el estado inicial del componente mayor, comparado con el estado final, y los números relacionados con cuántos nodos fueron nuevos en el periodo analizado y cuántos fueron absorbidos (es decir, nodos existentes que pasaron a ser parte del componente mayor).

#### Relación con actividades de investigación

De manera complementaria se intentó corroborar si: *“El componente gigante podría significar la principal actividad de investigación (mientras que los otros componentes pueden ser agrupamientos especializados o sub-comunidades)”* ([Fatt et al. 2010:3](#))

Este análisis se realizó sobre la red total, es decir, considerando todo el período de 20 años. Para ello, se revisaron los títulos de los artículos publicados por los cinco componentes con mayor cantidad de autores. Luego se verificó la frecuencia de palabras, previamente filtrando *stopwords* tanto en inglés como castellano dado que algunos artículos contenían ambos idiomas.

#### Búsqueda de comunidades

La búsqueda de comunidades se realizó sobre el total de la red dado que existen componentes grandes más allá del componente principal para la red resultante del periodo 1996-2016. La determinación de cuál algoritmo aplicaba mejor para la red estudiada se realizó a través del criterio de modularidad. Posteriormente se interpretó el resultado para el caso particular de la RAAB, se exploró qué comunidades surgieron y si presentaban cierto sustento temático al igual que en el caso de los componentes.

En primer lugar, se verificó si se encontraban agrupamientos bajo algún criterio determinado por los algoritmos disponibles y también se complementó con análisis de comunidades dentro de la red ([Fig. 2.3](#)).

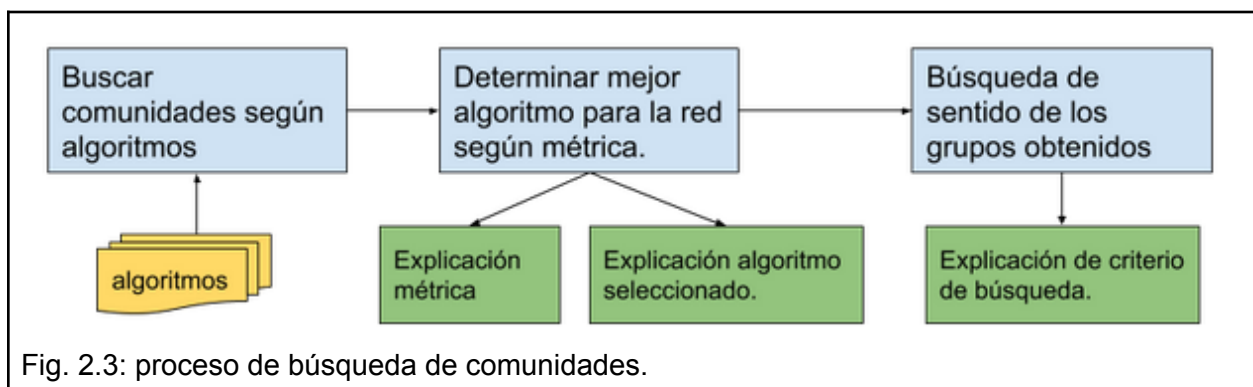


Fig. 2.3: proceso de búsqueda de comunidades.

## Métrica de evaluación de algoritmos: Modularidad

Para la selección del algoritmo se utilizó la métrica de modularidad, propuesta por [Newman-Girvan \(2004\)](#).

Siguiendo a [Barabási \(2016\)](#) puede comprenderse mejor la métrica de la siguiente forma:

Para calcular la modularidad de una red con  $N$  nodos y  $L$  conexiones, particionados en  $N_c$  comunidades; y cada una de las comunidades tiene  $N_c$  nodos y  $L_c$  conexiones, donde  $c$  es el número de comunidad 1 a  $N$ , si el número de  $L_c$  es mayor que el número de  $L$  esperado para la secuencia de grado de la red, entonces los  $N_c$  podrían ser parte de una comunidad.

Entonces para cada comunidad se calcula la diferencia entre la matriz de conexión de la red actual ( $A_{ij}$ ) y la matriz de conexión de los mismos nodos pero conectados aleatoriamente ( $p_{ij}$ ), preservando el grado para los nodos, y se obtiene la siguiente fórmula:

$$M_c = 1/2L * \sum_{(i,j) \in C_c} (A_{ij} - p_{ij})$$

Donde

$A_{ij}$  es la matriz de conexiones actuales,

$P_{ij}$  es la matriz conexiones aleatorias

$C$  representa una comunidad dentro de comunidades

$L$  son las conexiones totales de la red

$\frac{1}{2}$  se debe a que es un grafo no dirigido.

Y como resultado puede obtenerse:

$M_c > 0$ : cuanto mayor Modularidad, mejor el particionamiento.

$M_c == 0$ : todos los nodos en una sola partición

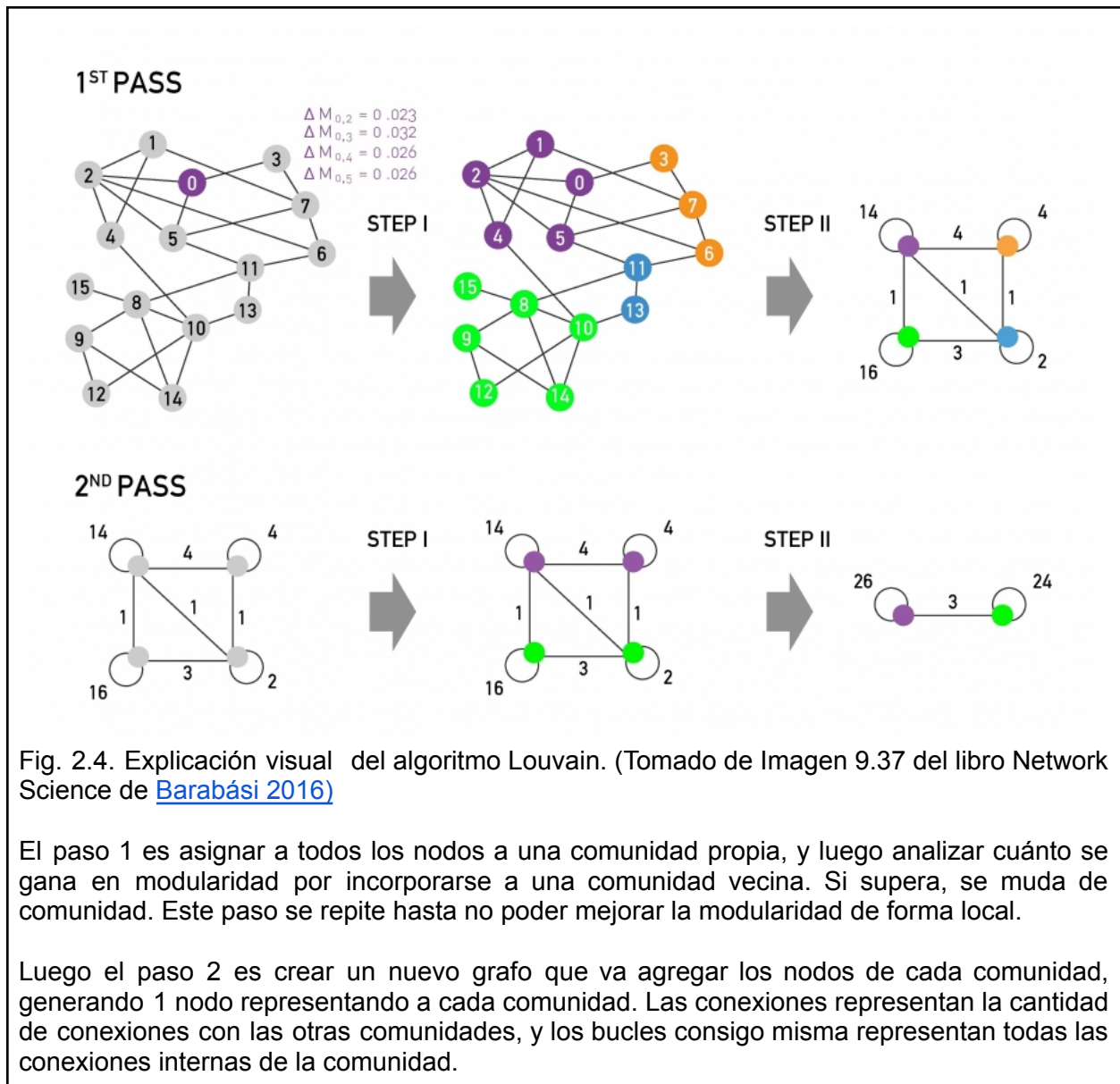
$M_c < 0$ : cada nodo como una partición separada.

Por lo tanto, cuanto mayor sea la Modularidad significa que la configuración de conexiones actual es mejor que una conexión aleatoria y mejor que otras configuraciones.

## Optimización de modularidad multi-nivel

Se evaluaron diferentes algoritmos de optimización que se detallan en el [Anexo: Detalle de algoritmos para búsqueda de comunidades](#). A continuación se detalla solamente cómo funciona el algoritmo de Louvain, que resultó ganador dentro de todos los evaluados ). Para comprender mejor cómo es que se realizan las divisiones con este algoritmo se analizó el trabajo de [Blondel \(2008\)](#) y la explicación de [Barabási \(2016\)](#).

El algoritmo tiene  $N$  Pases y cada Pase debe cumplir 2 pasos. Puede verse la explicación gráfica acompañada por su descripción en la [Figura 2.4](#). Luego de esto se termina el Pase, arrancando un siguiente Pase en el que se repiten los pasos 1 y 2. El algoritmo termina cuando no se puede aumentar más la modularidad ([Fig. 2.4](#)).



### Búsqueda de sentido en los grupos obtenidos

Estudiando los resultados del algoritmo ganador se profundizó en las primeras cinco comunidades según cantidad de integrantes. En ellas se buscó un sentido más allá de métricas estructurales, utilizando la información sobre qué trabajos fueron publicados por los autores pertenecientes a ellas.

Se evaluó cómo estaban compuestas las comunidades propuestas analizando los títulos de los artículos generados. Para ello, se realizó un breve análisis de frecuencias de palabras en los artículos publicados, tomando las primeras 10 palabras de cada comunidad. La idea de estos análisis fue tener una aproximación a qué temas se trabajan en las comunidades identificadas.

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

La evaluación de esta hipótesis se centró en el análisis comparado con modelos para verificar si ciertas propiedades de las redes están presentes en la red de la RAAB.

#### Mundo Pequeño (*small world*)

En primer lugar, se analizó la red por propiedades de Mundo Pequeño en su totalidad, de 1996 a 2016, para contrastar la hipótesis general y luego por periodos, para verificar si este comportamiento varió en el tiempo. Cabe recordar que los 20 años de la revista están repartidos en 18 años de publicaciones dado que en 1997, 1998 y 2000 no hubo ediciones.

La validación de Mundo Pequeño se realizó utilizando los dos métodos disponibles en la aplicación (la validación 1 de [Bialonski et al. \(2010\)](#) y la validación 2 utilizando la función *smallworldness*; ver sección anterior).

La configuración para realizar las simulaciones usando el modelo Ęrdos-Rényi para el caso de mundo pequeño de la red total entre 1996 y 2016 se puede ver en la [Tabla 2.1](#).

Tabla 2.1: Parámetros de simulación utilizados teniendo en cuenta la red total entre 1996 y 2016.

Parámetro	Valor
Iteraciones	1000
vértices	294
aristas	1082
semilla	12345

#### Libre Escala (*free scale*)

En el mismo conjunto de datos utilizado para Mundo Pequeño, se verificó si la red es de Libre Escala. Para ello, se ajustaron los grados de la red a la ley de potencia, luego se comprobó si el parámetro de la función de potencia para esta red superó el valor de 2 y se aplicó la prueba de bondad de ajuste de Kolmogorov-Smirnov para verificar si los datos originales se ajustan a la distribución de ley de potencia. Se empleó un valor crítico alfa de 0,05.

Finalmente, se analizaron ambas propiedades de manera comparativa para la red total de la RAAB (período 1996-2016) y para cuatro particiones equivalentes (1996-2001, 2002-2006, 2007-2001 y 2012-2016).

#### Comparación con otras redes de coautoría

Finalmente, a fin de complementar las comparaciones realizadas con modelos similares a la red obtenida generados en base a sus propiedades, se realizaron comparaciones con otras redes de coautoría de escritura científica, teniendo en cuenta que representan casos similares

de redes obtenidas en el mundo real. Se compararon distintas métricas con las obtenidas en otros trabajos ([Koseoglu 2016](#)), tales como: (a) artículos por autor, (b) coeficiente de clustering (transitividad), (c) tamaño del componente principal, (d) componente principal representado en porcentaje, y (e) distancia media.

Cabe destacar que la comparación realizada sólo está basada en métricas, y no en un análisis profundo de comparación de revistas.

## Disponibilidad de la herramienta: local, en la red, en internet

Inicialmente se trabajó de forma local, esto significa en la misma computadora donde se crean los *scripts*. Esta forma de trabajo es necesaria para pruebas, sin embargo no es suficiente en caso de querer disponibilizar la herramienta para otras personas. Existen opciones para disponibilizar aplicaciones *shiny* en servidores, pueden ser servidores propios con un *shiny server*, o con un *shiny server pro*, y en [shinyapps.io](#). De forma local o en *shiny server* son gratuitas, solo hay que tener la computadora para ejecutar la aplicación, por otro lado se puede pagar el servicio para *shiny server pro* o *shiny apps*. Para la opción de shinyapps.io existe una categoría gratis que da cierta cantidad de horas de uso por mes. Por este tema se consideró la opción de montar un servidor propio de *Shiny*; como plan de *backup* en caso de quedarse sin crédito y no poder pagar una categoría más alta.

Las tres opciones implican cierto nivel de aprendizaje, requisitos para poder lograr el objetivo de disponibilizar las aplicaciones. En relación a esto, [Wojciechowski et al. \(2015\)](#) especificaron más detalles acerca de las ventajas y desventajas.

## Código Fuente

El código fuente de la aplicación se encuentra disponible en [https://github.com/jas1/raab\\_coaut\\_tesis](https://github.com/jas1/raab_coaut_tesis). Dicho proyecto tiene ramas (*branches*) de GIT generadas acorde a puntos de entrega o revisión de la tesis, además de la rama principal que es periódicamente actualizada.



# Capítulo 3: Resultados

## Herramienta desarrollada para el análisis de redes de coautorías

En esta sección se presentarán algunas funcionalidades de la herramienta desarrollada como parte del Objetivo 1 de esta tesis. Se recomienda explorar la aplicación en paralelo a la lectura, utilizando la herramienta en vivo o reproduciendo el código de manera local. La herramienta se encuentra alojada en <https://jaspa.shinyapps.io/raab-coaut-tesis/>.

Algunas de las tareas contenidas en la aplicación son:

- visualización general de la red desde diferentes puntos de vista,
- selección de diferentes nodos/aristas/grupos,
- detalles de nodos y aristas, detalles de los datos, entre otras,
- filtros en los datos,
- variación de parámetros para filtros y/o algoritmos.

La aplicación está dividida en tres secciones principales: Sección 1: **consideraciones**; Sección 2: **análisis estático**; Sección 3: **análisis temporal**. El mapa de las opciones de navegación puede consultarse en el [Anexo: Mapa de la interfaz de usuario](#).

### Sección 1: Consideraciones

Esta sección fue desarrollada a fin de que al comenzar a utilizar la aplicación el usuario pueda relacionarse inicialmente con los datos, ver cómo están estructurados, y obtener una primera aproximación a algunas preguntas. El usuario puede descargar los datos en el caso de querer expandir sus análisis .

Luego de estar familiarizados con los datos se puede proceder tanto a la sección 2 como la 3 dependiendo de si se desea estudiar uno o varios períodos en particular, o analizar la evolución temporal de los datos, respectivamente. Un ejemplo de uso es explorar la evolución temporal (sección 3), para luego focalizarse en algún periodo particular (sección 2).

En esta primera sección se encuentran destacadas algunas consideraciones respecto de los datos utilizados, un análisis exploratorio de los mismos, y algunas preguntas para cada tabla de datos.

Las consideraciones a tener en cuenta a la hora de usar la aplicación corresponden específicamente a que se tomaron los trabajos originales publicados por la RAAB durante el periodo entre 1996 y 2016. En la sección de **análisis exploratorio** se encuentra un “checkbox” donde es posible seleccionar si se desea ver los datos disponibles considerando o no este acotamiento temporal y con respecto al trabajo propio de la presente tesis. Esta opción fue agregada para que quien realice un análisis utilizando la herramienta sepa que el conjunto de datos completo de la revista contiene más información. La utilidad principal se ve a la hora de evaluar la cantidad de artículos en otras secciones (este punto en particular se abordará más adelante).

El análisis exploratorio se focalizó en el acceso a los datos crudos, la composición de los datos, un resumen estadístico y algunas preguntas básicas de interés.



## Datos Crudos

Refiere a la visualización de los datos en forma de tabla interactiva en la cual se pueden aplicar distintos filtros. La herramienta permite bajar los datos, buscar por palabras que figuren en cualquier columna, entre otras.

## Composición

Esta opción muestra un análisis visual de cómo es la composición de las tablas, en cuanto a tipos de datos y cantidad de observaciones.

## Resumen

Permite ver un resumen estadístico de la distribución de los datos en la tabla. Mediante esta opción se pueden ver las estadísticas que se utilizan para cada tipo de dato.

## Preguntas

Incluye algunas preguntas que pueden utilizarse para familiarizarse con los datos. Para ver en detalle las que están disponibles en la aplicación se recomienda ver el [Anexo sección preguntas](#).

## Sección 2: Análisis Estático

La finalidad de esta sección es poder configurar, filtrar, visualizar, agrupar y estudiar métricas de la red en detalle. Para realizar estas tareas esta sección se dividió en varias subsecciones: Visualización de coautoría, Artículos asociados, Estructura de la red, Comparación con modelos y Comunidades. En la [Fig. 3.1](#) podemos ver la pantalla de Visualización de coautoría.



## Configuración

Para configurar el análisis a realizar, se debe hacer click sobre el engranaje como se ve en la [Fig. 3.1](#) en la esquina superior derecha. En esta sección se configura la semilla (*seed*) para asegurar la reproducibilidad en partes aleatorias de las ejecuciones, como podrían ser algunas de las “disposiciones” de la red (*layouts*) o las ejecuciones de algunos algoritmos, por ejemplo, en la parte de Comunidades, las “caminatas aleatorias cortas” (*short random walks*).

Luego, la opción periodos permite seleccionar los periodos a analizar. Al hacerlo, se desprenden opciones de la red resultante en particular, como ser la disposición de la visualización de Red, si se desea seleccionar algún autor en particular, y la opción de búsqueda de comunidades en la red actual.

## Visualización

Utilizando la aplicación es posible obtener diferentes tipos de visualización del grafo: la vista de red, Biofabric y Mapa de calor (*heatmap*).

La visualización en forma de red muestra a los Autores como nodos y su relación de coautoría en forma de vértices. El tamaño de los nodos representa la suma total de fuerza de coautoría del autor. El color de las aristas está basado en la suma de fuerza de colaboración de la relación y pueden seleccionarse tres categorías relativas al grafo actual (que representan baja, media o alta relación de coautoría) para filtrar la visualización de aristas.

La Visualización Biofabric ([Longabaugh, 2012](#)), es otra opción interactiva para ver las relaciones entre autores. En esta se toman los nodos como líneas horizontales para una visualización diferente cuando hay muchos vértices.

Finalmente, la visualización de tipo mapa de calor (*heatmap*) permite observar la fuerza de colaboración entre pares de autores.

## Artículos asociados

Esta sección está asociada a la visualización en formato de red. Al seleccionar un nodo Autor, se muestran los datos asociados a dicho Autor incluyendo las visualizaciones de subgrafo y *heatmap*. En el resumen se muestra quién es el Autor seleccionado, en qué periodos participó, la cantidad de artículos relacionados con dicho autor, la cantidad de autores asociados y la fuerza de colaboración total para dicho autor. Es importante considerar que cualquier métrica o valor tendrá aplicado el filtro correspondiente a los períodos seleccionados en la configuración de la sección de análisis estático.

En esta sección se muestra, además, un detalle de los datos asociados al autor en el período seleccionado, con la intención de comprender mejor con qué datos específicos están asociadas las métricas y visualizaciones obtenidas.

Por otro lado, al seleccionar un vértice se pueden ver los autores involucrados en la relación, los períodos de participación de esa relación, la cantidad de artículos relacionados y finalmente la fuerza colaboración asociada a esa relación.

## Estructura de la red

En esta sección se pueden obtener métricas asociadas a la red, los nodos y los componentes, que potencialmente sirvan para fundamentar conclusiones extraídas a partir de las visualizaciones.

## Métricas de la red

Esta sección dispone de diferentes métricas que se utilizan habitualmente en análisis de redes. Estos datos se pueden exportar mediante el botón *Bajar est. red.* que está ubicado arriba de la tabla. Para detalles de estas métricas se recomienda ver el [Anexo: métricas de red disponibles](#).

## Métricas de los nodos

Esta sección incluye distintas “solapas” donde se puede acceder a los datos (es decir, las métricas propiamente dichas), la distribución de grado, el histograma de variables, y la densidad de variables. Las métricas que se incluyen se describen en el [Anexo: métricas de nodos](#).

Para la sección de datos está disponible el cuadro de búsqueda que agiliza la búsqueda de algún autor en particular. La distribución de grado es un gráfico ampliamente utilizado en el estudio de redes. Analizando la distribución se puede ver qué tipo de nodos predomina más dentro de una red en particular. En la sección Histograma variable podemos obtener los histogramas de las métricas que se mencionaron anteriormente para poder visualizar de forma gráfica los datos. Estas métricas son muchas veces representadas en forma de gráfico de densidades en lugar de histogramas, por lo que la aplicación brinda ambas funcionalidades.

## Detalles de los componentes

En esta sección pueden obtenerse las métricas de grafo para cada componente (*Estructura subgrafo*), los autores pertenecientes a cada componente, y detalles del componente seleccionado y sus nodos asociados (*Detalle subgrafo seleccionado* y *Detalle nodos del subgrafo seleccionado*). En este último caso puede verse qué artículos están relacionados a dicho componente, así como distintas métricas de los nodos correspondientes.

Las métricas que pueden obtenerse para cada componente son las mismas que las que fueron utilizadas para la red total ([Anexo: métricas de red disponibles](#)).

En la subsección *Autores por subgrafo* se detalla la cantidad y los autores que participan de cada componente. Es posible realizar búsquedas por autores, ordenar por cantidades, y también exportar los datos.

En la subsección *Detalle del subgrafo* se encuentra el detalle de los artículos relacionados al componente seleccionado en la subsección *Autores por subgrafo*. Es posible realizar búsquedas, que aplican a todas las columnas, y exportar los datos asociados.

También se cuenta con secciones relacionadas con el análisis de texto. Puede verse, por ejemplo, la frecuencia de palabras en los títulos de los artículos asociados al componente seleccionado.

Finalmente, se ofrece un detalle de los nodos del subgrafo o componente seleccionado con las mismas funcionalidades descritas anteriormente: métricas de nodos, distribución de grado, histograma de variables y densidad de variables.

## Comparación con modelos

Esta subsección está compuesta por tres partes: Datos del modelo, Modelos y Verificaciones de propiedades.

En la solapa de *Modelos* disponemos de los tres modelos que se pueden simular para comparar con la red actual seleccionada previamente (Red Aleatoria, Mundo Pequeño y Libre Escala). Para cada uno de estos modelos existen distintas secciones: configuración de parámetros, ejecución del modelo, histogramas para comparar, y los datos resultantes de la simulación.

Luego de la ejecución del modelo es posible evaluarlo de forma visual a través de histogramas de los modelos en comparación con la red.

Por otro lado, pueden explorarse los datos asociados a la estructura de la red de cada una de las iteraciones del modelo teniendo en cuenta las métricas presentadas anteriormente al describir las métricas de interés para la comparación con modelos.

Finalmente en la sección de *Verificaciones* es posible contestar dos preguntas que se realizan frecuentemente en los análisis de redes: si la red es de Mundo Pequeño y si la red es de Libre Escala.

#### *Validación de Mundo Pequeño*

Visualmente, este análisis consta de tres paneles: dos de parámetros utilizados (uno por cada método de validación) y uno de la comparación de ambos métodos.

La diferencia de valores que aparece en la comparación de métodos está asociada al método de generación de las redes aleatorias (método 1 o 2).

#### *Validación de Libre Escala*

Al emplearse la función, la aplicación muestra el valor alfa correspondiente y los resultados de la prueba de Kolmogorov-Smirnov. Acorde a [Barabási \(2016\)](#), para que la red analizada sea de Libre Escala el valor alfa debe ser mayor a 2. Luego puede verificarse la hipótesis de ajuste de la distribución de grado a la distribución de ley de potencia mediante la prueba de Kolmogorov-Smirnov.

Por lo tanto, si el p-valor de la prueba de Kolmogorov-Smirnov es mayor a 0.05 y asimismo el alfa es mayor a 2, la red es considerada de Libre escala.

#### *Comunidades*

Para poder analizar comunidades dentro de la herramienta, primero debe seleccionarse el algoritmo con el cual se van a buscar dichas comunidades. Esta opción está disponible en el panel de configuración de la sección de Comunidades, así como la posibilidad de seleccionar cuántas comunidades se desean ver en la visualización de la red. Para más detalles de los algoritmos disponibles se recomienda ver el [Anexo: Detalle de algoritmos para búsqueda de comunidades](#).

Para complementar la visualización, se puede proceder con un análisis de la composición de los subgrafos resultantes de la búsqueda de comunidades, con ese objetivo se dispuso de la sección Comunidades. Como detalle se describe qué algoritmo fue utilizado para la búsqueda, la cantidad total de comunidades y, para evaluar la relevancia de los resultados de los algoritmos, está disponible la medida de modularidad ([Newman-Girvan 2004](#)). Una vez seleccionado el algoritmo se puede ver en el panel de comunidades el valor de esta medida teniendo en cuenta y sin tener en cuenta el peso de las relaciones. Los pesos de las relaciones en este caso están dados por la fuerza de colaboración de cada relación.

Además es posible realizar un análisis de la estructura de cada comunidad, en el que se puede ver qué autores forman parte de ella y los artículos asociados. Esto es similar a lo disponible para los componentes de la red, solo que esta vez el subgrafo es cada comunidad en vez de cada componente.

Asimismo, es posible obtener las mismas métricas que las descriptas para el análisis de componentes, solo que en este caso, los subgrafos son las comunidades en vez de los componentes.

Luego de seleccionar una comunidad, en la tabla de autores por comunidad se dispone de los detalles de la misma, entre los que se encuentran los artículos asociados. De dichos artículos se pueden ver algunos resultados del análisis de texto realizado sobre los títulos de los artículos, tales como: las palabras más frecuentes de esa comunidad, la cantidad de palabras frecuentes y los filtros aplicados de *stop words* en español e inglés (dado que existen algunos

títulos en dicho idioma). De manera similar a lo realizado para la sección de componentes, también se dispone del detalle de los nodos de la comunidad seleccionada.

## Sección 3: Análisis temporal

Esta sección permite realizar un análisis temporal de la red desde su primera publicación en 1996 hasta el periodo 2016. Cabe destacar que la red pasó por varias etapas diferentes para llegar a su estadio final, en este caso el año 2016. De esta manera, a medida que se avanza en los tiempos aparecen nuevos nodos y se modifican las relaciones. Las distintas redes de coautoría de la RAAB pueden ser analizadas entonces por etapa/año o bien de forma agregada, es decir, teniendo en cuenta lo acumulado en etapas anteriores hasta la etapa seleccionada.

Luego de tener una idea general de cómo se generaron las relaciones en el tiempo se pueden obtener diferentes medidas que sirven para analizar diversos aspectos de la red como por ejemplo qué nodos cumplen distintos roles. Dependiendo del objetivo, puede ponerse el foco en una o varias de estas medidas para contestar las preguntas de interés. La herramienta presenta dos enfoques de análisis de las medidas: de forma individual por periodo o de forma acumulada total. Esto nos permite analizar de forma general las relaciones de autoría dentro de la RAAB, de manera similar al trabajo realizado por [Koseoglu \(2016\)](#).

Finalmente la exploración de diferentes agregaciones temporales puede realizarse en la sección *Análisis estático* de periodos, mediante la selección de los periodos específicos de interés.

### Animación del grafo de manera acumulada

Esta herramienta muestra cómo se generaron las relaciones entre los nodos y cómo fueron conectándose los diversos componentes a lo largo del tiempo. Esto facilita focalizar en un periodo de interés para posteriormente realizar un análisis más detallado en la sección *Análisis estático*.

En la animación se pueden ver los nodos en escala de azules según su fuerza de colaboración y los colores de las aristas según la fuerza de relación entre los autores (medida en fuerza de colaboración). A la derecha de la animación hay una tabla para ver las equivalencias entre periodos y años.

En el detalle de los nodos de la animación puede verse quién es el autor, cuánta es la fuerza de colaboración acumulada, los periodos de participación y cantidad de artículos por periodo.

En el detalle de las aristas puede verse los autores participantes de dicha relación, la fuerza de colaboración resultante de las coautorías de los autores, los periodos en los cuales participan, y cuántos artículos contribuyeron en esos periodos.

### Medidas de cada periodo

En esta sección puede observarse la evolución anual de las métricas empleadas en el presente trabajo.

### Medidas Acumuladas a cada periodo

Esta sección contiene la visualización acumulada en el tiempo de las mismas métricas empleadas en las secciones anteriores.

Dado que en cada nueva edición de la revista se fueron incorporando más trabajos, esta sección facilita evaluar la interacción entre los viejos nodos y los nuevos nodos, cuándo aparecen los nuevos y qué tanto aporte realizan en el total de la revista.

## Ranking de autores en el tiempo

En esta sección es posible ver la evolución del ranking de autores, entre los primeros N (cantidad a seleccionar) por periodo, según alguna métrica seleccionada. Primero debe seleccionarse la métrica a evaluar y luego la cantidad de posiciones del ranking de autores (ej.: los 3 primeros, los 5 primeros, etc.).

Se obtiene como resultado un gráfico de un ranking de N autores por año, comparados según la métrica elegida (Fig. 3.2). Al analizar este gráfico se observa, por un lado, cómo evoluciona un puesto en el ranking siguiendo las líneas; cada puesto es representado por un color. Por otro lado, es posible enfocarse en analizar el comportamiento de un autor en el ranking. Cada punto representa una intervención del autor en un periodo. Al centrarse en una fila y observando cómo cambian de color los puntos puede visualizarse la evolución del autor. Cuando hay punto significa que ese autor estuvo en el ranking ese periodo, si no hay punto, el autor no participó en el ranking de ese periodo. Si el punto cambia de color, significa que cambió de posición en el ranking.

Cuando hay cambios en el ranking respecto de qué autor o autores están en una posición queda explícito en un cambio de pendiente en la línea. En caso de empates en un periodo se observan como una línea vertical.

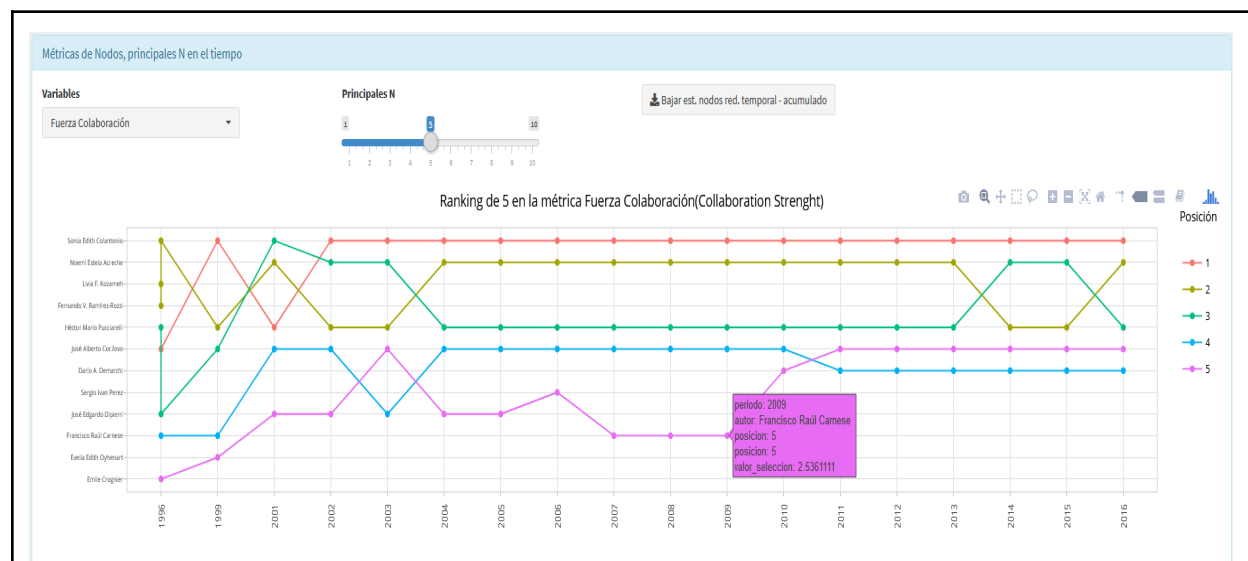


Fig. 3.2: Ranking de 5 autores según la métrica Fuerza de Colaboración. Se grafica la posición en el ranking. En el detalle interactivo se puede ver el valor de la variable elegida.

Analizando un puesto del ranking, vemos que el puesto 2 tuvo un empate entre 3 autores en el primer periodo (1996), luego cambió a otro autor que estuvo en el puesto 3 en el periodo anterior (1996), en el siguiente periodo (2001) lo tuvo una persona que no había participado en los primeros puestos del ranking en años anteriores. En 2002 y 2003, volvió a estar en puesto 2 el autor que estuvo en puesto 2 en 1999. En 2004, el puesto volvió a la persona que estuvo en 2001 y se mantuvo estable hasta 2013. En 2014 cambió nuevamente y se mantuvo en 2015, finalmente en 2016 volvió a cambiar.

Analizando un autor, por ejemplo la primera fila, vemos que empezó en el puesto 2 en 1996, en 1999 pasó al puesto 1. En 2001 pasó al puesto 3, luego desde 2002 hasta 2016 se mantuvo en puesto 1.

Otro ejemplo de autor, puede ser la última línea, que estuvo en puesto 5 en el primer periodo y luego no participó más del ranking de 5 para fuerza de colaboración.



## Ranking de autores en el tiempo para medidas acumuladas

Esta sección es similar a la anterior, con la diferencia de que se visualiza el color de la métrica acumulada en el tiempo. Al igual que en el caso anterior, es posible elegir una métrica de interés y la cantidad de autores a visualizar (N). En caso de empate aparecen todos los autores que estén entre los primeros N.

## Análisis de la red de coautorías de la RAAB

### Hipótesis 1: La cantidad de autores y artículos aumenta en el tiempo

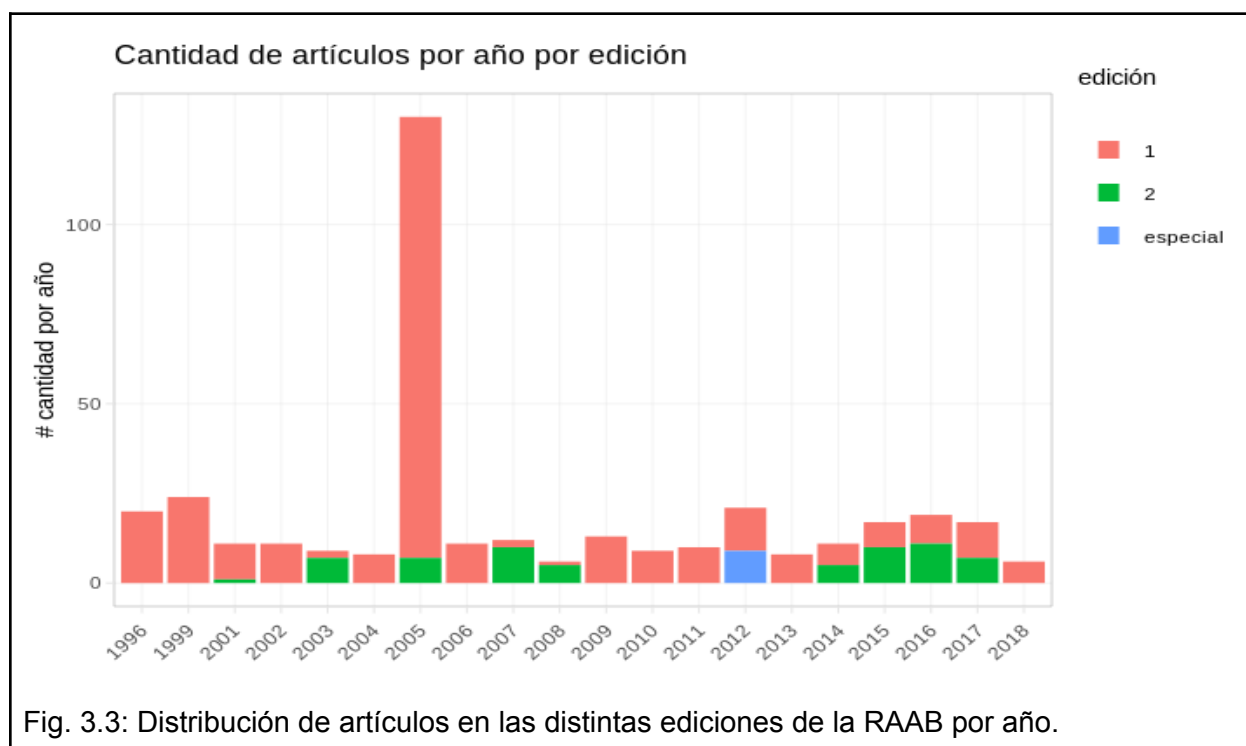
El cambio en el tiempo de la cantidad de autores y artículos puede brindar diferente información según si se analiza en forma acumulada o individualizando cada instante temporal. El análisis en cada instante temporal puede mostrar una mayor cantidad de autores según potenciales eventos en la existencia de la revista, como puede ser su inicio o congresos. Se estima que estos eventos afectan la producción de trabajos originales, haciendo crecer o reduciendo la cantidad de trabajos originales, y por ende, la cantidad de autores participantes en la red de coautoría de trabajos originales de la RAAB. Este punto de vista nos ayuda a comprender mejor cómo se comportan las dinámicas asociadas a los eventos de mayor o menor producción. Si tomamos las cantidades acumuladas de autores, estimamos que crecen en el tiempo en caso de crecer la revista, o mantenerse sin incrementos en caso de no crecer. Crecer implica que nuevos autores aparecen en la red, que estos pueden conformar nuevos equipos, o tal vez se integran a equipos existentes. Al revisar la cantidad acumulada resultan de interés los saltos que hay entre edición y edición viendo la variación que hay entre cada momento. Mayor variación entre periodos significa que nuevos autores participan de la red, menor variación significa que la red se mantiene por los autores existentes.

La cantidad de publicaciones por año en la RAAB se observa en la [Tabla 3.1](#) y las Figuras [3.3](#) y [3.4](#).

Tabla 3.1: Cantidad de publicaciones por año, discriminado en secciones y ediciones.\* Al momento de la confección de la tabla, el año 2018 no estaba completo.

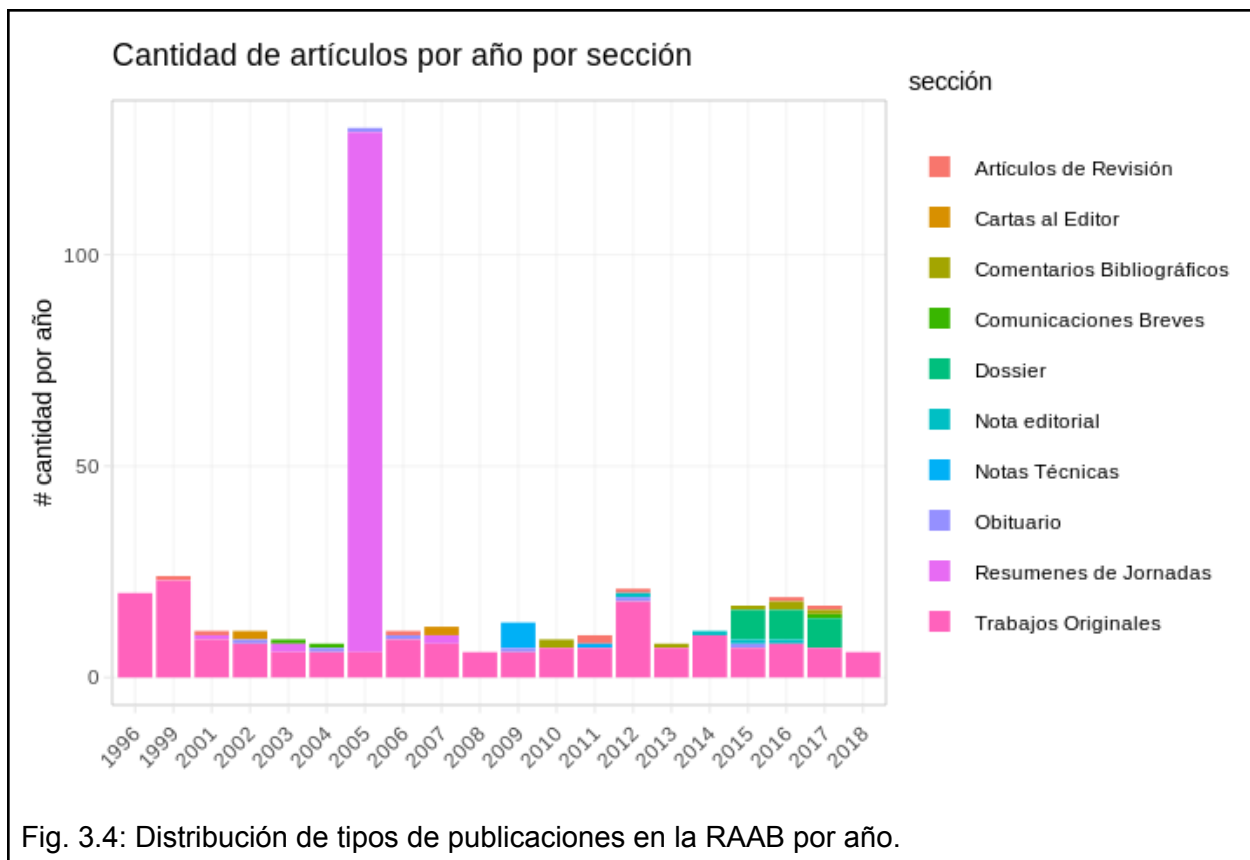
año	1996	1999	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018 *
total	20	24	11	11	9	8	130	11	12	6	13	9	10	21	8	11	17	19	17	6
sección	Artículos de Revisión	0	1	1	0	0	0	0	1	0	0	0	0	2	1	0	0	1	1	0
	Cartas al Editor	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
	Comentarios Bibliográficos	0	0	0	0	0	0	0	0	0	0	2	0	0	1	0	1	2	1	0
	Comunicaciones Breves	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0
	Dossier	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	7	7	0
	Nota editorial	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0
	Notas Técnicas	0	0	0	0	0	0	0	0	0	6	0	1	0	0	0	0	0	0	0
	Obituario	0	0	0	1	0	1	1	1	0	0	1	0	0	1	0	0	1	0	0
	Resúmenes de Jornadas	0	0	1	0	2	0	123	0	2	0	0	0	0	0	0	0	0	0	0
	Trabajos Originales	20	23	9	8	6	6	6	9	8	6	6	7	7	18	7	10	7	8	6
edición	1	20	24	10	11	2	8	123	11	2	1	13	9	10	12	8	6	7	8	6
	2	0	0	1	0	7	0	7	0	10	5	0	0	0	0	5	10	11	7	0
	especial	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0

Al observar la cantidad total de publicaciones en la RAAB por año y discriminada en ediciones (primera, segunda y especial) ([Fig. 3.3](#)) pudo verse una gran disparidad para el año 2005.

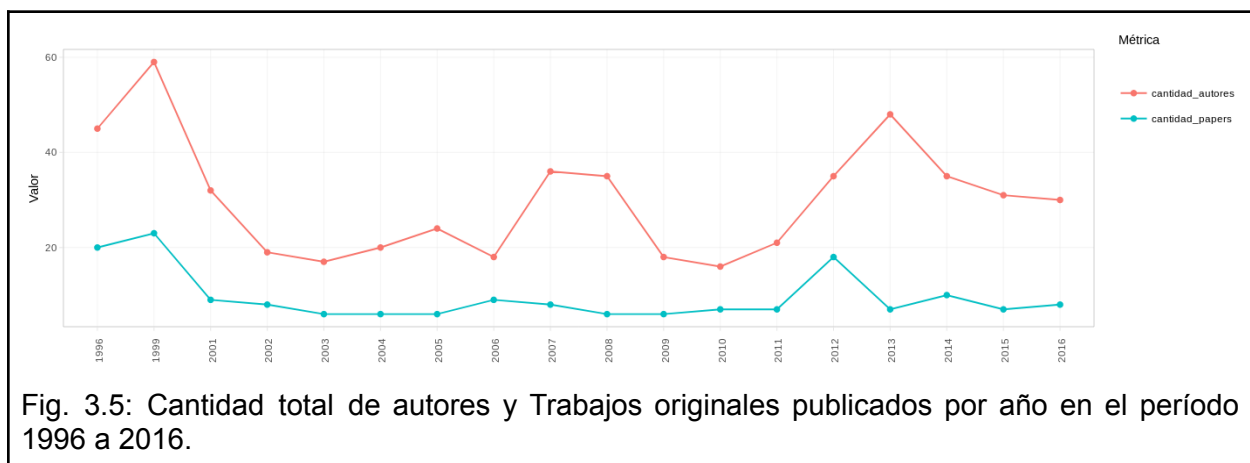


Cuando se evaluó la composición por secciones (es decir, tipo de publicación) se observó que esta diferencia en la cantidad total de publicaciones en el año 2005 se debe a la inclusión de los resúmenes de los trabajos presentados en las Jornadas Nacionales de Antropología Biológica ([Fig. 3.4](#)).





Esta contextualización inicial puso en evidencia la complejidad de incorporar distintos tipos de publicaciones al análisis. Por lo tanto, y retomando lo planteado originalmente, en esta tesis se analizó la cantidad de autores y artículos de tipo Trabajo original, en el período de tiempo acotado entre 1996 y 2016 (Fig. 3.5).

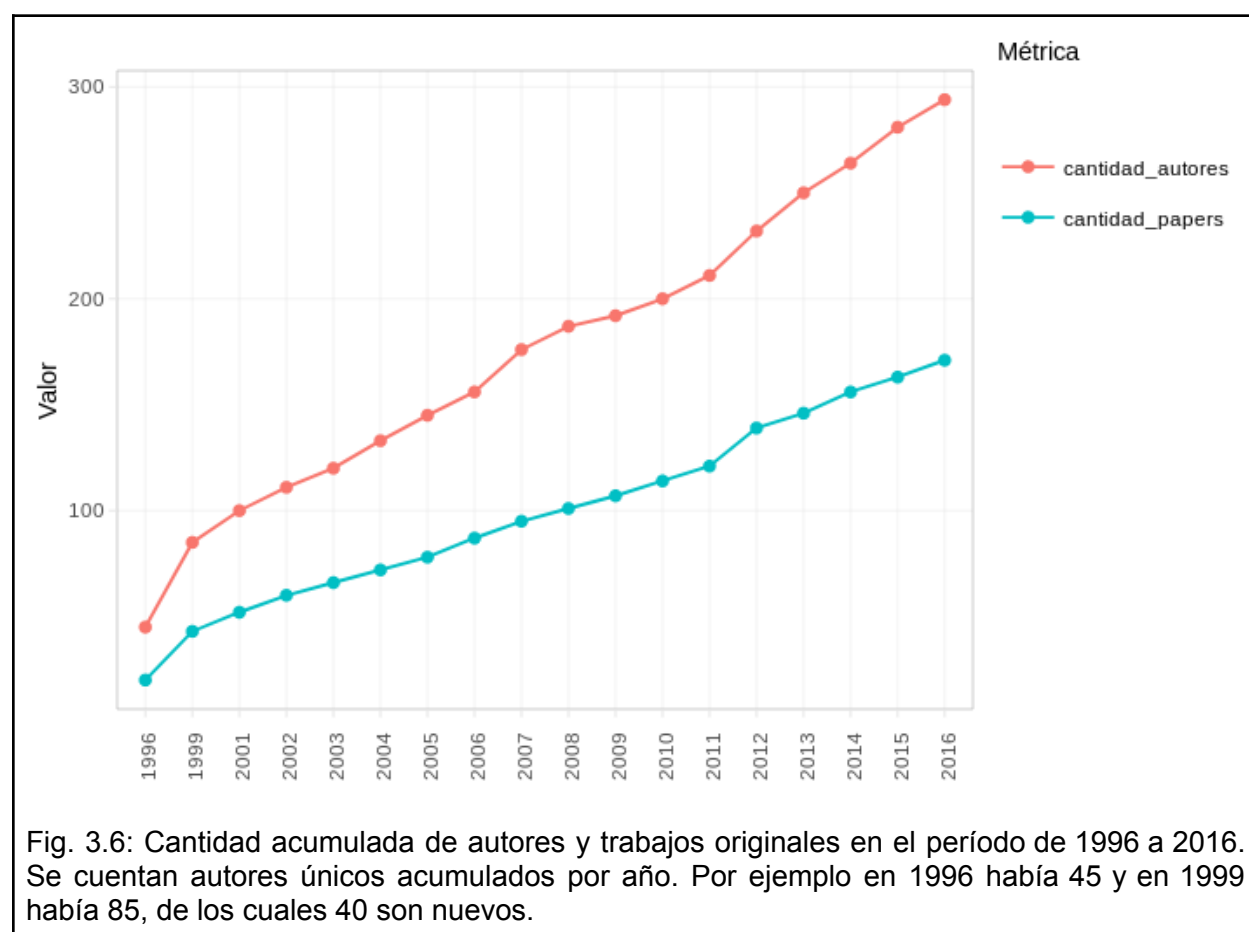


Se observa que en la etapa inicial la RAAB tuvo un fuerte impulso, tanto en cantidad de autores como artículos. Sin embargo, el patrón general no es el mismo para ambas medidas, lo cual amerita su análisis por separado.

Por un lado, con respecto al número total de Trabajos originales, se observó que a partir del 2001 se alcanzó un valor que corresponde a la media para la RAAB, y que se mantuvo sin grandes variaciones en los años que siguieron.

En el caso de los autores, descartando el impulso inicial de los dos primeros años, parece haber un pequeño aumento en líneas generales, pero con la existencia de picos en determinados años, no necesariamente asociados a un incremento en el número de publicaciones. Por ejemplo, en 2013 hubo un fuerte aumento en la cantidad de autores, quedando como segundo año con mayor cantidad de autores para trabajos originales, luego del impulso inicial. Esto podría estar asociado al incremento en la cantidad de artículos originales publicados en 2012, que corresponde al tercer lugar en la línea de vida de la RAAB. Sin embargo, otro aumento en la cantidad de autores se observó en los años 2007 y 2008 pero sin asociarse a un incremento marcado en el número de publicaciones.

El análisis de cómo se comportan los valores acumulados de la cantidad de autores y artículos a lo largo del tiempo se muestra en la [Figura 3.6](#).



Pueden observarse las diferencias mencionadas anteriormente acerca del ritmo de crecimiento en la cantidad de autores y de artículos. En particular, para la cantidad de autores existen ciertos ciclos de cambios en la pendiente de crecimiento acumulado asociados a los picos señalados para los años 1999 (impulso inicial), 2007 y 2012-2013. En todos los casos, estos incrementos en la pendiente parecen sucederse por períodos de aumento relativamente

constante. Se observa, además, que los aumentos en la cantidad de autores son mayores y más variables que en el número de artículos, cuyo incremento acumulado se mantiene relativamente constante por períodos mayores de tiempo. En función de esto, es posible plantear que estas diferencias podrían deberse a una mayor cantidad de artículos trabajados en conjunto.

A pesar de estas tendencias que se observaron inicialmente, 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 ( $K=-0.013$ ,  $p=0.93$ ) ni para la cantidad de publicaciones originales ( $K=-0.013$ ,  $p=0.69$ ) a lo largo de los años.

### Cantidad de autores en artículos por año

En la [Figura 3.7](#) (y [Tabla A.14](#) del [Anexo: Cantidad de autores en artículos por año](#)) se observa que en general predominan los trabajos originales que son realizados en equipo, entre 2 y 7 personas. Sin embargo, en diferentes momentos varían los patrones de cantidad de autores y cantidad de artículos. y estas relaciones no parecen ser lineales, puesto que existen momentos donde se publican más artículos con múltiples autores y momentos donde la faceta individual prevalece. A pesar de esto, puede decirse que existe una tendencia a que, con el pasar del tiempo, se reduzcan los artículos individuales y se fortalezcan los artículos en equipo. Se observa, por ejemplo, la desaparición de artículos con un solo autor en los últimos dos años analizados. Además, se observa la aparición de artículos con gran cantidad de autores (más de 8) a partir de la segunda mitad del período total analizado.

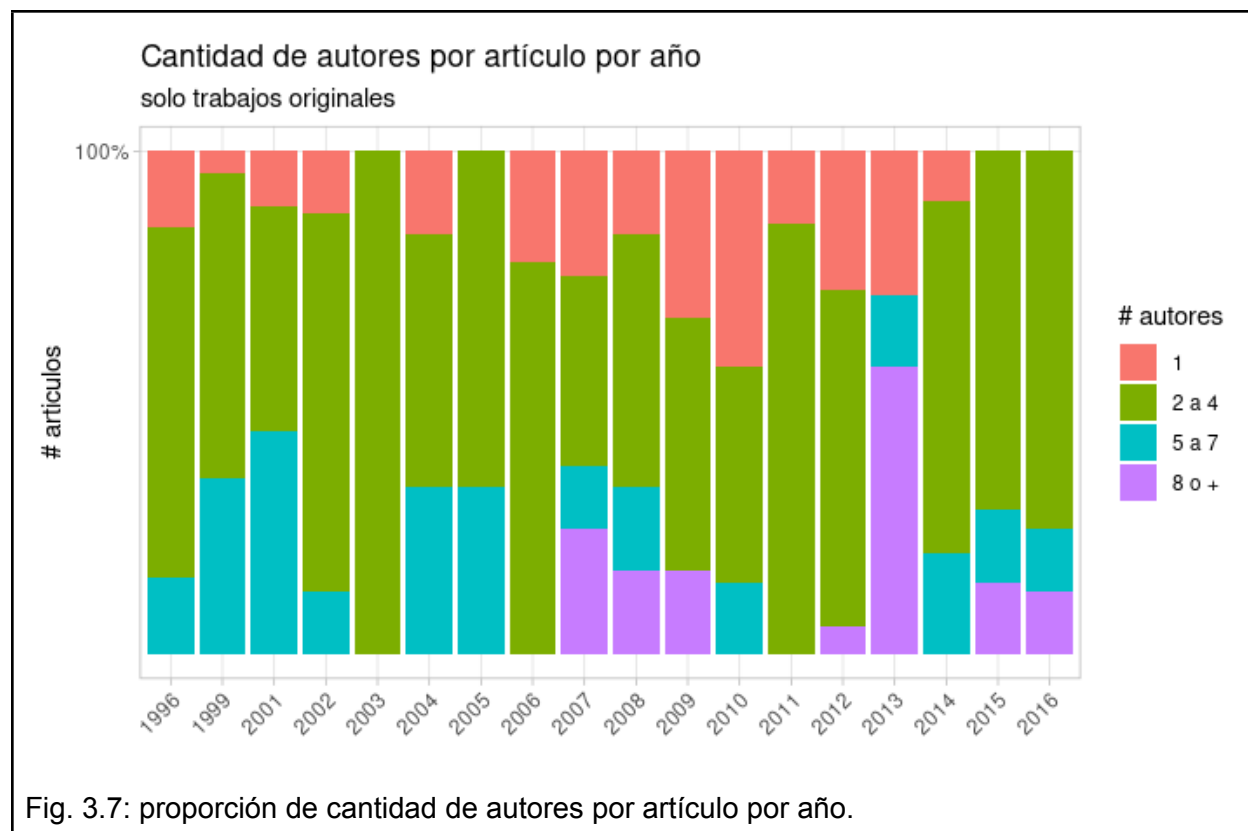
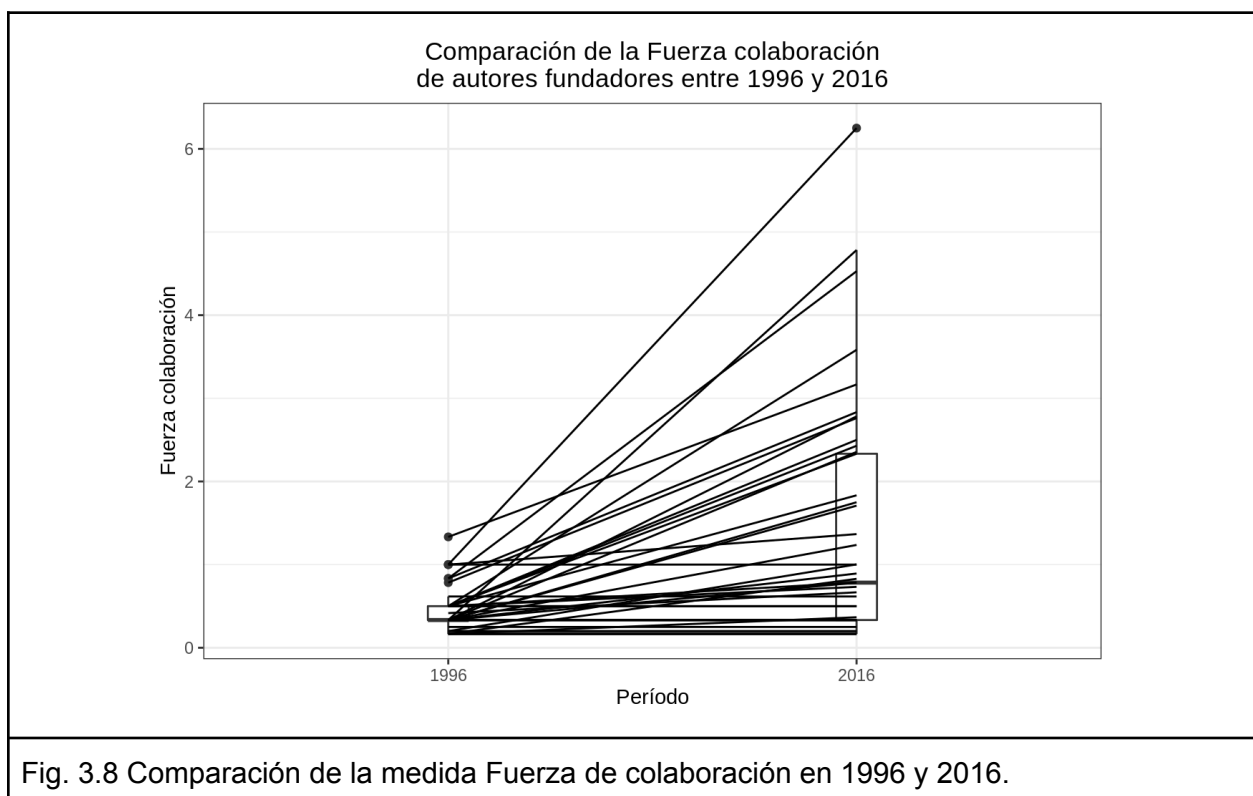


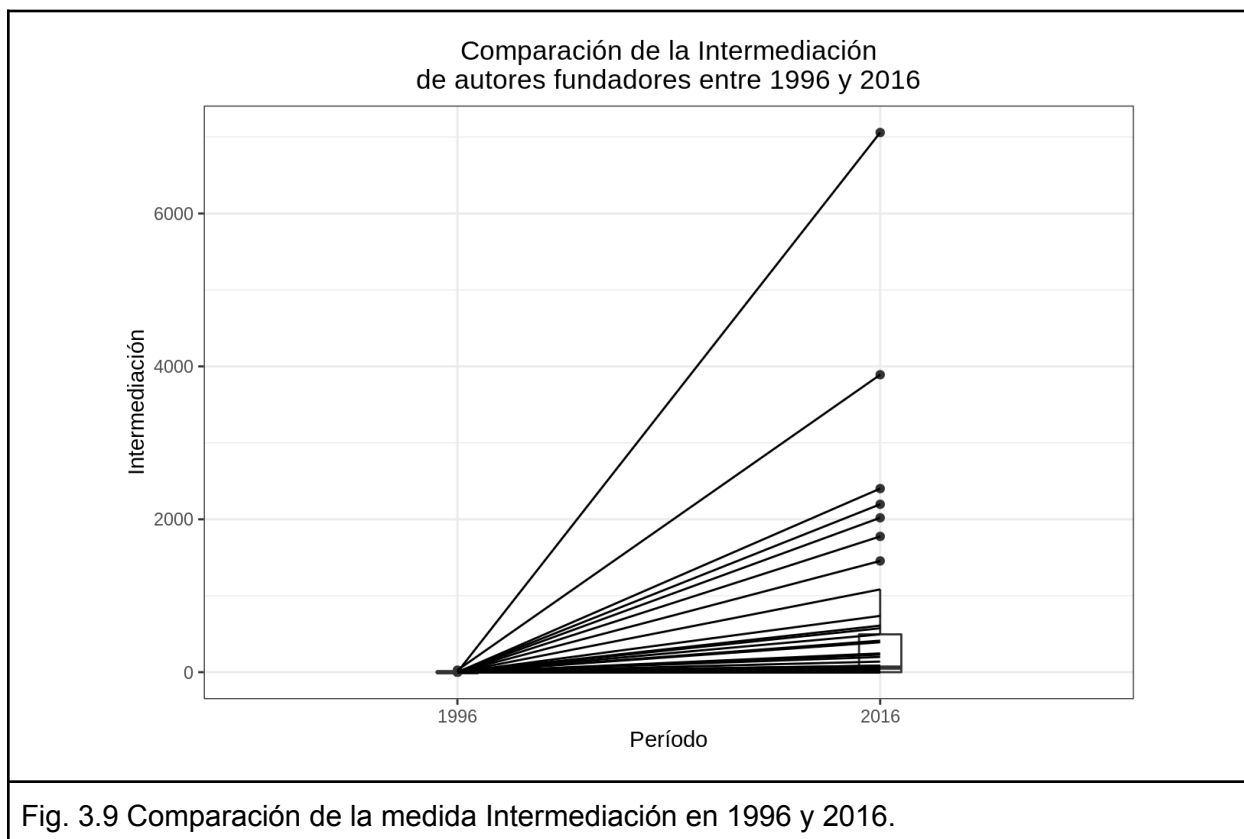
Fig. 3.7: proporción de cantidad de autores por artículo por año.

Hipótesis 2: Los autores fundadores se refuerzan en el tiempo, lo cual se relaciona con un aumento en alguna de las métricas (Grado, Intermediación o Fuerza de colaboración).

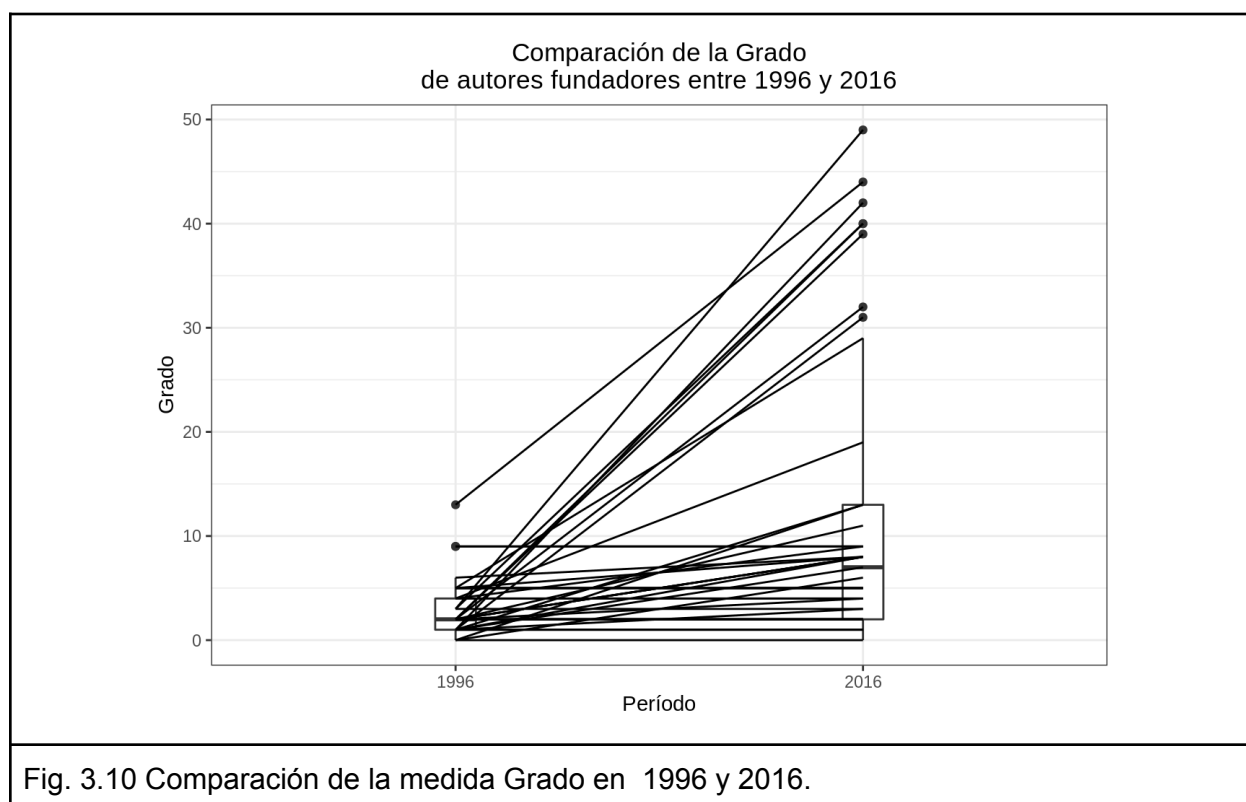
Analizando la fuerza de colaboración en el total de autores fundadores, empleando la prueba de Rangos con Signo de Wilcoxon para muestras pareadas, se observa que existe evidencia estadística para afirmar que la fuerza de colaboración aumentó comparando el año 1996 con el 2016 ( $V=0$ ,  $p < 0.001$ ,  $n = 45$ ), lo que también se evidencia en el gráfico de cajas para datos pareados ([Fig. 3.8](#)).



Análogamente con el caso anterior, para la Intermediación en el total de autores fundadores, también se observó que existe evidencia estadística para afirmar que esta métrica aumentó comparando el año 1996 con el 2016 ( $V=0$ ,  $p < 0.001$ ,  $n = 45$ ) ([Fig. 3.9](#)).



Por último, se aplicó el mismo procedimiento, pero esta vez sobre el Grado, y se observó un aumento significativo comparando el año 1996 con el 2016 ( $V=0$ ,  $p < 0.001$ ,  $n = 45$ ). Nuevamente, las diferencias observadas se muestran en la [Figura 3.10](#).

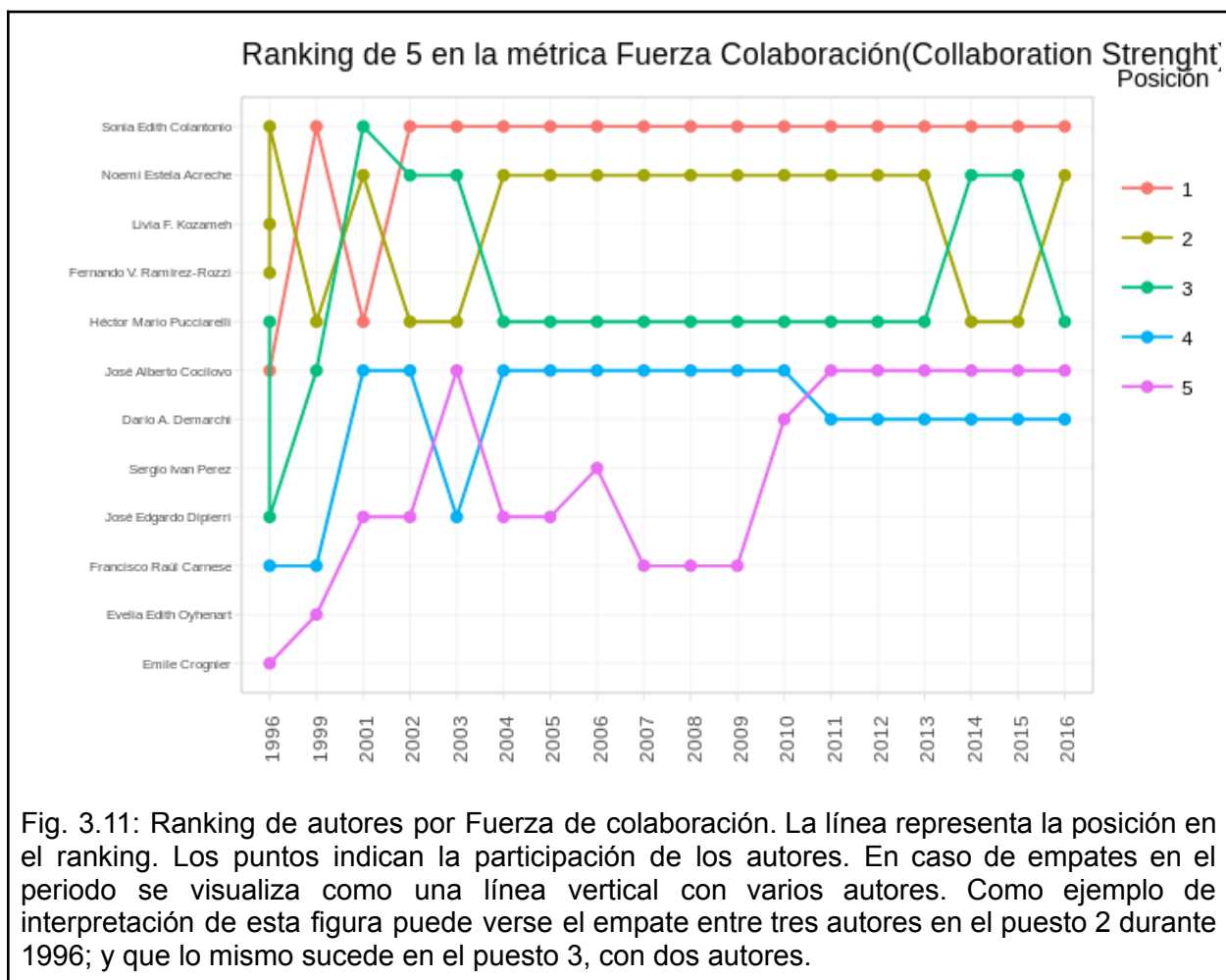


### Ranking de 5 autores por año y su evolución temporal

La clasificación de los autores en el ranking de cinco puestos según cada una de las métricas se observa en las Figuras [3.11](#), [3.12](#) y [3.13](#) (Los valores numéricos utilizados se encuentran en el [Anexo: Tablas numéricas complementarias para la Hipótesis 2](#)). Para interpretar correctamente las figuras es necesario comprender que las líneas de colores están asociadas a la posición en el ranking, esto ayuda a visualizar si los autores se mantienen o varían en el tiempo (eje X). Cuando vemos líneas rectas horizontales significa que ese autor se mantiene por un tiempo en la misma posición del ranking. Mientras que la ausencia de las mismas significa que hay mucha variación y que los autores no se mantienen estables en una misma posición a lo largo del tiempo. Por otro lado, las líneas verticales indican que se comparte el puesto del ranking entre varios autores.

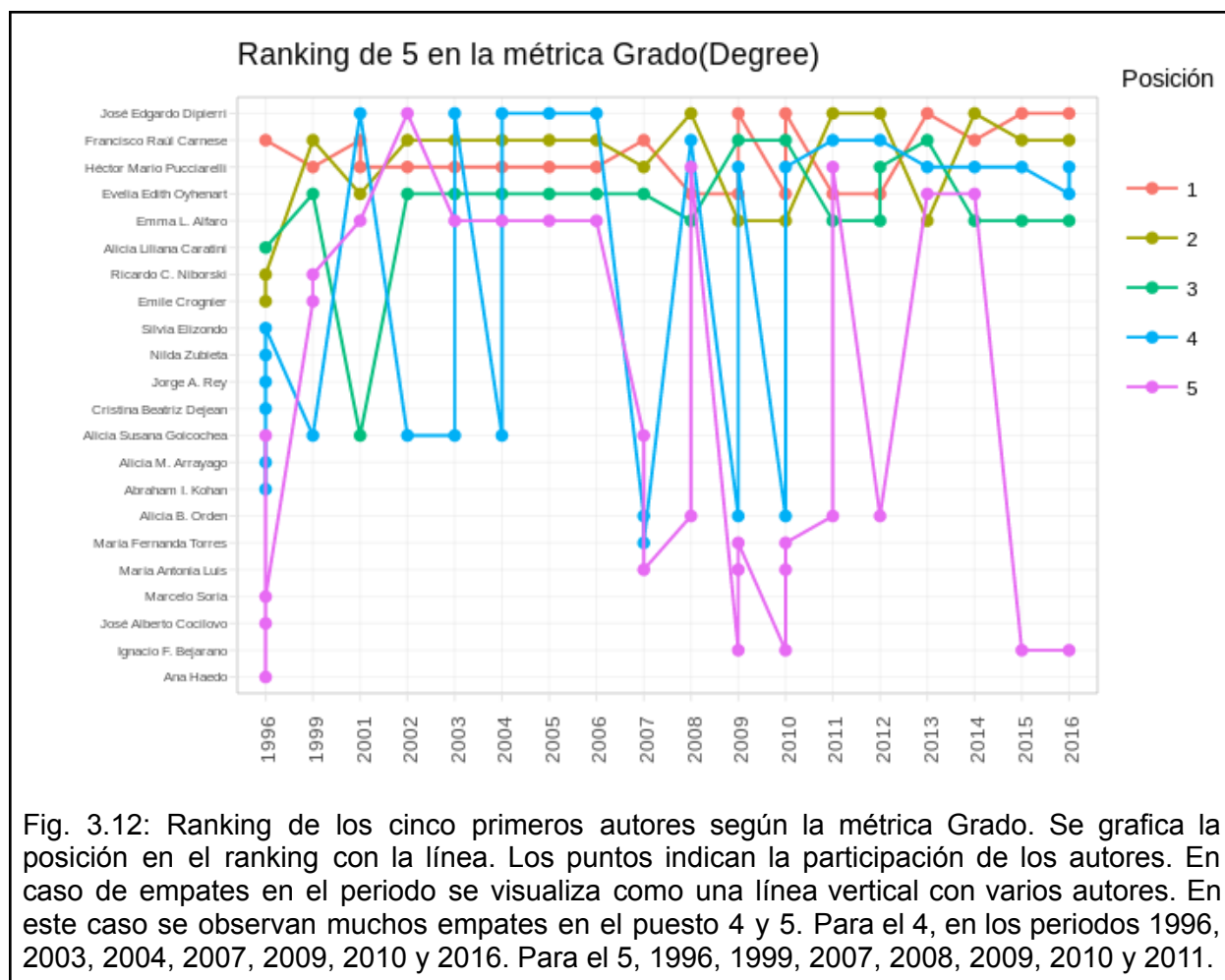
En el caso de la Fuerza de colaboración cabe recordar que está definida como artículo/cantidad de autores. Esto significa que el ranking de la métrica representa quienes mejor están posicionados respecto a producción de contenido de trabajos originales. Si en un año se ven dos puntos en la figura, significa que esos autores “empataron” en cuanto a Fuerza de colaboración.

Se observa que los autores fundadores compiten por sus posiciones en los primeros tres puestos y se refuerzan en el tiempo ([Fig. 3.11](#)). Los puestos 4 y 5 son más variables, pero a partir de 2011 se mantienen. Por lo tanto, puede decirse que la generación de contenido se mantiene bastante estable a partir del 2002.



Como análisis complementario quedaría pendiente ver la variación interanual para evaluar cuánto crecen estos valores año a año.

En la [Fig. 3.12](#) se observa la métrica Grado que indica cuánto se conecta directamente un determinado nodo con otros nodos. En el contexto de red de coautoría hace referencia a con cuántos otros autores publicó ese año. Cabe recordar que el Grado es la cantidad de nuevos nexos que genera el autor (son nuevos porque no estamos considerando fortalecimiento de una relación por hacer más trabajos en conjunto, como indica la Fuerza de Colaboración).

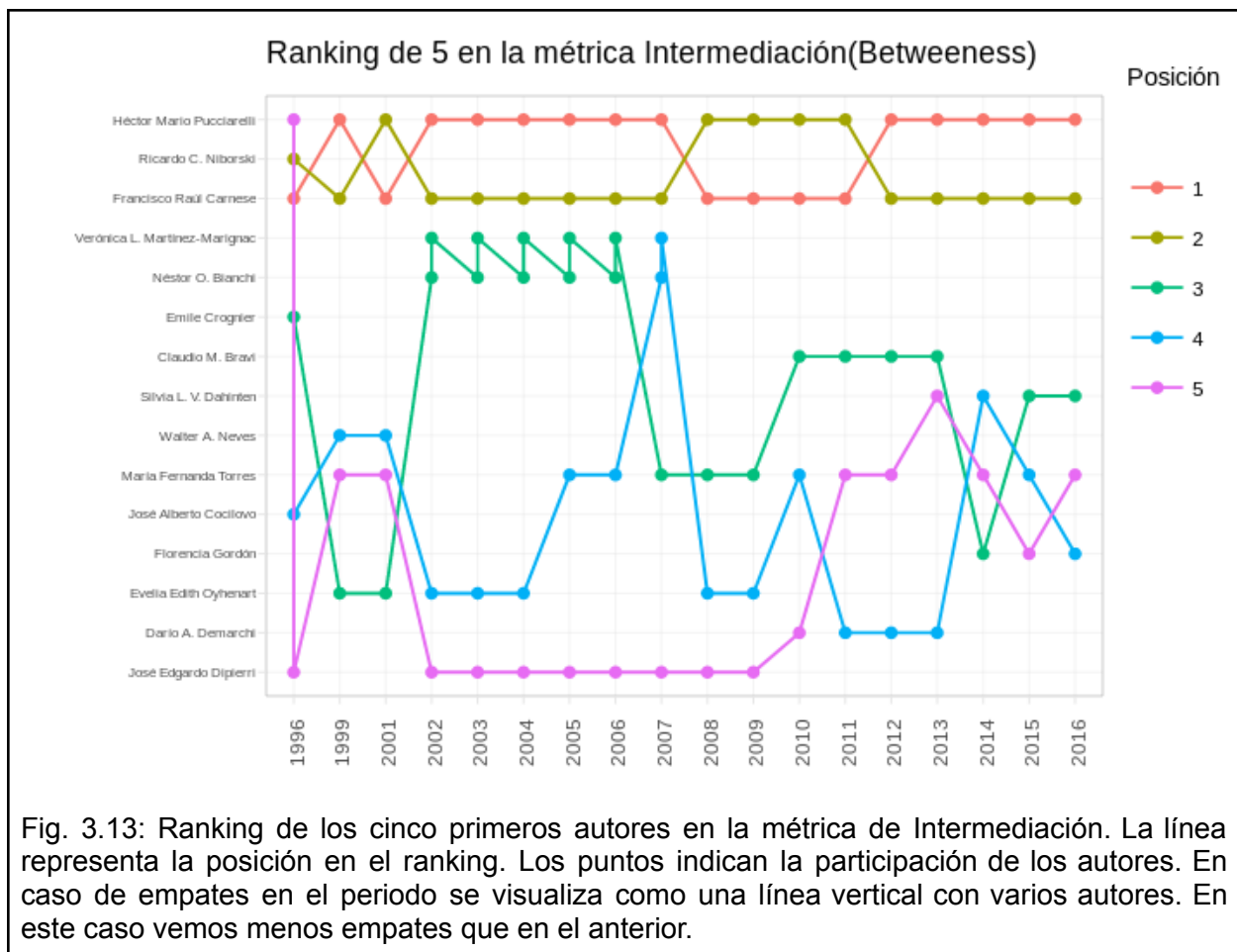


Se observa que este ranking es competitivo, anualmente existen más empates que en el de Fuerza de colaboración ([Fig. 3.11](#)). Con respecto a los autores fundadores, se observa que para los primeros tres puestos compiten bastante, aunque entre 2002 y 2006 se mantuvieron estables, mientras que los puestos 4 y 5 presentan mayor variación ([Fig. 3.12](#)).

En cuanto a la evolución temporal, el primer nodo, por ejemplo, aparece en 2001 en el puesto 4, luego en 2002 baja al puesto 5, en 2003 vuelve al 4 y se mantiene hasta 2006. En 2007, no aparece. Luego en 2008 aparece con más fuerza, saltando al puesto 2, en 2009 y 2010 escaló hasta el primer puesto y en los años siguientes, se mantiene entre el puesto 1 y el 2. Esto significa que, en el contexto de la revista analizada, dicho autor es un gran generador de colaboraciones e integrador de nuevos nodos a la red.

Finalmente, los resultados obtenidos para la Intermediación se muestran en la [Figura 3.13](#). Esta métrica mide la importancia de un autor según su participación en caminos para llegar de un autor a otro.



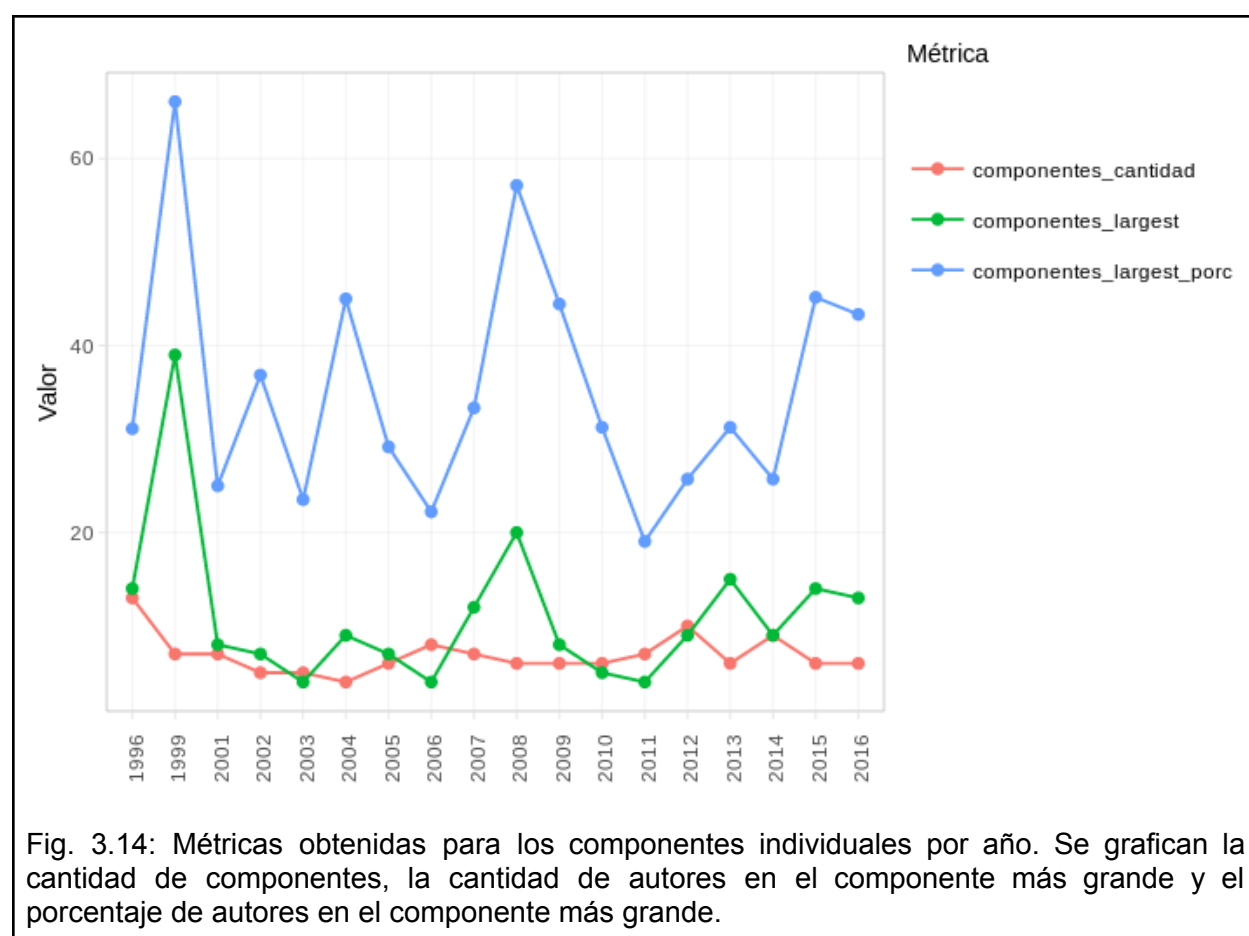


Se observa que hay más empates para esta métrica que en comparación a Fuerza de Colaboración, pero menos que en el caso de Grado. Los primeros dos puestos se mantienen bastante estables sin embargo los puestos 3, 4, y 5 varían bastante a lo largo del tiempo. Con respecto al puesto 3, dos autores lo comparten desde 2002 a 2006; el puesto 4, tiene un empate en 2007, que son los mismos autores que venían en el puesto 3. El puesto 5 tiene un empate en el primer periodo y se mantiene estable durante el periodo de 2002 a 2009. Finalmente, resulta interesante cómo a partir de 1999 el puesto 1 y 2 va oscilando entre dos autores (Fig. 3.13). Respecto de los autores fundadores, se mantienen los primeros 2 puestos mientras que el resto cambia.

En líneas generales se observó mucha variación entre los autores que figuran en el ranking según la métrica considerada. En los ranking que involucran Grado e Intermediación (Figs. 3.12 y 3.13) se comparten más autores, sin embargo en Fuerza de Colaboración varían (Fig. 3.11). Esto podría interpretarse como que diferentes autores han cumplido diferentes funciones en la RAAB dentro del periodo analizado.

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

La evolución temporal de las tres métricas analizadas para los componentes se observa en la [Figura 3.14](#). Cabe recordar que un componente es un grupo de nodos que están conectados entre sí y está desconectado de otros grupos y que los autores individuales son componentes en sí mismos. Cuando existe más de un componente se califica a la red total como una red inconexa, lo cual es natural en redes de coautoría dado que dentro de una comunidad científica, existen autores individuales y grupos que todavía no realizaron colaboraciones a través de ninguno de sus autores, o que publican en otros lugares además de la revista analizada.

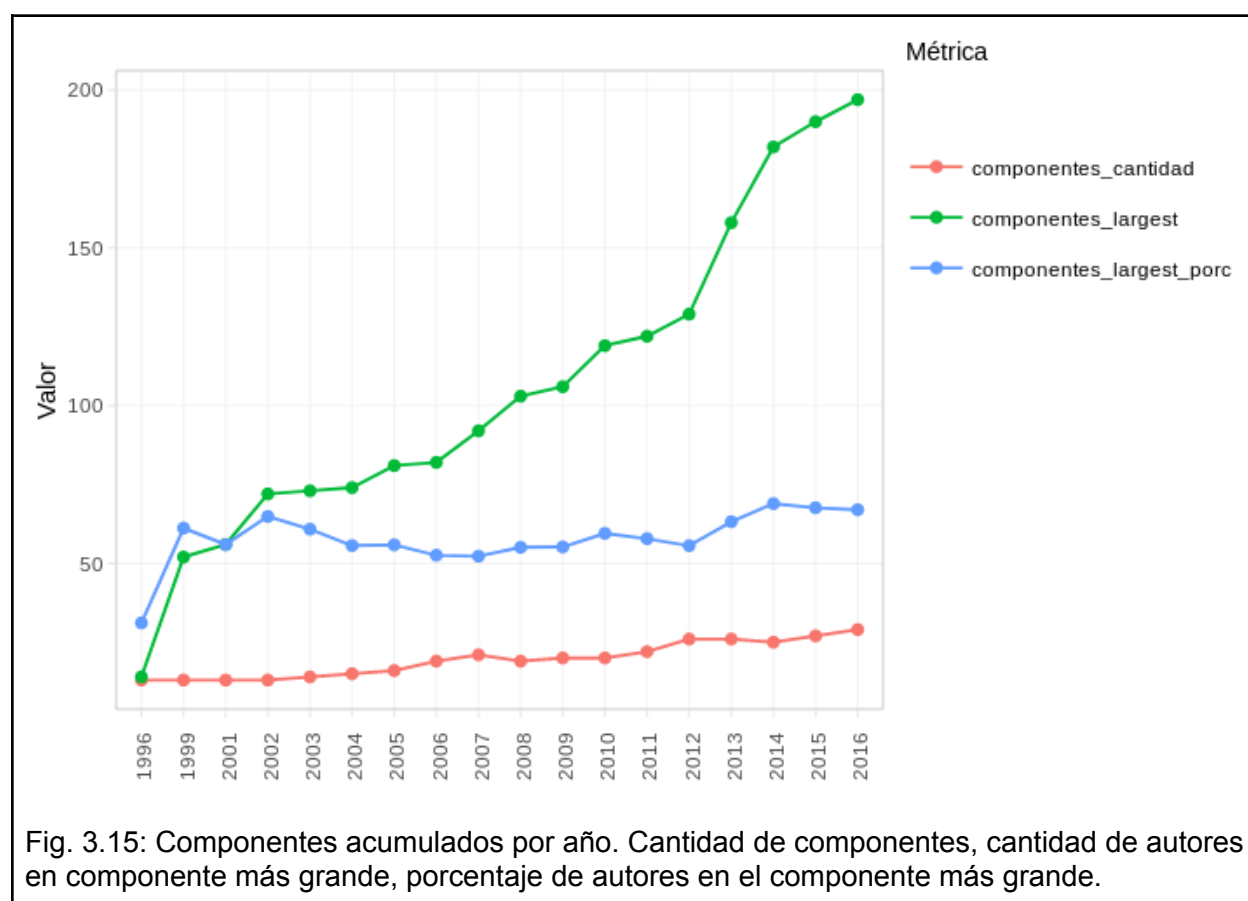


Desde un punto de vista anual ([Fig. 3.14](#)), se observa que en 1999 la RAAB tuvo su pico de participación de autores involucrados en el componente principal para una misma edición. Observando ese año en el gráfico de autores por artículo ([Fig. 3.7](#)), vemos que es el año con

mayor cantidad de trabajos originales presentados, de estos su mayoría corresponde a artículos entre 2 y 7 autores.

Por otro lado, se observa un segundo pico en relación al componente más grande en 2008 (Fig. 3.14). Si vemos el detalle de composición mencionado en la Hipótesis 1, este fenómeno puede interpretarse como resultado de un crecimiento en la cantidad de autores y cantidad de artículos, que podría estar relacionado con que en ese año se publicó un número especial de la revista. Se observa que éste fue el tercer número de mayor publicación en la historia de la RAAB, que no contó esa cantidad de artículos desde los números iniciales del 1996 y 1999. En cuanto a cantidad de autores, la de ese año es similar a la de los primeros números, con la diferencia de que existe un artículo de 8 o más personas, hubo más artículos individuales, y no se publicaron artículos con entre 5 y 7 personas participantes (Fig. 3.7).

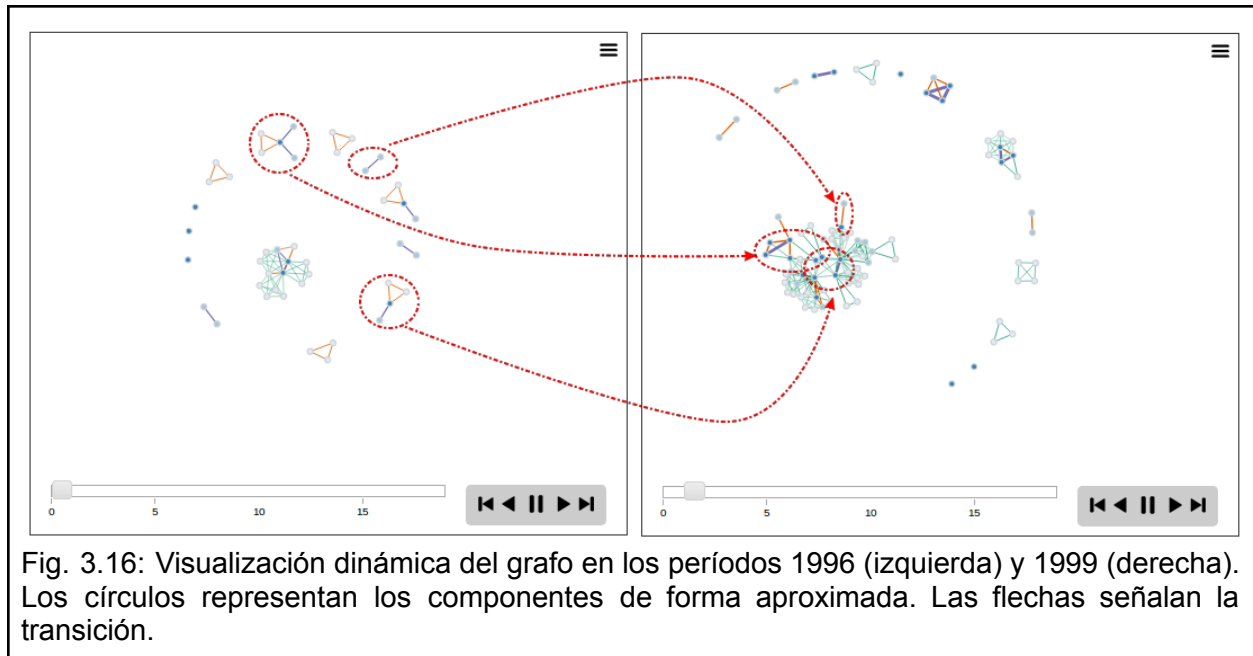
El análisis de las métricas en componentes acumulados por año (Fig. 3.15) facilita la observación de esos “saltos” en la cantidad de autores en el componente mayor. Estos saltos implican que un componente grande fue absorbido por el componente mayor. Lo podemos ver en las transiciones de 1996 a 1999, de 2012 a 2013 y de 2013 a 2014.



Teniendo en cuenta las observaciones de 2012, se observa que alguno de los componentes originados fue absorbido en el 2013 al realizar trabajos de colaboración. A continuación se muestran los resultados obtenidos tras verificar estas supuestas absorciones a través de revisar las transiciones estructurales de los grafos de forma visual utilizando la funcionalidad disponible herramienta desarrollada.

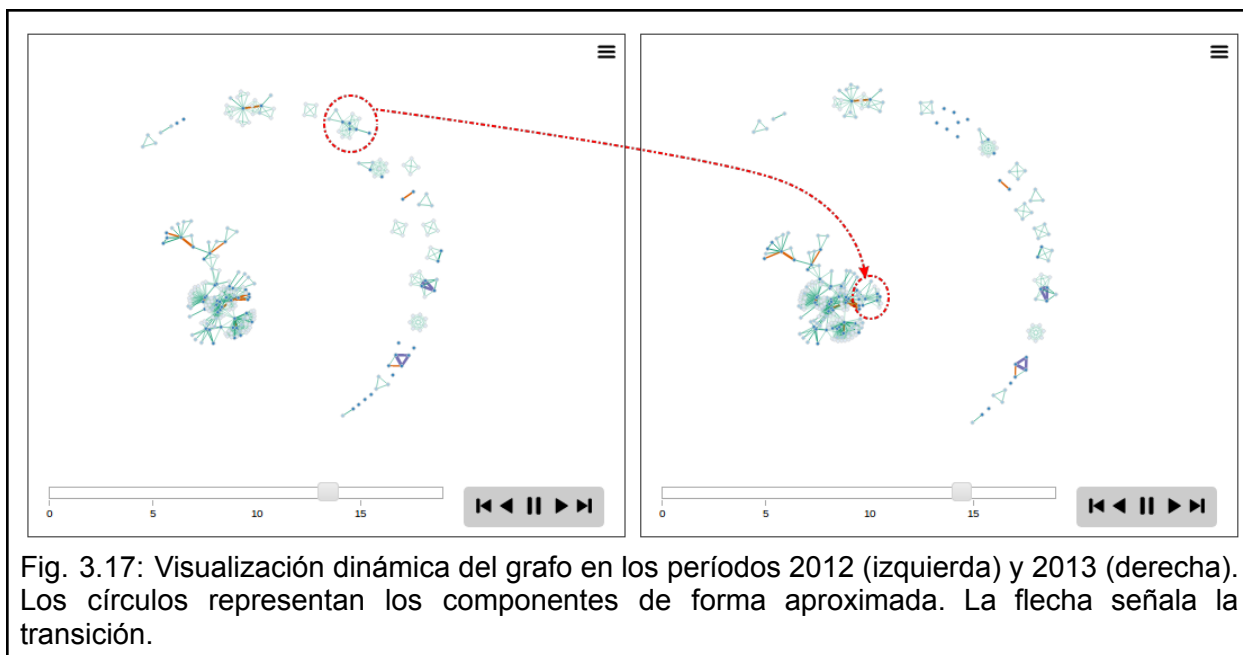
### Verificación de absorción de componentes en la red

La [Figura 3.16](#) muestra un ejemplo de verificación de absorción de componentes entre los años 1996 y 1999. En la aplicación pueden verse de forma dinámica estas transiciones. Cabe recordar que se entiende por absorbidos a aquellos nodos que ya existían en la red y que pasan a formar parte del componente mayor.

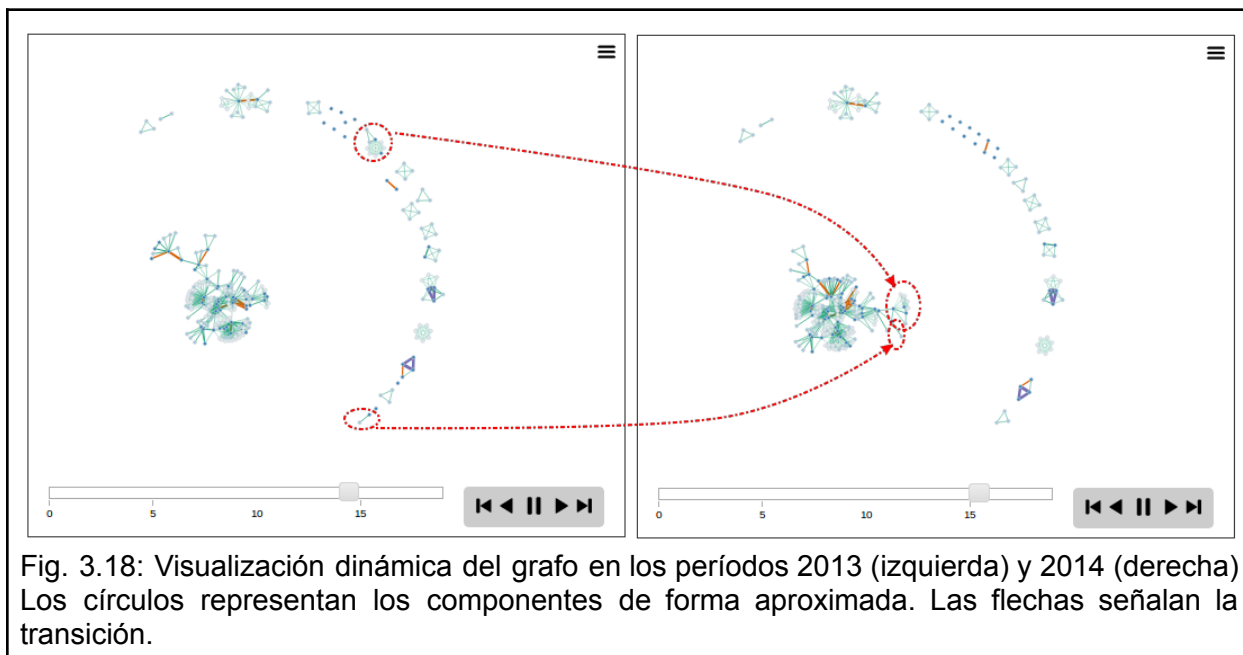


Se observa que en la primera transición el componente mayor absorbe dos componentes muy importantes del tiempo anterior de cinco y cuatro autores, y un tercer componente de dos nodos. Además aparecen nuevos nodos en el nuevo año haciendo que pase de 14 a 52 autores en 1999 (crecimiento de 38 autores).

Para el período siguiente ([Fig. 3.17](#)) se observa que uno de los componentes más grandes (12 autores) es absorbido por el componente mayor. De forma complementaria se agregan directamente nuevos nodos al componente mayor pasando de 129 a 158 autores en 2013 representando un crecimiento de 29 autores.



Finalmente, en el período de 2013 a 2014 ([Fig. 3.18](#)), se observa que uno de los componentes más grandes (9 autores) es absorbido por el componente mayor. De forma complementaria también absorbe un componente menor de dos autores y además se agregan de manera directa nuevos nodos al componente mayor pasando de 158 a 182 en 2014, lo cual representa un crecimiento de 24 autores.



La verificación numérica de lo observado anteriormente en los grafos se encuentra en la [Tabla 3.3](#).

Tabla 3.3: Cantidades de nodos Autores en el componente mayor en los tres períodos analizados.

Periodo ( t0 a t1)	Cantidad total t1	Cantidad componente mayor t0	Cantidad componente mayor t1	Representa % Red (t1)	Nuevos nodos	Nodos absorbidos	Total nuevos nodos en t1
1996 a 1999	85	14	52	61.18	27	11	38
2012 a 2013	250	129	158	63.2	17	12	29
2013 a 2014	264	158	182	68.9	13	11	24

De esta forma, 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. Asimismo se observó que de los nuevos nodos que se agregan a este componente mayor, una parte corresponde a nuevos autores publicando en la revista, además de los nodos ya existentes.

#### Relación con actividades de investigación

Los resultados del análisis de minería de texto realizado sobre los títulos de los artículos publicados por los autores pertenecientes a cada componente para la red total se presentan en la [Tabla 3.4](#).

Tabla 3.4: Artículos, autores y palabras de mayor frecuencia en los títulos de los artículos correspondientes a los primeros cinco componentes de la red total (período 1996-2016).

componente	Cantidad autores	Cantidad artículos	Primeras 10 palabras por frecuencia en el grupo
1	197	114	Argentina, poblaciones, análisis, provincia, cordoba, población, jujuy, influencia, populations, craneofacial
2	21	10	Tierra, fuego, grande, regional, fagnano, río, argentina, resultados, isótopos, estables
3	9	11	Salta, argentina, genética, poblacional, tamaño, localidades, homogamia, genéticas, calchaquí, demografía
4	7	1	Rio, province, population, racial, polymorphisms, pinar, genetic, cuba
5	5	1	Risk, riesgo, grupos, etnicos, ethnic, congenitas, congenital, anomalies, anomalias, america

Se observa que el componente principal tiene 114 artículos, mientras que el segundo tiene sólo 10 artículos y el tercero 11. Además, los dos últimos componentes están representados solamente por un artículo cada uno en la red ([Tabla 3.4](#)). Por este motivo, a continuación se presentan únicamente los resultados de frecuencias de palabras para los primeros tres componentes.

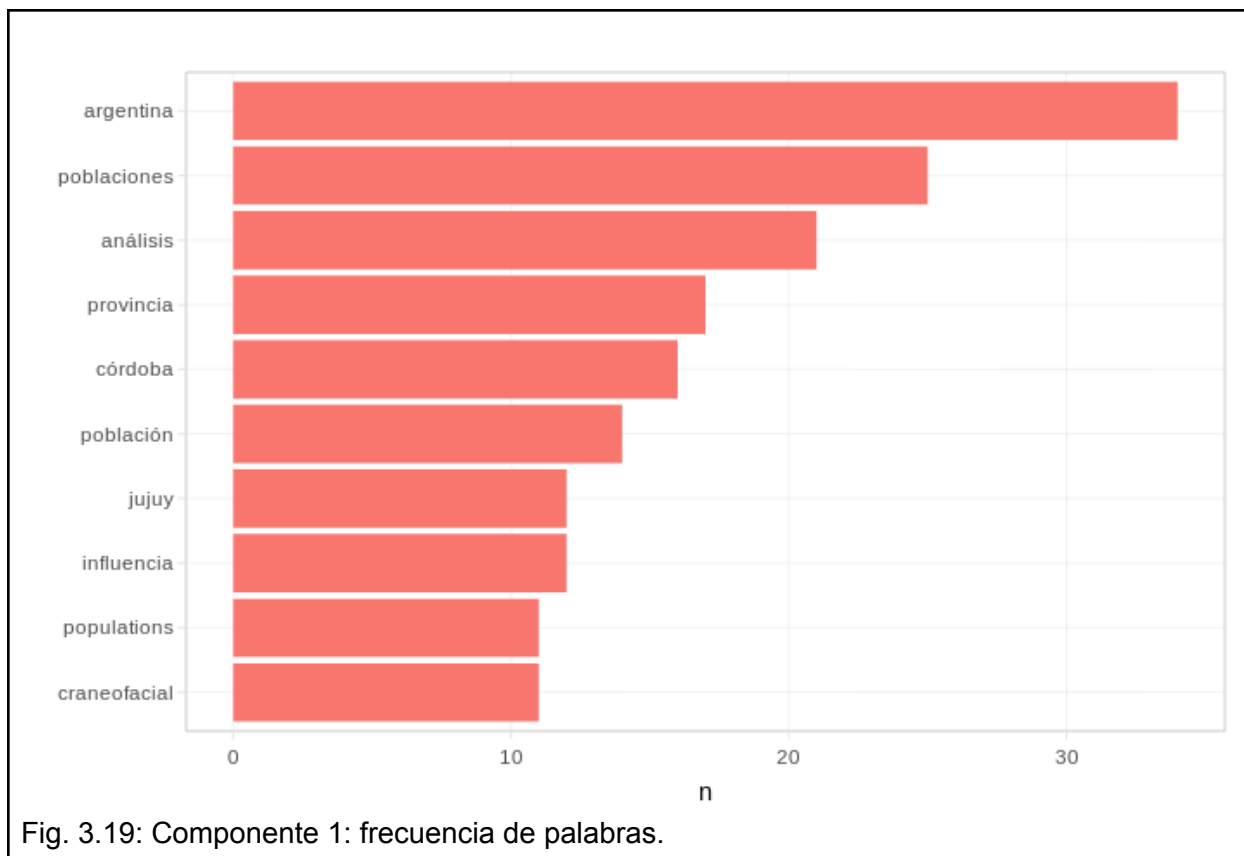
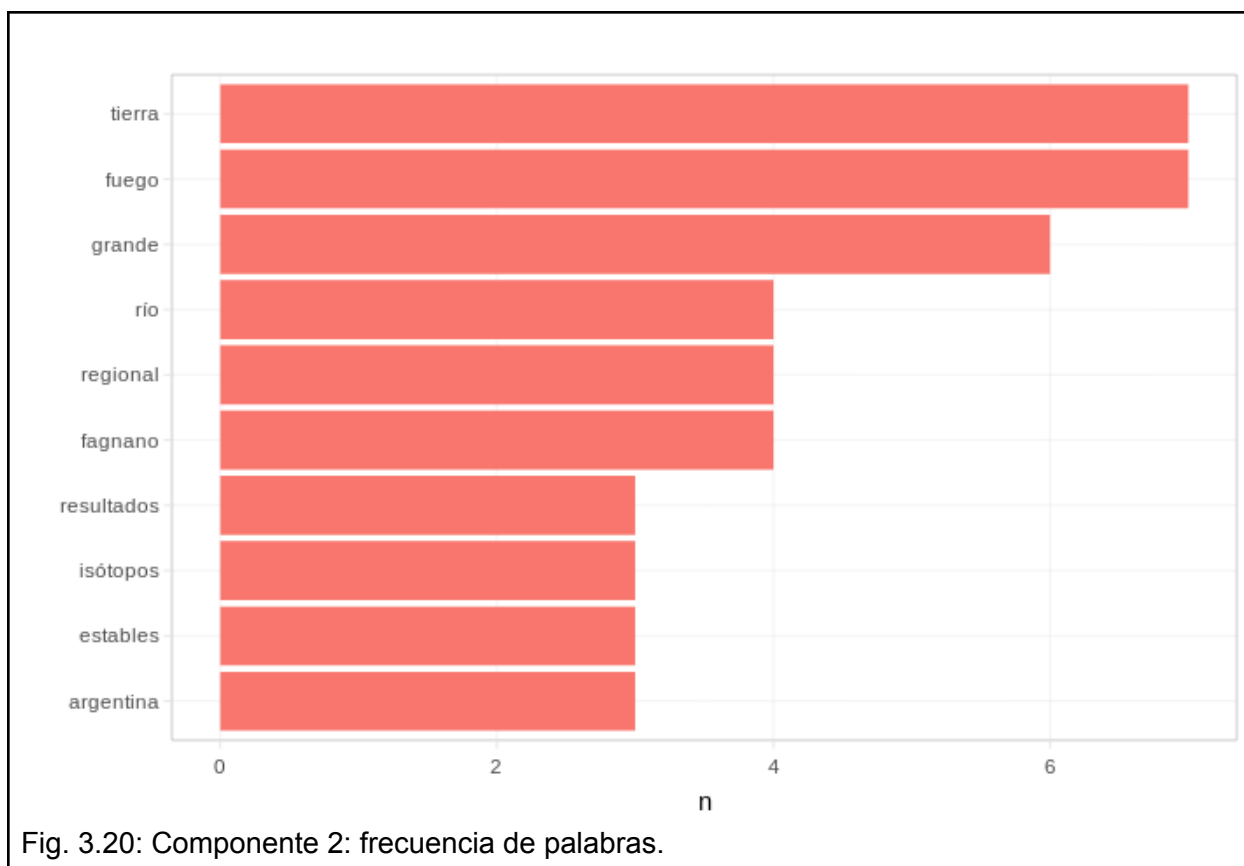


Fig. 3.19: Componente 1: frecuencia de palabras.

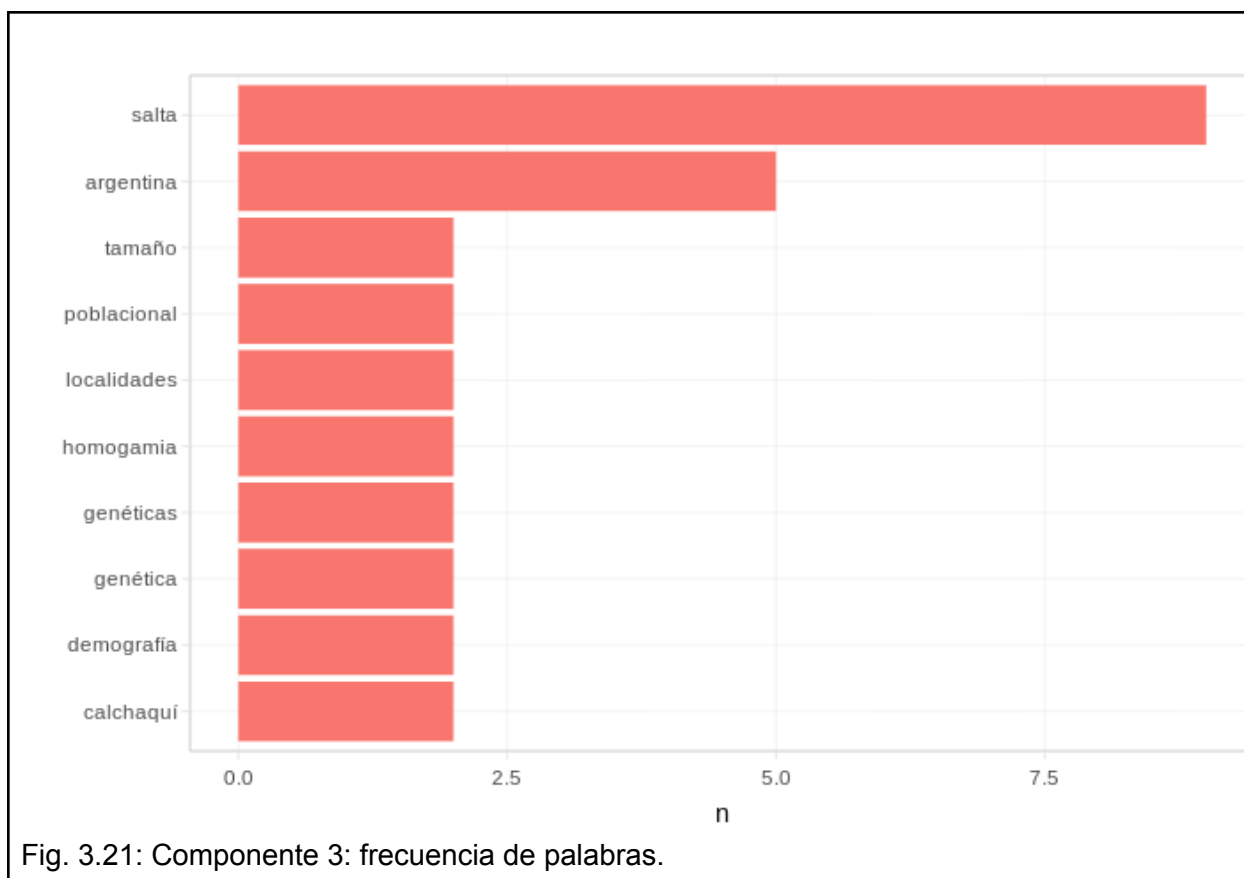
En el componente principal ([Fig. 3.19](#)) se destacan entre los términos más frecuentes algunas provincias, como Córdoba y Jujuy, a pesar de ser el componente más grande y heterogéneo a nivel geográfico. Sin embargo, la presencia del término “craneofacial”, estaría indicando que este componente tiene una relación estrecha con esa línea de investigación en varias regiones del país. Se observa además que aparecen palabras en inglés, como “populations” pero también “población” se recuperó como un término independiente, lo cual remarca que este análisis representa solamente una primera aproximación que puede mejorarse.

Con respecto al segundo componente ([Fig. 3.20](#)), puede verse que está asociado a estudios a nivel local en Tierra del Fuego (“tierra del fuego”, “río grande”), aunque también parece vinculado a una línea de investigación particular por la presencia de las palabras “isótopos” y “estables”.



Finalmente, en el componente 3 ([Fig. 3.21](#)) también se observa que se destaca la provincia de Salta, haciendo que lo geográfico sea el factor común, mientras que parecería ser heterogéneo en términos de las líneas de investigación que incluye.





En síntesis, los resultados obtenidos permiten interpretar a los componentes principalmente en términos geográficos. El componente principal abarca actividades de investigación en todo el país (se destaca la palabra “argentina”), mientras que los secundarios abarcan áreas geográficas más acotadas.

En relación a las actividades de investigación, en este caso es probable que la línea que surge del componente más grande (asociada a la palabra “craneofacial”) sea la más frecuente dentro de la disciplina a nivel nacional, lo cual repercute en una mayor cantidad de artículos publicados en la RAAB.

### Búsqueda de comunidades

En la [Tabla 3.5](#) se observa la comparación de los algoritmos de búsqueda de comunidades en la red agregada total (periodo 1996 a 2016) viendo que el algoritmo mejor posicionado es *Optimización de modularidad multinivel*, sea con o sin considerar los pesos de las aristas. En el segundo puesto podemos ver que si no consideramos los pesos, *Intermediación de aristas* queda seleccionado, mientras que si consideramos los pesos, *Búsqueda de comunidades infomap* queda mejor ranqueado.

Tabla 3.5: Comparación de los algoritmos utilizados para la búsqueda de comunidades.

Algoritmo Utilizado	Cantidad de comunidades	Valor modularidad	Valor modularidad con peso (dado 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

Estudiando los resultados del algoritmo de *Optimización de modularidad multinivel* pueden visualizarse las primeras cinco comunidades halladas ([Fig. 3.22](#)).



Fig. 3.22: primeras cinco comunidades según el algoritmo de *Optimización de modularidad multinivel* ( rojo [id comunidad: 10], azul [id comunidad: 37], verde [id comunidad: 23], violeta [id comunidad: 33], naranja [id comunidad: 18]).

## Búsqueda de sentido en los grupos obtenidos

Los resultados del análisis de texto realizado sobre los títulos de los artículos generados por las comunidades propuestas se muestran en la [Tabla 3.6](#).

Tabla 3.6: Artículos, autores y palabras de mayor frecuencia en los títulos para las primeras cinco 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)

En la comunidad 1 se destacan principalmente las palabras “craneofacial”, “crecimiento” e “influencia”. Con respecto a la palabra “influencia”, al leer los títulos de los artículos se observó que muchos de ellos comienzan con esa palabra (como por ejemplo: “*influencia de X factor en crecimiento/nutrición de Y población*”), lo cual podría explicar su relativamente alta frecuencia y remarca las limitaciones de este análisis basado únicamente en los títulos de los artículos.

En la comunidad 2, los nodos parecen estar vinculados por un componente geográfico asociado a la provincia de Jujuy, y por estudios poblacionales sobre dicha provincia.

En la comunidad 3 se destacan las palabras “influencia”, “argentina”, “crecimiento”, “nutricional” y “población”. Es posible que esta comunidad esté relacionada con el análisis de diversos factores que afectan el crecimiento en poblaciones de Argentina .

En la comunidad 4 nuevamente se resalta la temática “craneofacial”, “crecimiento” y la palabra “influencia”. Al leer algunos títulos se puede ver que los temas desarrollados por esta comunidad se asocian a cómo repercuten las dietas en el desarrollo humano. Cabe destacar que los temas asociados a las comunidades 3 y 4, según el análisis de las 10 palabras más frecuentes en los títulos, poseen similitudes por lo que resulta necesario a futuro realizar otro tipo de análisis de texto que permita profundizar e identificar mejor las líneas de investigación posiblemente asociadas a estas comunidades.

Finalmente, la comunidad 5 parece estar relacionada principalmente a la provincia Tierra del Fuego y, más específicamente, a la localidad de Río Grande. Se puede observar que la frecuencia de palabras de forma simple no resultó suficiente para este caso y que es necesario analizar compuestos de palabras, como ser “tierra del fuego” en su completitud en lugar de cada palabra por separado.

Si bien los resultados obtenidos confirman que, a futuro, sería necesario implementar análisis de texto más complejos y detallados, en muchos casos tanto los componentes como las comunidades halladas han podido asociarse con diferentes temas de investigación y regiones geográficas del país.

**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

#### Comparación con modelos

La red total de coautorías de la RAAB (es decir, desde 1996 hasta 2016) según el criterio de validación 1 expuesto por [Bielanski et al. \(2010\)](#) es de Mundo Pequeño. Sin embargo, para el criterio de validación 2 utilizando la métrica *smallworldness*, no lo es, tanto para una cota laxa de 1, como para una cota estricta de 3, según lo planteado por [Humphries-Gurney \(2008\)](#).

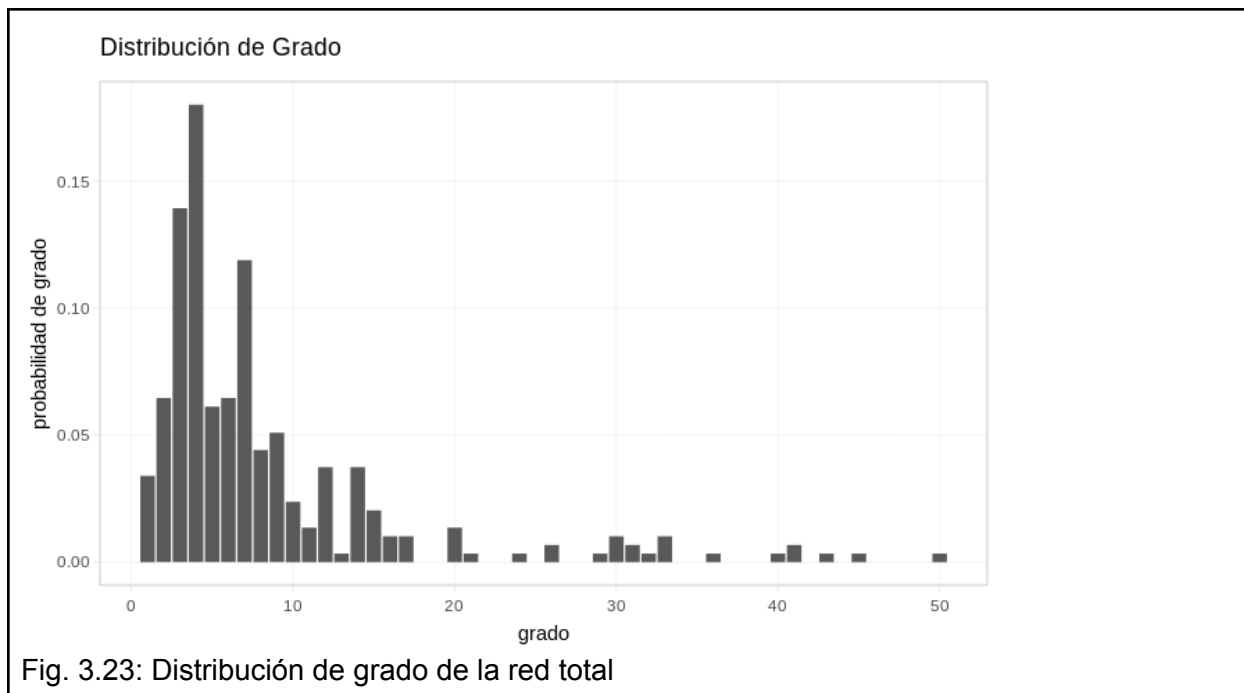
Al analizar los resultados comparados, se observa que la generación de las simulaciones a través de la función *smallworldness* varía considerablemente con lo obtenido en las simulaciones creadas utilizando *Érdos-Rényi*, según lo descrito en el criterio de validación de [Bialonski et al. \(2010\)](#). Ver ([Tabla 3.7](#)).

Tabla 3.7: Comparación de valores obtenidos para cada método de validación de 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

Por otro lado, en los análisis realizados se observa claramente que la red de coautorías de la RAAB posee muchos nodos con pocas conexiones y pocos nodos con muchas conexiones ([Fig. 3.23](#)).

Los resultados del análisis de ley de potencia (exponente alfa = 2,5156; p-valor de la prueba de Kolmogorov-Smirnov = 0,5082) confirman que en la red de coautorías de la RAAB se da el fenómeno de Libre Escala ([Fig. 3.23](#)).



Si bien para la red total (período completo de 1996 a 2016) se verificaron las características de Mundo Pequeño y Libre escala para la RAAB, el análisis particionado por períodos mostró resultados dispares ([Anexo: Detalle periodos inferiores para Mundo pequeño y Libre Escala](#)). El resumen de la comparación de las redes en el tiempo para ambas propiedades se presenta en la [Tabla 3.8](#).

Tabla 3.8: Resumen de los análisis de Mundo Pequeño y Libre Escala para todos los períodos considerados.

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

Se observa que salvo el periodo 2007-2011, la red siempre se comporta como Libre Escala. Sin embargo, para el caso de Mundo Pequeño el comportamiento depende de los diferentes métodos de validación y del período analizado. Según el método 1, es de Mundo Pequeño excepto en el período 2002-2006. En cambio, para el criterio 2, nunca cumple con la característica de mundo pequeño bajo el umbral estricto mientras que, bajo el umbral laxo los

resultados fueron más variados y, para algunos períodos, sí se validó como Mundo Pequeño ([Tabla 3.8](#)).

### Comparación con otras redes de coautoría

La [Tabla 3.9](#) muestra los resultados de la red de coautorías de la RAAB (período completo) en comparación con las compiladas por [Koseoglu \(2016\)](#).

Tabla 3.9: Métricas obtenidas para la RAAB y otras redes de coautorías de revistas científicas. Se resaltan los valores más próximos a la red de la RAAB.

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)	Revista Argentina de Antropología Biológica (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	1395693	6396	4625	1376	296	197

En cuanto a cantidad de artículos por autor, se observa que la RAAB se posiciona entre Computer Science ([Newman, 2001](#)) y Management and organization ([Acedo et al., 2006](#)). Esta métrica representa que en el tiempo analizado, los autores participan en promedio con 2.14 artículos a la revista para el caso de la RAAB.

Respecto de autores por artículo, se posiciona cerca de Biomedical ([Newman, 2001](#)). Considerando que esta métrica indica qué tanto trabajan en conjunto los autores, este resultado indica que para ambas revistas realizar un trabajo original en equipo es algo común. Este resultado es de esperarse teniendo en cuenta que las tradiciones disciplinares de aquellos grupos de investigación que publican en Biomedical son posiblemente similares a las de los autores de la RAAB.

Al revisar la transitividad, se observa que el valor más próximo corresponde a Computer Science ([Newman, 2001](#)). La transitividad es la probabilidad media de que dos coautores también sean coautores con otro en común. Esta métrica refuerza lo observado en el punto anterior de que los autores que publican en la RAAB tienden a trabajar en equipo y podría relacionarse además con el hecho de que se trata de una comunidad de autores relativamente pequeña.

Respecto de cuánto representa el componente principal en porcentaje de la red, se obtuvo un valor similar al de Strategic Management ([Koseoglu, 2016](#)). Esta métrica representa qué tan conectada se encuentra la red total. Cuanto más conectada esté significa que hay menos

equipos aislados, por trabajar en trabajos originales en conjunto por lo menos alguna vez a lo largo de la historia de la revista.

Nuevamente, se observa que según la distancia media la RAAB se posiciona más próxima a Biomedical ([Newman, 2001](#)), aunque es la que posee el valor más bajo respecto de las otras redes. La distancia media es el promedio de la distancia entre nodos conectados en la red, lo cual estaría confirmando que en este caso se trata de una comunidad de autores pequeña e interconectada

Finalmente analizando el tamaño del componente mayor se observa que la RAAB es la red más pequeña respecto a las demás, apoyando nuevamente la idea de que esta revista representa una comunidad científica relativamente pequeña.

Más allá de las métricas, para realizar una mejor comparación se debe tratar de entender el contexto de las otras revistas con las cuales se compara, por lo tanto, el análisis realizado en esta tesis representa una primera aproximación y se espera poder profundizarlo en trabajos futuros.



## Capítulo 4: Discusión y Conclusiones

En la presente tesis se evaluaron las características de la red de coautorías de la Revista Argentina de Antropología Biológica (RAAB), como caso de aplicación para el estudio de redes de coautorías mediante el desarrollo de una aplicación *web* que facilitó el análisis a través del cálculo automatizado de grafos, métricas y modelos, entre otras herramientas.

En el capítulo uno se presentó una introducción general al análisis de redes de coautoría científica incluyendo algunos conceptos claves. Además se describieron características del caso de aplicación a modo de contexto y antecedentes. Finalmente, se presentaron los objetivos, preguntas e hipótesis que se trataron en esta tesis.

En el capítulo dos se describieron las principales metodologías empleadas para la obtención del conjunto de datos, el desarrollo de la aplicación y el análisis de la red. Se definieron las fuentes de entrada, las transformaciones realizadas y las salidas de esas transformaciones. También fue descripto cómo se realizó el proceso de armado de la herramienta para explorar los datos, lo que ayudó a investigar las inquietudes iniciales de forma reproducible, interactiva y accesible por otros pares. Además, se detalló qué herramientas se usaron para su creación, diferentes consideraciones de armado y detalles técnicos de validación de hipótesis.

El capítulo tres se focalizó en utilizar esta herramienta para analizar el caso puntual de la RAAB. Este análisis se realizó contextualizando y armando un hilo conductor entre las preguntas y sus respuestas. También se centró en verificar las hipótesis planteadas al inicio de la tesis a través de los datos extraídos desde la *web* de la RAAB utilizando la aplicación creada y análisis complementarios necesarios.

Finalmente, en este capítulo se presentan las principales conclusiones y discusión asociadas a los dos objetivos planteados en esta tesis, junto con algunas consideraciones finales.

### Desarrollo de una herramienta para el análisis de redes de coautoría

En esta tesis se cumplió con el objetivo de desarrollar una aplicación *web* para facilitar el análisis de redes de coautoría científica tomando como caso de aplicación a la RAAB. Como resultado de este trabajo, surgieron los siguientes productos que pueden ser de utilidad en otras tareas y en otros ámbitos:

- **Base de datos** de los artículos de la RAAB, el link a donde están publicados y el trabajo de desambiguación de autores.
- **Aplicación web para análisis de redes de coautoría:** La aplicación en sí para realizar el análisis en la RAAB.
- **Código abierto:** El código fuente de la aplicación web para análisis de coautoría.
- **Proceso de armado de la Base de datos:** El detalle del proceso de armado de la base de datos.
- **Proceso de armado de la aplicación:** El detalle del proceso de armado de la aplicación.

Es importante destacar que la generación de la base de artículos analizados fue posible gracias a que el caso de aplicación se trata de una revista local de acceso abierto. De esta forma, al haberse documentado el proceso de creación de la base de datos, esta tesis ha proveído una forma de crear bases a partir de revistas existentes que proveen sus artículos en internet.

Asimismo, al brindar un caso de aplicación y código abierto, en esta tesis se comparten metodologías para poder reproducir este caso en otras revistas científicas, y puede ser de utilidad también para la comparación entre revistas, utilizando las herramientas implementadas en este estudio. Por otro lado, también se pueden analizar grupos que no necesariamente sean de revistas, sino círculos de publicación o de generación de trabajos conjuntos, como lo mencionado por [Price-Beaver \(1966\)](#). También se espera poder mejorar y expandir el actual proyecto por parte de futuras comunidades interesadas.

El desarrollo de esta herramienta brinda la posibilidad de construir una red de coautorías para cada revista o grupo dentro de su círculo de autores, más allá de si parte de los circuitos científicos principales, pudiendo analizar y comparar comportamientos de autores dentro de revistas independientemente de su nivel de madurez.

## Análisis del caso aplicado: redes de coautorías de la RAAB

La herramienta desarrollada permitió el análisis de la red de coautorías de la RAAB y la puesta a prueba de las hipótesis planteadas.

Bajo la hipótesis 1 se intentó verificar si la cantidad de autores y artículos creció a lo largo del tiempo en la Revista Argentina de Antropología Biológica (RAAB). A pesar de algunas tendencias encontradas para ciertos períodos dentro del rango estudiado (1996-2016), se rechazó la hipótesis de crecimiento para ambas métricas evaluadas. Esto resultó principalmente evidente en cuanto a la cantidad de artículos, puesto que en líneas generales se observó que se mantuvo relativamente constante a lo largo de los años analizados. Es posible que este hecho se relacione a ciertas características particulares de la RAAB como, por ejemplo, que durante gran parte del período analizado se trataba de una revista impresa, lo cual suele limitar el número de artículos publicados. Es posible que este límite se haya mantenido incluso posteriormente a la migración a formato digital por diferentes motivos. Cabe recordar que estas medidas corresponden generalmente a dos ediciones de la revista por año, lo cual en cierto sentido es una restricción a la cantidad de artículos publicables y probablemente también a la de autores. Además, en esta primera aproximación no solo se tuvieron en cuenta trabajos originales, sino también otro tipo de publicaciones.

Considerando la métrica la cantidad de autores en el tiempo podrían plantearse además nuevas preguntas como cuánto tiempo tardan en aparecer nuevos trabajos siendo disparados por interacción en congresos, entre otras.

Es importante destacar que se acotó el estudio a dicha publicación como una primera aproximación a la representación de la materia a nivel local y considerando que se trata del órgano de difusión oficial de la Asociación de Antropología Biológica Argentina. Sin embargo, queda de manifiesto que la puesta a prueba de una hipótesis a nivel disciplinar requiere de la expansión de los análisis a otras revistas locales e incluso publicaciones internacionales realizadas por autores locales dentro de la Antropología Biológica.

Luego de ese primer análisis de autores y artículos asociados a la RAAB, se evaluó si el avance de la revista dependía de los autores fundadores, o si estuvo impulsado por nuevos autores (Hipótesis 2). Al analizar las métricas Grado, Intermediación y Fuerza de colaboración se observó un aumento en el tiempo en los tres casos, lo cual permite concluir que los resultados son concordantes con la hipótesis de que los autores fundadores se refuerzan en el tiempo. Esto estaría indicando que la revista dependió en gran medida de los autores participantes en las primeras ediciones, al menos dentro del período de 20 años analizado. Es posible que esto se relacione con el hecho de que la creación de la RAAB y la consolidación de la disciplina en el país son fenómenos relativamente recientes, y la comunidad local de investigadores en Antropología Biológica es relativamente pequeña.

Para tratar de entender en mayor profundidad el comportamiento interno de la red se evaluó el comportamiento de los componentes bajo la 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”*). Al analizar la evolución temporal del tamaño del componente principal se identificaron momentos específicos de absorción de componentes y fue posible detallar qué componentes fueron absorbidos o integrados, en concordancia con la hipótesis planteada. De forma complementaria al análisis de componentes se verificó si existía relación con las actividades de investigación, para validar que el componente mayor sea la rama principal de investigación, mientras que los otros componentes representan ramas especializadas. Los resultados obtenidos indicaron que, en líneas generales, los componentes identificados se corresponden con distintas regiones geográficas y no pudo verificarse una asociación clara con determinadas líneas de investigación. Sin embargo, este enfoque representó una primera aproximación realizando un análisis de minería de texto sobre los títulos de los artículos publicados y se espera en trabajos futuros realizar un análisis sobre los textos completos. Asimismo, se puede profundizar el análisis para comprender qué estudios o temas fueron los que crearon puentes entre los componentes. Estas propuestas a futuro también resultan de utilidad para la búsqueda de comunidades que se realizó de forma complementaria para dar otro enfoque a la búsqueda de la rama principal de investigación y ramas especializadas. Esta búsqueda dio resultados similares a los obtenidos para los componentes, remarcando la necesidad de profundizar el estudio analizando los textos completos.

Luego de tener un panorama interno de la red de coautorías de la RAAB, se compararon algunas de sus propiedades con puntos de vista “externos”. Inicialmente se realizó la validación de propiedades comunes que se dan en redes del mundo real como Mundo Pequeño y Libre Escala para corroborar la Hipótesis 4. Los resultados obtenidos indicaron que la red de coautorías de la RAAB presenta características de Mundo Pequeño y Libre Escala dependiendo del período analizado y, en el caso de Mundo Pequeño, dependiendo también del método de validación utilizado. Finalmente, de forma complementaria a la comparación con modelos se realizó una comparación con otras redes de coautoría existentes en trabajos previos. Este análisis puso de manifiesto nuevamente algunas características particulares del caso de aplicación evaluado como, por ejemplo, que se trata de una revista cuya comunidad de autores es relativamente pequeña. Además, se observó que presenta algunas propiedades similares a revistas del área de la biomedicina, como el número de autores por artículo, lo cual puede estar relacionado a las tradiciones disciplinares y las exigencias del sistema científico.

## Consideraciones metodológicas y perspectivas futuras

Durante el desarrollo de la presente tesis hemos encontrado algunos desafíos y dificultades que vale la pena mencionar. Uno de ellos fue que los datos fueron obtenidos en 2017 pero a la hora de la escritura final de la tesis, la estructura de la página de la revista cambió y, por ejemplo, los archivos bibTex ya no se encuentran más accesibles.

Además, actualmente las referencias que tiene cada artículo son incluidas en el HTML. Sin embargo, en etapas tempranas de este estudio se descartó hacer análisis de referencias dado que, en aquel momento, estaban embebidas en los PDF y estos no eran accesibles de no ser por tareas manuales. A partir del cambio de estructura del sitio, las referencias de cada artículo están disponibles en la página de cada uno pero esto no era así al momento de extracción de los datos.

### Perspectivas futuras

Más allá de las ideas propuestas en la presente tesis, existe una amplia bibliografía en el campo de análisis de redes de coautorías y, tomadas en su conjunto, pueden expandirse a futuros trabajos, ya sean de desarrollo de herramientas para facilitar tareas o de aplicación de las técnicas para el soporte en la toma de decisiones fundamentadas en un mapa estratégico más complejo. Como menciona [Kumar \(2015\)](#), el análisis de redes de coautoría es una parte de un análisis que se debe complementar con otros estudios para lograr mejores resultados.

Finalmente el concepto de nodos que se asocian para generar un producto, y el conjunto de herramientas utilizadas para analizar este fenómeno puede abrir otras implementaciones en otras ramas diferentes de coautoría de artículos científicos.

A partir de los desafíos y los análisis realizados en la presente tesis, es posible sugerir varias líneas para explorar a futuro. A continuación mencionamos algunas de ellas.

#### Actualización de la base de datos

Un producto deseable podría ser expandir la base de datos y quitar la cota temporal. De esta forma, sería posible reutilizar la aplicación para expandir el análisis a más periodos de publicación en la RAAB.

También podría replicarse con la misma estructura para otras revistas, logrando que los análisis sean comparables a través del mismo proceso. Esto no es un trabajo trivial dado que una nueva base de datos implica el proceso de desambiguación de autores, la carga de datos en la aplicación, y luego sacarle la cota temporal en los lugares donde está filtrado. Sin embargo, el trabajo realizado en esta tesis agilizaría mucho la realización de estudios similares.

#### Aspectos técnicos

Como propuesta a futuro se planea una actualización de la aplicación en cuanto a su estructura, procesos y performance. Así como también aspectos estéticos y de calidad de la herramienta, como puede ser navegabilidad, y sugerencias de trabajo para realizar análisis específicos. También pueden incluirse actualizaciones en las evaluaciones y métricas disponibles. Se propone además transformar esta herramienta en uno o varios paquetes, haciendo que sea más fácilmente portable para otras revistas, así como para separar los procesos de pre-procesamiento para la obtención de datos de la aplicación para análisis propiamente dicho. Por este motivo se optó por liberar el código fuente para que sea accesible

y pueda ir evolucionando como comunidad para poder aceptar propuestas de interesados en el proyecto.

### Proyección sobre artículos

Dada la red de naturaleza bipartita (autores - escriben - artículos), el trabajo actual se basó en la proyección sobre autores, pudiéndose leer: “*autor - trabaja con - autor*”. La otra proyección basada sobre los nodos artículos, que se puede leer como “*artículo - comparte autor con - artículo*”, representa un interesante análisis para tener en cuenta en trabajos futuros.

### Análisis de textos y tópicos en el tiempo

En etapas muy tempranas del desarrollo de la presente tesis otro objetivo de interés era el análisis de texto en base a los artículos publicados en la RAAB en el período estudiado. Sin embargo, se decidió no incluir estos análisis, dado que extenderían considerablemente los tiempos de trabajo y forman parte de otro tipo de hipótesis. Además, otro aspecto que ayudó a tomar la decisión es que al momento de escritura de la tesis se está desarrollando una “revolución” con respecto a la interpretación de imágenes, que avista que se obtendrán mejores resultados facilitando esta tarea en trabajos futuros.

Por lo tanto, sería interesante realizar análisis textual de los PDF para la extracción de tópicos, evaluar su evolución en el tiempo y su relación con la red de coautoría. En particular, sería posible verificar si los tópicos resultantes cuadran con las descripciones de las líneas de investigación en Antropología Biológica planteadas por [Cesani et al.\(2013\)](#), por ejemplo. Además, se podría validar si los grupos de coautores identificados en la red están asociados a los tópicos determinados, de manera similar al trabajo realizado por [Janssens \(2007\)](#).

### Análisis de conectividad / puentes

Se denomina “puente” a aquella arista que conecta, por ejemplo, a los dos únicos componentes que conforman una red. El puente transforma a la red en “conectada” y, de esta manera, representa “*cualquier link que si se corta, desconecta la red*” ([Barabási \(2016\):Section 2.9 Connectedness](#)). Este análisis permite plantear varias preguntas estratégicas.

La herramienta desarrollada en esta tesis ayuda a analizar mejor cómo se fueron dando los puentes en el tiempo, así como qué elementos generaron los mismos. También puede utilizarse para analizar componentes candidatos para incentivar a generar nuevos “Puentes” entre los componentes.

Algunas preguntas asociadas serían:

- ¿Existen posibilidades de generar estos puentes vía coautoría?
- ¿Qué tipo de estudios “podrían generar” dichos puentes?
- ¿Vale la pena tratar de unir estos componentes?
- ¿Qué beneficios se obtendrán de dichos puentes?
- ¿Cuánto cuesta crear esos puentes?

Así como se generan puentes para componentes, se puede tratar de fomentar la conectividad entre comunidades, fomentando nuevas áreas de investigación o complemento entre equipos. En este sentido, algunas preguntas que pueden surgir son:

- ¿Quiénes son los puentes entre las comunidades?
- ¿Qué trabajos se encuentran asociados a los puentes?
- ¿Qué temáticas incluyen estos trabajos?

Dentro del análisis de relaciones se puede evaluar el vínculo entre los autores en el tiempo, conocer quiénes están vinculados, qué temas tratan, en qué época, si existe una frecuencia, entre otras.

## Otros Análisis

Como análisis complementario a la evolución de métricas para los autores fundadores se podría incorporar la variación interanual para ver cuánto crecen estos valores año a año. Recordemos que el Grado es la cantidad de nuevos nexos que genera el autor (son nuevos porque no estamos considerando el fortalecimiento de una relación por hacer más trabajos en conjunto, para eso está la Fuerza de Colaboración).

Para la sección de componentes, queda profundizar el enfoque para comprender con mayor detalle estos componentes así como su composición de temas, sobre todo porque el componente mayor contiene múltiples artículos que seguramente no fueron representados en las primeras 10 palabras. De la misma forma sería interesante, a futuro, revisar en el tiempo los diferentes componentes y sus temas para evaluar si mientras se va armando el componente gigante va absorbiendo las diferentes comunidades, enriqueciendo así los temas tratados a nivel revista. Otro aspecto a explorar en estas temáticas es un análisis de texto más complejo, teniendo en cuenta, por ejemplo, los conceptos compuestos por varias palabras (como 'genética poblacional' o 'tierra del fuego').

## Comparación con otras revistas

La comprensión de los contextos de las revistas entre las cuales se realizan comparaciones representa un complemento fundamental al análisis de métricas como el realizado en la presente tesis. Para profundizar en la comparación de la revista estudiada con otras se pueden plantear diversas preguntas que faciliten la comprensión de los contextos, como por ejemplo: ¿de qué temáticas tratan?, ¿se acercan más a lo técnico o a las ciencias humanas?, ¿cuál es la cota de artículos por edición?, ¿con qué periodicidad se publica?, ¿publican autores de todo el mundo?, ¿publican solo artículos locales?, ¿en qué idioma publican?, ¿qué políticas de publicación tienen? Las respuestas a estas preguntas pueden incluso variar en el tiempo para cada revista analizada.

## Consideraciones finales

El desarrollo de la herramienta para análisis de redes de coautoría realizado en esta tesis representa un 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. En relación a esto, 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.

Finalmente, cabe destacar que esta tesis se realizó en un marco de colaboración entre profesionales de la Ciencia de Datos y la Antropología Biológica y, en este sentido, refleja los desafíos y limitaciones que presenta el trabajo multidisciplinar, como así también sus ventajas y las múltiples posibilidades que se abren a futuro.



# Bibliografía

AABA, Asoc. Antrop. Biolog. Arg. Revista Argentina de Antropología Biológica. URL: <https://revistas.unlp.edu.ar/raab/>. Accessed: 2020-01-06.

Acedo, F. J., Barroso, C., Casanueva, C., & Galan, J. L. (2006). Co-Authorship in Management and Organizational Studies: An Empirical and Network Analysis\*. *Journal of Management Studies*, 43(5), 957–983. doi:10.1111/J.1467-6486.2006.00625.X

Aria, M. and Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis, *Journal of Informetrics*, 11(4), pp 959-975, Elsevier, DOI: 10.1016/j.joi.2017.08.007

Barabási, A. L. (2016). *The power of network science, the beauty of network visualization*. Cambridge University Press. ISBN: 978-1107076266. URL: <http://networksciencebook.com/>.

Beigel F. (2014) *Publishing from the periphery: Structural heterogeneity and segmented circuits. The evaluation of scientific publications for tenure in Argentina's CONICET. INCIHUSA-CONICET, Universidad Nacional de Cuyo, Argentina* URL: <https://doi.org/10.1177/0011392114533977>

Bialonski, S, M. Horstmann, and K. Lehnertz (2010). “From brain to earth and climate systems: Small-world interaction networks or not?” In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 20.1, p. 013134. DOI: 10.1063/1.3360561. URL: <https://doi.org/10.1063/1.3360561>

Blischak, J. D, E. R. Davenport, and G. Wilson (2016). “A Quick Introduction to Version Control with Git and GitHub”. In: *PLOS Computational Biology* 12.1, pp. 1-18. DOI: 10.1371/journal.pcbi.1004668. URL: <https://doi.org/10.1371/journal.pcbi.1004668>.

Blondel, V. D, J. Guillaume, R. Lambiotte, et al. (2008). “Fast unfolding of communities in large networks”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, p. P10008. DOI: 10.1088/1742-5468/2008/10/p10008. URL: <https://doi.org/10.1088/1742-5468/2008/10/p10008>.

Bourgeois, D. T. (2014). *Information Systems for Business and Beyond*. Open Textbook Challenge by the Saylor Academy. URL: <https://bus206.pressbooks.com/>.

Cesani Rossi, M. F, M. Sardi, S. E. Colantonio, et al. (2013). “Líneas de investigación actuales de la antropología biológica argentina/Current researches in Argentine biological anthropology”. In: *Revista Argentina de Antropología Biológica* 16.1, pp. 31-37. URL: <https://revistas.unlp.edu.ar/raab/article/view/749>.

Clauset, A, M. E. J. Newman, and C. Moore (2004). “Finding community structure in very large networks”. In: *\_Phys. Rev. E\_* 70 (6), p. 066111. DOI: 10.1103/PhysRevE.70.066111. URL: <https://link.aps.org/doi/10.1103/PhysRevE.70.066111>.

Clauset, A, C. R. Shalizi, and M. E. J. Newman (2009). “Power-Law Distributions in Empirical Data”. In: *SIAM Review* 51.4, pp. 661-703. DOI: 10.1137/070710111. URL: <https://doi.org/10.1137/070710111>.



Cobo, M., López-Herrera, A., Herrera-Viedma, E. and Herrera, F. (2011), *Science mapping software tools: Review, analysis, and cooperative study among tools*. *J. Am. Soc. Inf. Sci.*, 62: 1382-1402. <https://doi.org/10.1002/asi.21525>

Cobo, M., López-Herrera, A., Herrera-Viedma, E. and Herrera, F. (2012), *SciMAT: A new science mapping analysis software tool*. *J Am Soc Inf Sci Tec*, 63: 1609-1630. <https://doi.org/10.1002/asi.22688>

Erdős, P. and A. Rényi (1959). "On random graphs I". In: *Publ. Math. Debrecen*.

Fatt, C. K, E. A. Ujum, and K. Ratnavelu (2010). "The structure of collaboration in the *Journal of Finance*". In: *Scientometrics* 85.3, pp. 849-860. ISSN: 1588-2861. DOI: 10.1007/s11192-010-0254-0. URL: <https://doi.org/10.1007/s11192-010-0254-0>

Freeman, L. C. (1977). *A Set of Measures of Centrality Based on Betweenness*. *Sociometry*, 40(1), 35–41. <https://doi.org/10.2307/3033543>

Freeman, L. C. (1978). "Centrality in social networks conceptual clarification". In: *\_Social Networks\_* 1.3, pp. 215 - 239. ISSN: 0378-8733. DOI: [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7). URL: <http://www.sciencedirect.com/science/article/pii/0378873378900217>

Freeman, L.C. (2004) *The Development of Social Network Analysis: A Study in the Sociology of Science*. ISBN: 9781594577147

Glänzel W., Schubert A. (2004) *Analysing Scientific Networks Through Co-Authorship*. In: Moed H.F., Glänzel W., Schmoch U. (eds) *Handbook of Quantitative Science and Technology Research*. Springer, Dordrecht. [https://doi.org/10.1007/1-4020-2755-9\\_12](https://doi.org/10.1007/1-4020-2755-9_12)

Humphries, M. D. and K. Gurney (2008). "Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence". In: *PLOS ONE* 3.4, pp. 1-10. DOI: 10.1371/journal.pone.0002051. URL: <https://doi.org/10.1371/journal.pone.0002051>.

Janssens, F. (2007). *Clustering of scientific fields by integrating text mining and bibliometrics*. Katholieke Universiteit Leuven, Open Access publications from Katholieke Universiteit Leuven.

Kendall, M. (1975). *Rank Correlation Methods*. Griffin, London, 202 pp.

Khalik, M. N., Ouarda, T. B. M. J., Gachon, P., Sushama, L., and St-Hilaire, A. (2009). *Identification of hydrological trends in the presence of serial and cross correlations: A review of selected methods and their application to annual flow regimes of Canadian rivers*. *Journal of Hydrology*, 368: 117-130.

Kleinberg, J. M. (1999). "Authoritative Sources in a Hyperlinked Environment". In: *J. ACM* 46.5, p. 604–632. ISSN: 0004-5411. DOI: 10.1145/324133.324140. URL: <https://doi.org/10.1145/324133.324140>.

Kolaczyk, E. D. and G. Csárdi (2020). *\_Statistical Analysis of Network Data with R\_*. En. Springer International Publishing. DOI: <https://doi.org/10.1007/978-3-030-44129-6>.

- Koseoglu, M. A. (2016). "Growth and structure of authorship and co-authorship network in the strategic management realm: Evidence from the Strategic Management Journal". In: *BRQ Business Research Quarterly* 19.3, pp. 153 - 170. ISSN: 2340-9436.  
DOI: <https://doi.org/10.1016/j.brq.2016.02.001>.  
URL: <http://www.sciencedirect.com/science/article/pii/S2340943616300019>.
- Kundzewicz, Z.W. and Robson, A. (2000) *Detecting Trend and Other Changes in Hydrological Data*. World Meteorological Organization, Geneva, 157.
- Kumar, S. (2015). "Co-authorship networks: a review of the literature". In: *Aslib Journal of Information Management* 67.1, pp. 55-73. ISSN: 2050-3806. DOI: 10.1108/AJIM-09-2014-0116.  
URL: <https://doi.org/10.1108/AJIM-09-2014-0116>.
- Longabaugh, W. J. (2012). "Combing the hairball with BioFabric: a new approach for visualization of large networks". In: *BMC Bioinformatics* 13.1, p. 275. ISSN: 1471-2105. DOI: 10.1186/1471-2105-13-275. URL: <https://doi.org/10.1186/1471-2105-13-275>.
- Madrigal, L. and J. R. González (2016). *Introducción a la antropología biológica*. Asociación Latinoamericana de Antropología Biológica (ALAB).  
URL: [https://scholarcommons.usf.edu/islac\\_alab\\_antropologia/1](https://scholarcommons.usf.edu/islac_alab_antropologia/1).
- Mallig, Nicolai (2010), "A relational database for bibliometric analysis"  
*Journal of Informetrics*, Volume 4, Issue 4, 2010, <https://doi.org/10.1016/j.joi.2010.06.007>.
- Mann, H. B. (1945). *Nonparametric Tests Against Trend*. *Econometrica*, 13(3): 245-259.
- Milgram, S. (1967). "The small world problem". In: *Psychol. Today* 2.1, pp. 60-67.
- Morel, C. M, S. J. Serruya, G. O. Penna, et al. (2009). "Co-authorship Network Analysis: A Powerful Tool for Strategic Planning of Research, Development and Capacity Building Programs on Neglected Diseases". In: *PLOS Neglected Tropical Diseases* 3.8, pp. 1-7.  
DOI: 10.1371/journal.pntd.0000501. URL: <https://doi.org/10.1371/journal.pntd.0000501>.
- Newman ME, Moore C, Watts DJ (2000) *Mean-field solution of the small-world network model*. *Phys Rev Lett* 84: 3201–3204. DOI: 10.1103/PhysRevLett.84.3201
- Newman, M. E. J. (2001). "The structure of scientific collaboration networks". In: *Proceedings of the National Academy of Sciences* 98.2, pp. 404-409. ISSN: 0027-8424.  
DOI: 10.1073/pnas.98.2.404. eprint: <https://www.pnas.org/content/98/2/404.full.pdf>.  
URL: <https://www.pnas.org/content/98/2/404>.
- Newman, M. E. J. (2004). "Coauthorship networks and patterns of scientific collaboration". In: *Proceedings of the National Academy of Sciences* 101.suppl 1, pp. 5200-5205.  
ISSN: 0027-8424. DOI: 10.1073/pnas.0307545100.  
URL: [https://www.pnas.org/content/101/suppl\\_1/5200](https://www.pnas.org/content/101/suppl_1/5200).
- Newman, M. E. J. and M. Girvan (2004). "Finding and evaluating community structure in networks". In: *Phys. Rev. E* 69 (2), p. 026113. DOI: 10.1103/PhysRevE.69.026113.  
URL: <https://link.aps.org/doi/10.1103/PhysRevE.69.026113>.

Newman, M. E. J. (2005). "Power laws, Pareto distributions and Zipf's law". In: *Contemporary Physics* 46.5, pp. 323-351. DOI: 10.1080/00107510500052444. URL: <https://doi.org/10.1080/00107510500052444>.

Newman, M. E. J. (2006). "Finding community structure in networks using the eigenvectors of matrices". In: *Phys. Rev. E* 74 (3), p. 036104. DOI: 10.1103/PhysRevE.74.036104. URL: <https://link.aps.org/doi/10.1103/PhysRevE.74.036104>.

Newman, M. E. J. (2016). "Mathematics of Networks". In: *The New Palgrave Dictionary of Economics*. London: Palgrave Macmillan UK, pp. 1-8. ISBN: 978-1-349-95121-5. DOI: 10.1057/978-1-349-95121-5\_2565-1. URL: [https://doi.org/10.1057/978-1-349-95121-5\\_2565-1](https://doi.org/10.1057/978-1-349-95121-5_2565-1).

Önöz , B. and Bayazit M. (2012). Block bootstrap for Mann-Kendall trend test of serially dependent data. *Hydrological Processes*, 26: 3552-3560.

Ordoñez, G, A. Hernández Badillo, C. Hernández, et al. (2009). "Análisis bibliométrico de la Revista de Economía Institucional en sus primeros diez años". In: *Revista de Economía Institucional* 11.20. URL: <https://revistas.uexternado.edu.co/index.php/ecoins/article/view/354>.

Pons, P. and M. Latapy (2005). Computing communities in large networks using random walks (long version). arXiv: physics/0512106 [physics.soc-ph].

Price, D. and Beaver, D. (1966). Collaboration in an Invisible College. *The American psychologist*. 21. 1011-8. DOI:10.1037/h0024051.

Raghavan, U. N, R. Albert, and S. Kumara (2007). "Near linear time algorithm to detect community structures in large-scale networks". In: *Physical Review E* 76.3. ISSN: 1550-2376. DOI: 10.1103/physreve.76.036106. URL: <http://dx.doi.org/10.1103/PhysRevE.76.036106>.

Rosvall, M, D. Axelsson, and C. T. Bergstrom (2009). "The map equation". In: *The European Physical Journal Special Topics* 178.1, p. 13–23. ISSN: 1951-6401. DOI: 10.1140/epjst/e2010-01179-1. URL: <http://dx.doi.org/10.1140/epjst/e2010-01179-1>.

Santa, S., & Herrero-Solana, V. (2010). Coverage in scopus vs. web of science off research produced in Latin America and the Caribbean. *Investigación bibliotecológica*, 24(52), 13-27. DOI:10.22201/iibi.0187358xp.2010.52.27451

Sardi, M., Bailliet, G., & Carnese, F. (2016). 20 años. *Revista Argentina De Antropología Biológica*, 18(2). Recuperado a partir de <https://revistas.unlp.edu.ar/raab/article/view/2541>

Saurkar, A. V., K. G. Pathare, and S. A. Gode (2018). n Overview on Web Scraping Techniques and Tools. URL: [http://www.ijfrsce.org/download/browse/Volume\\_4/April\\_18\\_Volume\\_4\\_Issue\\_4/1524638955\\_25-04-2018.pdf](http://www.ijfrsce.org/download/browse/Volume_4/April_18_Volume_4_Issue_4/1524638955_25-04-2018.pdf).

Savić, M., M. Ivanović, M. Radovanović, et al. (2015). "Exploratory Analysis of Communities in Co-authorship Networks: A Case Study". In: *ICT Innovations 2014. Springer International Publishing*, pp. 55-64. DOI: [https://doi.org/10.1007/978-3-319-09879-1\\_6](https://doi.org/10.1007/978-3-319-09879-1_6). URL: [https://link.springer.com/chapter/10.1007/978-3-319-09879-1\\_6](https://link.springer.com/chapter/10.1007/978-3-319-09879-1_6).

Siegel, S., & Castellan, N. J. (1972). *Estadística no paramétrica aplicada a las ciencias de la conducta* (2a ed, pp. 117-120). Mexico: Trillas.

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*. *Química Nova* [online]. 2012, v. 35, n. 4, pp. 671-676. URL: <https://doi.org/10.1590/S0100-40422012000400003>. ISSN 1678-7064.

van Eck, N.J., Waltman, L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84, 523–538 (2010). <https://doi.org/10.1007/s11192-009-0146-3>

Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393: 440–442. DOI: 10.1038/30918.

Wilson, G, J. Bryan, K. Cranston, et al. (2017). “Good enough practices in scientific computing”. In: *PLOS Computational Biology* 13.6, pp. 1-20. DOI: 10.1371/journal.pcbi.1005510. URL: <https://doi.org/10.1371/journal.pcbi.1005510>.

Wojciechowski, J, A. M. Hopkins, and R. N. Upton (2015). “Interactive Pharmacometric Applications Using R and the Shiny Package”. Eng. In: *CPT: pharmacometrics & systems pharmacology* 4.3. 26225240[pmid], pp. e00021-e00021. ISSN: 2163-8306. DOI: 10.1002/psp4.21. URL:<https://pubmed.ncbi.nlm.nih.gov/26225240>.

# Anexos

## Anexo: resumen proceso obtención y generación de la base de datos

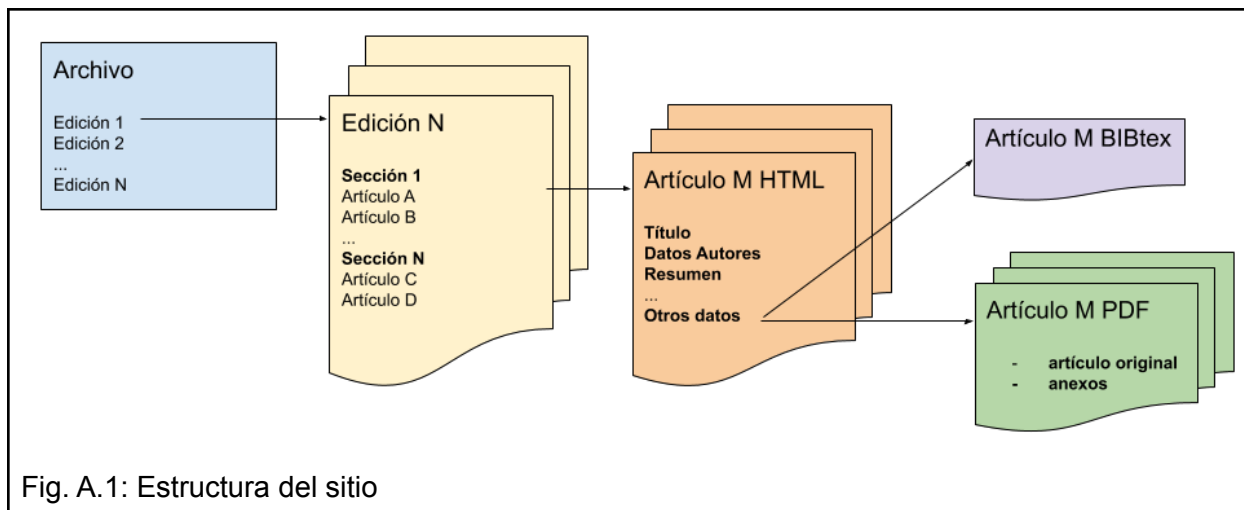
El proceso de adquisición de los datos quedó definido de la siguiente forma:

Cuadro A.1: Descripción del proceso de adquisición y generación de la base de datos

1. Análisis de estructura del sitio (código fuente de las páginas, navegación del sitio)
2. Adquisición de los datos
  - a. Descarga del archivo de ediciones
    - i. Generación de un índice de ediciones (tabla intermedia)
  - b. Descarga de cada edición
    - i. Generación de un índice de artículos (tabla intermedia)
  - c. Descarga de los artículos asociados a las ediciones
    - i. Descarga de HTML
    - ii. Descarga de PDF
    - iii. Descarga de bibtex
  - d. Interpretación de los datos
    - i. Interpretación de ediciones
    - ii. Interpretación de artículos
    - iii. Interpretación de bibtex
    - iv. Interpretación de PDF
3. Generación de la base
  - a. Generación de Resultados intermedios
  - b. Generación de la base relacional
  - c. Generación de los grafos

## Anexo: web scraping

Para cada artículo existe una página en la que se publican datos del artículo junto con un resumen del mismo, los PDF asociados y las citas a los artículos con un *link* para la generación de citas en diferentes formatos ([Fig. A.1](#))



Se puede ver en la [Tabla A.2](#) los datos extraídos, sus orígenes y la razón por la cual resultan de interés.

Tabla A.1: Datos extraídos del sitio web de la RAAB.

Origen de datos	Razón de interés en el origen de datos	Datos extraídos
Página de archivo	Para poder navegar a las ediciones de la revista, en donde se extraen los datos considerados de interés para este estudio.	Listado de ediciones de la revista, y para cada edición: Nombre de edición, <i>link</i> a la edición
Páginas de cada edición	Para cada edición de la revista se encuentran los datos propios de la edición, secciones que la componen y links a cada artículo.	Fecha publicación edición, Nombre edición, Secciones de la edición: Nombre sección, Pertenencia de artículos a secciones; Artículos en la edición: Título de artículo, <i>Link</i> a artículo, <i>Link</i> a pdfs por artículo, Autores de artículo, Páginas de artículo.
Páginas de Artículos	Para cada artículo se encuentran los datos de interés detallados en la columna de datos extraídos.	Título del artículo, Nombre edición, Sección edición, Autores del artículo, Afiliación Autores ( opcional ), Resumen ( opcional ), Abstract ( opcional ), Pdfs asociados ( opcional ), Fecha de publicación del artículo,

		Forma de cita, doi ( opcional ), Link a articulo ( esta página ),
Páginas de citas (*1)	Estas páginas no están directamente visibles, pero se pueden extraer por el javascript del <i>link</i> de cambio de formato de cita; esto permite bajar un HTML con diferentes formatos de citas. Se decidió usar el formato bibtex dado que es muy amigable para la extracción de los datos; además de ser un estándar que se puede transformar a otros en caso de ser necesario. Para cada página de citas se puede extraer el texto de formato bibtex; pudiendo generar el archivo .bib a posterior.	author title journal volume number year Keywords ( opcional ) Abstract ( opcional ) issn pages url
Archivos PDF	Cada artículo, su texto completo se encuentra en formato PDF. Hay artículos que tienen más de un pdf dado que tienen anexos de información complementaria al principal. Si bien se bajan todos los PDF, todavía no se decide si se va a trabajar solo el principal o la información complementaria también.	El texto completo del artículo

(\*1) las páginas de citas ya no están disponibles desde 2018

## Adquisición de los datos

La obtención de los datos se realizó siguiendo el plan de acción detallado en la [Fig. A.2](#). El detalle de los pasos está en la [Tabla A.3](#).

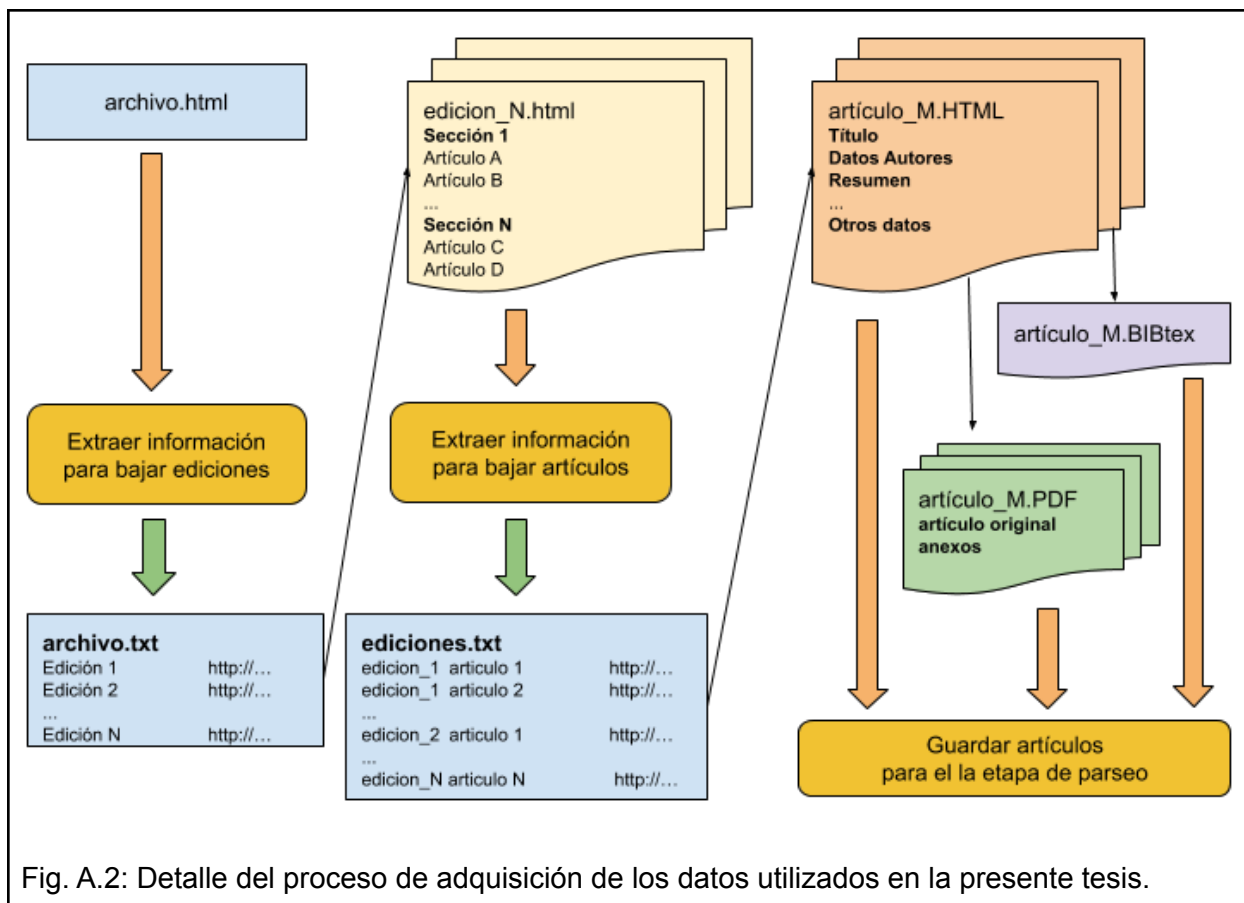


Fig. A.2: Detalle del proceso de adquisición de los datos utilizados en la presente tesis.

Tabla A.2: Detalle de las tareas del proceso de adquisición de datos

tarea	descripción	entrada	resultado
Descargar el archivo	Descargar la página HTML del archivo de ediciones de la revista	Url del archivo de ediciones de la revista	archivo.html
Generar un índice de ediciones	Obtiene del archivo las ediciones y la URL de cada una	archivo.html	indice_ediciones.txt
Descargar ediciones	Descargar la página HTML de cada edición.	indice_ediciones.txt	edicion_####.html
Generar un índice de artículos	Obtiene de cada edición, las url de los artículos.	indice_ediciones.txt	indice_articulos.txt
Descargar artículos (HTML) (PDF) (BIBTEX)	Descargar de cada artículo, los archivos HTML / PDF y BIBTEX.	indice_articulos.txt	articulo_####.html articulo_####.bibtex articulo_####.pdf articulo_####_anexo_### #.pdf



## Interpretación de los archivos

Se habla de *Interpretación (parsing)* para referirse a la lectura de los archivos para su transformación desde su formato original a uno más conveniente para la extracción de los datos que ayude a generar la información necesaria para el estudio.

## Interpretación de los archivos HTML

Para los datos en archivos HTML se utilizó, en el lenguaje R, la [librería rvest](#) que facilita la extracción de este tipo de archivos.

## Interpretación de los archivos bibTex

Para los datos bibTex, se utilizó primero la [librería rvest](#) para extraer el HTML, luego se corrigieron errores y por último se procesaron con la [librería RefManageR](#). Para más detalles ver [ANEXO: Detalle procesamiento bibTEX](#).

## Interpretación de los archivos PDF

Los archivos PDF debieron ser procesados para extraer los textos en un formato más flexible para la interpretación por una computadora. Por lo tanto, los textos se extraen de formato .pdf a texto plano.

Sin embargo, la particularidad del conjunto de datos en este caso fue que no todos los PDF tenían la misma estructura. Por ejemplo, desde el punto de vista del almacenamiento, algunos archivos son el resultado de un escaneo, mientras que otros son texto. Además, desde el punto de vista de la disposición (*Layout*) del texto, no todos tienen el mismo formato de texto, como espaciados, formato bicolumnar o simple, entre otros detalles. Para más detalle acerca de las dificultades relacionadas al trabajo con archivos PDF ver [ANEXO: Detalle consideraciones PDF](#).

## Anexo: Base de datos: Estructura relacional

Diagrama entidad relación (Fig. A.3). En la tabla intermedia existieron campos replicados para facilitar las consultas a la base (Tabla A.3).

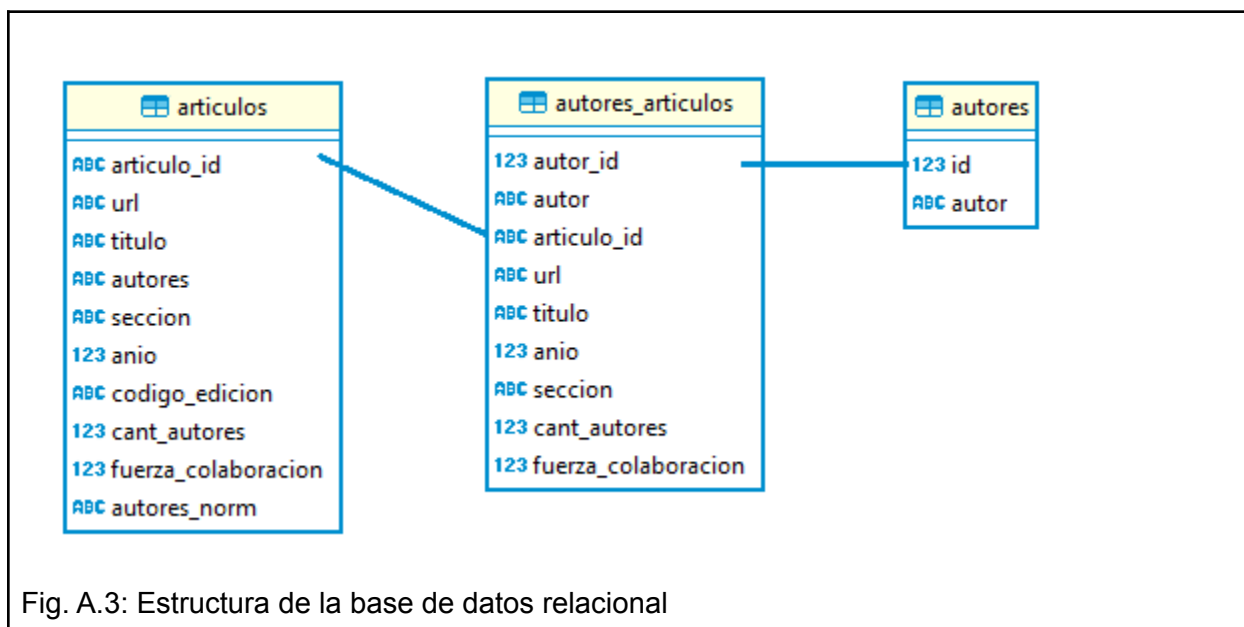


Tabla A.3: Propiedades de las tablas que conforman la base de datos relacional.

tabla	campo	descripción	Ejemplos / Observaciones
articulos	articulo_id	Identificador de artículo	Ejemplo: 1090, extraído de la página.
articulos	url	Url del artículo	Ejemplo: <a href="https://revistas.unlp.edu.ar/raab/article/view/1090">https://revistas.unlp.edu.ar/raab/article/view/1090</a>
articulos	titulo	Título del artículo	Ejemplo: Riesgo de anomalías congénitas en grupos étnicos de Sudamérica / Risk of congenital anomalies in ethnic groups from South America
articulos	autores	Autores del artículo separados por “,”	Ejemplo: Maria I. Villalba;Hebe Campaña;Sandra C. Scala;Mariela S. Pawluk;Jorge S. López-Camelo
articulos	seccion	Sección de la edición	Ejemplo: Trabajos Originales
articulos	anio	Año de publicación en la revista	Ejemplo: 2015
articulos	codigo_edicion	Código de edición	Ejemplo: 2015_17_1
articulos	cant_autores	Cantidad de autores que participaron en el artículo	Ejemplo: 5

articulos	fuerza_colaboracion	Fuerza de colaboración que otorga el artículo por haber participado en él.	Ejemplo: 0.2
articulos	autores_norm	Autores normalizados, ordenados en orden alfabético.	Ejemplo: Hebe Campaña; Jorge S. López-Camelo; María I. Villalba; Mariela S. Pawluk; Sandra C. Scala
autores	id	Id del autor	Ejemplo: 1
autores	autor	Nombre del autor	Ejemplo: A. Benítez
autores_articulos	autor_id	Id del autor	1
autores_articulos	autor	Nombre del autor	Campo replicado de la tabla articulos, para facilitar consultas
autores_articulos	articulo_id	Id del articulo	1821
autores_articulos	url	Url del articulo	Campo replicado de la tabla articulos, para facilitar consultas
autores_articulos	titulo	Titulo del articulo	Campo replicado de la tabla articulos, para facilitar consultas
autores_articulos	anio	Año de publicación del artículo en la revista	Campo replicado de la tabla articulos, para facilitar consultas
autores_articulos	seccion	Sección del artículo en la edición	Campo replicado de la tabla articulos, para facilitar consultas
autores_articulos	cant_autores	Cantidad de autores en el artículo.	Campo replicado de la tabla articulos, para facilitar consultas
autores_articulos	fuerza_colaboracion	Fuerza de colaboración que otorga el artículo por haber participado en él.	Campo replicado de la tabla articulos, para facilitar consultas

## Anexo: Generación de grafos

La transformación de la base de datos relacional a un grafo, se realizó de la siguiente manera ([Fig. A.4](#)):

1. Recepción de un parámetro de filtro para el grafo (ej.: periodos a analizar -ver [Resultados - Análisis estático - Configuración](#) )
2. Aplicación del filtro a la tabla “autores\_articulos” para la preparación de un data.frame que representa los datos necesarios para crear el grafo.
3. Utilización de la [librería igraph](#), para la creación de un grafo bipartito de nodos autores y nodos artículos.
4. Utilización de la [librería igraph](#) para la realización de la proyección sobre los nodos autores, pasando a significar “*estos autores están juntos por haber realizado un artículo en conjunto*”
5. Adjudicación de propiedades a los nodos autores y las aristas relaciones de coautoría.

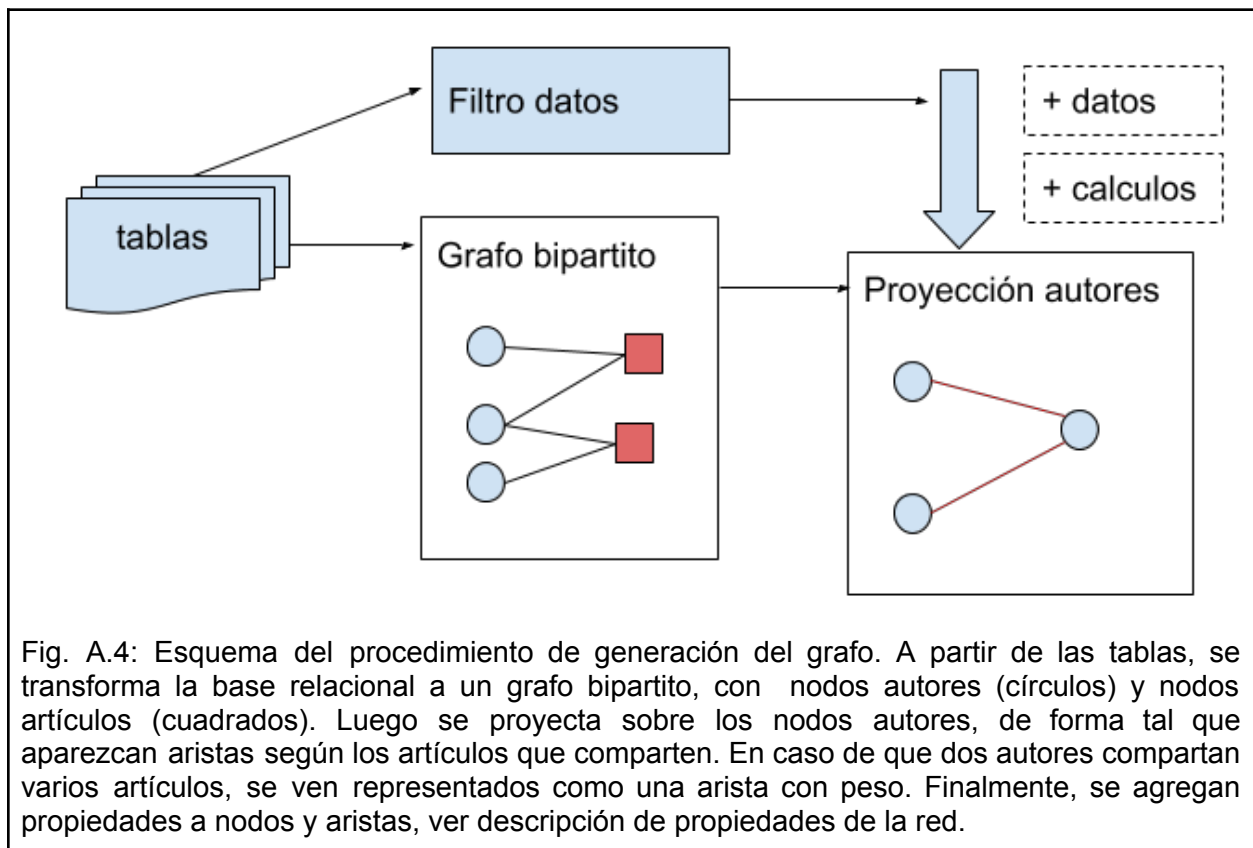


Fig. A.4: Esquema del procedimiento de generación del grafo. A partir de las tablas, se transforma la base relacional a un grafo bipartito, con nodos autores (círculos) y nodos artículos (cuadrados). Luego se proyecta sobre los nodos autores, de forma tal que aparezcan aristas según los artículos que comparten. En caso de que dos autores compartan varios artículos, se ven representados como una arista con peso. Finalmente, se agregan propiedades a nodos y aristas, ver descripción de propiedades de la red.

### Descripción de las propiedades de los nodos y aristas

En este apartado se detallan las propiedades disponibles para los nodos y las aristas. Recordemos que los nodos representan a los Autores, y las aristas representan la relación de coautoría entre los autores que unen. Cada uno de estos, tienen características propias como en el caso de autores pueden ser el nombre y apellido del autor, y en el caso de las aristas pueden ser los autores que la generan. En la [Tabla A.4](#) podemos ver las propiedades de los autores y en la [Tabla A.5](#) de las aristas.

Debemos recordar que la generación del grafo se realiza luego de aplicar filtros a los datos, por lo tanto algunos valores son calculados según los filtros aplicados en la sección 2 de la aplicación web.

### Propiedades de los Nodos

Estas propiedades están disponibles para cada nodo autor ([Tabla A.4](#)).

Tabla A.4: Propiedades de los Nodos

nombre	descripción	ejemplo
name id	Identificador generado para poder seguir mejor los nodos y los cruces de los mismos.	"a0001"
id_old	Identificador inicial al generar el grafo	"4"
fuerza_colaboracion	Es la suma de fuerza de colaboración del autor en todos los artículos en los que participa en el grafo actual.	0.1666667
cant_autores	Cantidad de autores con los que se relaciona en el grafo actual	6
size	Tamaño que va a ser visualizado. Sirve para que se note la diferencia visual. La fórmula aplicada es: $10 + (\text{fuerza\_colaboracion} * 10)$	11.66667
label	Etiqueta del nodo, con el nombre del autor. Se conservó la propiedad como "label" porque no siempre se puede configurar por parámetros a funciones. Como el label es default en muchos lados, se uso esto para simplificar.	"Abraham I. Kohan"
anio	Propiedad que dice el año de participación y cuántos artículos escribió	"1996[1]"

### Propiedades de las Aristas

Estas propiedades están disponibles para cada arista que representa la relación de coautoría entre autor 1 y autor 2 ([Tabla A.5](#)).

Tabla A.5: Propiedades de las Aristas

nombre	descripción	ejemplo
id	Id generado para la relación	"a0001--a0016"
weight	Peso de la relación, resultante de la proyección del grafo bipartito sobre los nodos autores. Si hay 2 artículos compartidos , el peso es 2	1
color	Código hexadecimal del color a aplicar al edge en la visualización. Se genera a partir fuerza_colaboracion en 3 separaciones. ( leves , medias , fuertes )	"#1B9E77"
width	Ancho de la arista en la visualización.	1

	Se genera a partir fuerza_colaboracion en 3 separaciones. ( leves , medias , fuertes )	
fuerza_colaboracion	La suma de fuerza de colaboración de los artículos en esa relación. Ej: Articulo 1: fuerza_colaboracion = 0.5 Articulo 2: fuerza_colaboracion = 0.33 Total: fuerza_colaboracion = 0.5+0.33 = 0.83	0.1666667
anios	Años en los que fueron realizados los artículos de la relación y cantidad de artículos en esos años	"1996[1]" "2001[1];2005[1];2007[1]" "1996[2];2005[1];"
autores	Autores entre los cuales es la relación	"Abraham I. Kohan - Alicia M. Arrayago"
autor1_label	Nombre del autor 1	"Abraham I. Kohan"
autor2_label	Nombre del autor 2	"Alicia M. Arrayago"

## Anexo: sección preguntas

Las preguntas que se hicieron en este análisis están en la [Tabla A.6](#). Si el usuario desea expandir sus preguntas puede exportar los datos desde la aplicación.

Tabla A.6: Preguntas Análisis Exploratorio.

Tabla	Pregunta
Autores	¿Cuántos autores hay en total de la base?
Artículos	¿Cuántos artículos hay en total de la base?
Artículos	Cantidad de artículos por año
Artículos	Cantidad de artículos por sección (tiene sentido cuando no se filtra por sólo Trabajos originales)
Artículos	Cantidad de artículos por año por edición
Autores - Artículos	¿Cuántas autorías hay en total de la base?
Autores - Artículos	¿Quiénes son los 20 autores con más artículos?
Autores - Artículos	¿Cuál es la frecuencia de artículos por autor?
Autores - Artículos	¿Cuáles son los 20 artículos que tienen más autores?
Autores - Artículos	¿Cuál es la frecuencia de autores por artículo?
Autores - Artículos	Cantidad de autores por artículo por periodo agrupados (1, 2 a 4, 5 a 7, 8 o +)

## Anexo: métricas de red disponibles

Tabla A.7. Descripción de las métricas disponibles en la sección Métricas de la red.

Métrica	Descripción
# Aut.	Cantidad de autores en la red
# Art.	Cantidad de artículos en la red
# Contrib.	Cantidad de contribuciones en la red: la suma de las cantidades de autores por artículo.
# Rel.	Cantidad de relaciones en la red: por cada artículo en conjunto se genera una relación entre los diferentes autores. La suma por artículo de: $n_{\text{autores}} * (n_{\text{autores}} - 1) / 2$
# Media de Art. por Autor	Media de artículos por autor
# Media de Aut. por Art.	Media de autores por artículo
# de Comp.	Cantidad de componentes en la red. Un componente es un grupo de nodos conectados.
# Aut. en Comp. Grande	Cantidad de autores en el componente más grande, es decir, la mayor cantidad de autores conectados.
Comp. más Grande (%)	Porcentaje del componente más grande. (respecto del total de nodos en la red)
Densidad	Densidad de la red: La densidad de un grafo es la relación del número de aristas y el número de aristas posibles.
Dist. Media	Distancia media entre nodos conectados en la red.
Diámetro	Diámetro de la red: Es la geodésica más larga. La geodésica es la mínima longitud para conectar 2 puntos.
Cliqué más grande (%)	Cliqué más grande expresado en porcentaje $n_{\text{autores\_clique\_maximo}} / n_{\text{autores}} * 100$ , recordemos que un clique es un subgrafo completo dentro de la red en cuestión. Completo significa que todos los que son parte de ese subgrafo están conectados entre sí. Esta medida representa que tanto de la red representa el máximo subgrafo completo
Tran.	Transitividad: es la probabilidad media que dos coautores también sean coautores con otro en común.
Asort.	Asortatividad: es el coeficiente de correlación de Pearson entre los grados de los nodos adyacentes en la red.



## Anexo: métricas de nodos

Tabla A.8. Descripción de las métricas de los nodos.

Columna	Descripción
Autor	El autor representado por ese nodo.
Fuerza Colaboración	<i>(Collaboration Strength)</i> Suma total de la fuerza de colaboración por cada artículo del autor. Recordemos fuerza de colaboración = 1 artículo / cantidad de autores. Fuerza de colaboración total: suma de fuerza colaboración de cada artículo que participó el autor
Grado	<i>(Degree)</i> El grado del nodo: la cantidad de conexiones directas del nodo.
Intermediación	<i>(Betweenness)</i> La importancia de un nodo según su participación en caminos para llegar de un nodo a otro.
Centralidad Autovector	<i>(Eigen centrality)</i> Además de la cantidad de conexiones valora la calidad de las mismas. Si un nodo se conecta con otros nodos con gran centralidad de autovector, se posiciona mejor.
Cercanía	<i>(Closeness)</i> Es la media de la distancia geodésica entre un nodo y el resto de los nodos. La distancia medida en cantidad de nodos participantes de los caminos. Se recomienda utilizarla cuando todos los nodos están conectados.
<i>Page Rank</i>	Es una variante de centralidad autovector.
# Triángulos	<i>(# Triangles)</i> Cantidad de triángulos de los que participa el autor. Un triángulo son 3 nodos completamente conectados.

## Anexo: Propiedades modelos

Tabla A.9. Propiedades de la red que pueden analizarse en esta subsección.

Propiedad	Razón de interés
Transitividad	De utilidad para diagnosticar mundo pequeño.
Diámetro	De utilidad para diagnosticar mundo pequeño.
Distancia camino medio	De utilidad para diagnosticar mundo pequeño.
Valor Alfa de la distribución de ley de potencia	De utilidad para diagnosticar Libre Escala. El exponente de la función para ajustar la distribución de ley de potencia
Verificación si alfa es mayor a 1	De utilidad para diagnosticar Libre Escala. Es un resultado de aplicar la función de ajuste a la ley de potencia según los parámetros.
Valor del <i>test</i> de Kolmogorov-Smirnov para ley de potencia	De utilidad para diagnosticar Libre Escala. Es un resultado de aplicar la función de ajuste a la ley de potencia según los parámetros.
Significancia del <i>test</i> Kolmogorov-Smirnov	De utilidad para diagnosticar Libre Escala. Es un resultado de aplicar la función de ajuste a la ley de potencia según los parámetros.

## Anexo: Mapa de la Interfaz de usuario

Cuadro A.2 mapa de navegación de interfaz de usuario

- Consideraciones
  - General
  - Análisis Exploratorio de datos
    - Autores
      - Datos Crudos
      - Composición
      - Resumen
      - Preguntas
    - Artículos
      - Datos Crudos
      - Composición
      - Resumen
      - Preguntas
    - Autores y Artículos
      - Datos Crudos
      - Composición
      - Resumen
      - Preguntas
- Análisis Estático
  - Configuración
    - Semilla
    - Selección periodos de análisis
    - Selección de nodo por nombre
    - Disposición de la visualización de red
    - Comunidades
      - Selección algoritmo
      - Selección visualización de comunidades
  - Visualización de coautoría
    - Red
      - Filtrado por aristas
    - Biofabric
    - Heatmap
  - Artículos Asociados
    - Datos del Nodo o Relación
    - Subred de coautoría
    - Heatmap de coautoría
    - Listado de artículos asociados
  - Estructura Red
    - Listado de Métricas de Red
    - Métricas de Nodos
      - Listado de Métricas de Nodos
      - Distribución de grado
      - Distribución de Grado ( log / log )
      - Histograma variables
      - Densidades variables
    - Detalle de Componentes
      - Listado de métricas de Estructura de los componentes
      - Listado de Autores por Componente
      - Detalles del Componente seleccionado
        - Artículos Analizados
        - Stopwords ES
        - Stopwords EN
        - Palabras analizadas
        - Palabras analizadas cantidad
        - Gráfico palabras analizadas cantidad
          - configuración de cantidad de palabras a considerar
    - Detalle nodos del componente seleccionado
      - Métricas de Nodos
      - Distribución de grado
      - Distribución de Grado ( log / log )
      - Histograma variables
      - Densidades variables
- Comparación con modelos

- Datos Red para comparación vs modelos
- Modelos
  - Red aleatoria
    - Párametros
    - Histogramas
    - Datos simulación
  - Mundo pequeño
    - Párametros
    - Histogramas
    - Datos simulación
  - Libre escala
    - Párametros
    - Histogramas
    - Datos simulación
- Validaciones
  - Mundo Pequeño
  - Libre Escala
- Comunidades
  - Detalles de la ejecución del algoritmo de comunidad seleccionado
  - Listado de métricas de Estructura de las comunidades
  - Listado de Autores por Comunidad
  - Detalles de la Comunidad seleccionada
    - Artículos Analizados
    - Stopwords ES
    - Stopwords EN
    - Palabras analizadas
    - Palabras analizadas cantidad
    - Gráfico palabras analizadas cantidad
      - configuración de cantidad de palabras a considerar
  - Detalle nodos de la comunidad seleccionada
    - Metricas de Nodos
    - Distribución de grado
    - Distribución de Grado ( log / log )
    - Histograma variables
    - Densidades variables
- Análisis Temporal
  - Visualización dinámica - Acumulado
  - Métricas de Red en el tiempo por año
  - Métricas de Red en el tiempo, valores acumulados
  - Métricas de Nodos, principales N en el tiempo
  - Métricas de Nodos, principales N en el tiempo, valores acumulados
- Contacto

## Anexo: Tablas numéricas complementarias para la Hipótesis 2

De manera complementaria se presentan tablas, para verlo desde un punto de vista numérico ([Tabla A.10](#), [Tabla A.11](#) y [Tabla A.12](#)), se puede ver una tabla a lo largo del tiempo de los primeros 5 puestos. Se toman en cuenta 3 indicadores por cada puesto: ***mantiene*** con valor 0 para no y 1 para sí, respecto de si el autor se mantiene en el puesto respecto del tiempo anterior.

***es\_fundador***, para representar si el autor estuvo en la primera edición de la revista o no, con valor 0 para representar que NO es de la primera edición, y valor 1 para representar que sí es de la primera edición. Finalmente la variable ***valor***, que representa el valor obtenido de la métrica analizada para el puesto en ese año.

En la última columna se muestra ***ratio\_fundador*** que representa del total de los años analizados, cuánto corresponde a si fue fundador o no. Para esta métrica, un valor de 1 equivale a 100 %, significa que en todos los años evaluados fue un autor fundador el que estuvo en esa posición. Si fuese menor significa que hubo momentos en los cuales el autor no fue fundador.

Tabla A.10: Valores del ranking de Fuerza de Colaboración. Análisis de si se mantiene y si es fundador.

métrica	Fuerza de colaboración														
#	1			2			3			4			5		
variable	mantiene	es_fundador	valor	mantiene	es_fundador	valor	mantiene	es_fundador	valor	mantiene	es_fundador	valor	mantiene	es_fundador	valor
1996	0	1	1.333	0	1	1.000	0	1	0.833	0	1	0.783	0	1	0.617
1999	0	1	2.000	0	1	1.995	0	1	1.917	1	1	1.383	0	1	1.343
2001	0	1	2.662	0	1	2.167	0	1	2.000	0	1	1.917	0	1	1.726
2002	0	1	3.333	0	1	2.912	0	1	2.167	1	1	1.917	1	1	1.726
2003	1	1	3.833	1	1	2.912	1	1	2.167	0	1	1.976	0	1	1.917
2004	1	1	3.833	0	1	3.417	0	1	2.912	0	1	2.167	0	1	1.976
2005	1	1	4.333	1	1	3.417	1	1	2.912	1	1	2.167	1	1	2.143
2006	1	1	4.667	1	1	3.417	1	1	2.912	1	1	2.500	0	0	2.310
2007	1	1	4.667	1	1	3.417	1	1	2.912	1	1	2.500	0	1	2.411
2008	1	1	4.667	1	1	3.750	1	1	3.112	1	1	2.500	1	1	2.411
2009	1	1	4.667	1	1	3.750	1	1	3.112	1	1	2.833	1	1	2.536
2010	1	1	4.667	1	1	4.450	1	1	3.112	1	1	2.833	0	1	2.583
2011	1	1	4.667	1	1	4.450	1	1	3.362	0	1	3.083	0	1	2.833
2012	1	1	5.333	1	1	4.450	1	1	4.112	1	1	3.583	1	1	2.833
2013	1	1	5.333	1	1	4.450	1	1	4.279	1	1	3.583	1	1	2.833
2014	1	1	5.583	0	1	4.529	0	1	4.450	1	1	3.583	1	1	2.833
2015	1	1	6.250	1	1	4.529	1	1	4.450	1	1	3.583	1	1	3.167
2016	1	1	6.250	0	1	4.783	0	1	4.529	1	1	3.583	1	1	3.167
ratio fundador	1.00			1.00			1.00			1.00			0.94		

Para Fuerza de Colaboración, se aprecia que salvo el quinto puesto del ranking, se mantienen los fundadores en el tiempo; y en el quinto puesto el 94% del tiempo son autores fundadores.

Tabla A.11: Valores del ranking de Grado. Análisis de si se mantiene y si es fundador.

métrica	Grado											
#	1			2			3			4		
variable	mantiene	es_fundador	valor	mantiene	es_fundador	valor	mantiene	es_fundador	valor	mantiene	es_fundador	valor
1996	0	1	13.000	0	1	9.000	0	1	6.000	0	1	5.000
1999	0	1	23.000	0	1	21.000	0	1	13.000	0	1	10.000
2001	1	1	24.000	0	1	15.000	0	1	13.000	0	1	12.000
2002	1	1	27.000	0	1	24.000	0	1	15.000	0	1	13.000
2003	1	1	27.000	1	1	24.000	1	1	15.000	1	1	13.000
2004	1	1	27.000	1	1	24.000	1	1	15.000	1	1	13.000
2005	1	1	27.000	1	1	26.000	1	1	18.000	1	1	16.000
2006	1	1	27.000	1	1	26.000	1	1	18.000	1	1	16.000
2007	0	1	31.000	0	1	27.000	1	1	23.000	0	0	18.000
2008	0	1	34.000	0	1	33.000	0	1	32.000	0	1	31.000
2009	1	1	34.000	0	1	33.000	0	1	32.000	0	0	30.000
2010	1	1	34.000	1	1	33.000	1	1	32.000	0	0	30.000
2011	1	1	36.000	0	1	34.000	0	1	33.000	0	1	32.000
2012	1	1	36.000	1	1	34.000	1	1	33.000	1	1	32.000
2013	0	1	43.000	0	1	42.000	0	1	39.000	0	1	38.000
2014	0	1	44.000	0	1	43.000	0	1	42.000	1	1	40.000
2015	0	1	49.000	0	1	44.000	1	1	42.000	1	1	40.000
2016	1	1	49.000	1	1	44.000	1	1	42.000	1	1	40.000
ratio fundador	1.00			1.00			1.00			0.83		

Para Grado, se aprecia que salvo el cuarto y quinto puestos del ranking, se mantienen los fundadores en el tiempo; y en el cuarto puesto el 83% del tiempo son fundadores y el 94% del tiempo son fundadores para el quinto puesto ([Tabla A.11](#)).

Tabla A.12: Valores del ranking de intermediación. Análisis de si se mantiene y si es fundador.

métrica	intermediación														
#	1			2			3			4			5		
variable	mantiene	es_fundador	valor	mantiene	es_fundador	valor	mantiene	es_fundador	valor	mantiene	es_fundador	valor	mantiene	es_fundador	valor
1996	0	1	26.500	0	1	12.533	0	1	10.333	0	1	5.000	0	1	2.000
1999	0	1	658.500	0	1	591.979	0	1	210.500	0	0	158.667	0	1	145.794
2001	0	1	736.018	0	1	721.052	1	1	259.614	1	0	180.000	1	1	154.595
2002	0	1	1601.177	0	1	1064.907	0	0	381.683	0	1	291.614	0	1	259.465
2003	1	1	1656.177	1	1	1085.463	1	0	411.163	1	1	293.614	1	1	292.673
2004	1	1	1691.292	1	1	1069.504	1	0	418.163	1	1	300.500	1	1	297.212
2005	1	1	1853.668	1	1	1297.290	1	0	542.122	0	1	482.103	1	1	424.005
2006	1	1	1885.902	1	1	1309.977	1	0	550.622	1	1	540.090	1	1	430.032
2007	1	1	2272.349	1	1	1777.675	0	1	723.950	0	0	635.622	1	1	490.295
2008	0	1	1990.606	0	1	1398.516	1	1	916.141	0	1	727.105	1	1	490.767
2009	1	1	2087.649	1	1	1452.230	1	1	1027.498	1	1	758.956	1	1	553.594
2010	1	1	2438.818	1	1	2161.094	0	0	1368.792	0	1	1141.215	0	1	1096.000
2011	1	1	2522.497	1	1	2225.381	1	0	1502.125	0	1	1236.000	0	1	1179.789
2012	0	1	2777.240	0	1	2715.160	1	0	1985.458	1	1	1749.000	1	1	1252.436
2013	1	1	4124.605	1	1	2663.442	1	0	2554.467	1	1	2184.000	0	0	2029.277
2014	1	1	6355.776	1	1	3595.673	0	0	2223.364	0	0	2223.058	0	1	2011.820
2015	1	1	6614.722	1	1	3728.957	0	0	2435.435	0	1	2357.307	0	0	2332.303
2016	1	1	7059.152	1	1	3891.051	1	0	2503.777	0	0	2444.122	0	1	2401.666
ratio fundador	1.00			1.00			0.33			0.72			0.89		

Para Intermediación, el primero y segundo puesto son estables respecto de los fundadores. Sin embargo, el tercer puesto, tiene un 33% del tiempo autores fundadores. El cuarto puesto tiene autores fundadores un 72% del tiempo, y el quinto puesto tiene autores fundadores un 89% del tiempo ([Tabla A.12](#)).



## Anexo: Detalle periodos inferiores para Mundo pequeño y Libre Escala

Para ello, se utilizó la división especificada en la [Tabla A.13](#). Recordemos que los 20 años de la revista están repartidos en 18 años de publicaciones dado que en 1997, 1998 y 2000 no hubo ediciones.

Tabla 4.13: División de años de la revista para el análisis.

Período	desde	hasta	incluye	# artículos afectados	# autores afectados
1	1996	2001	1996,1999,2001	52	100
2	2002	2006	2002,2003,2004,2005,2006	35	81
3	2007	2011	2007,2008,2009,2010,2011	34	94
4	2012	2016	2012,2013,2014,2015,2016	50	143

### Mundo pequeño para el periodo 1996 - 2001

Podemos ver que para el periodo 1996-2001, La [Figura A.5](#) muestra los parámetros utilizados para realizar las simulaciones usando el modelo modelo Érdos-Rényi para el caso de mundo pequeño.

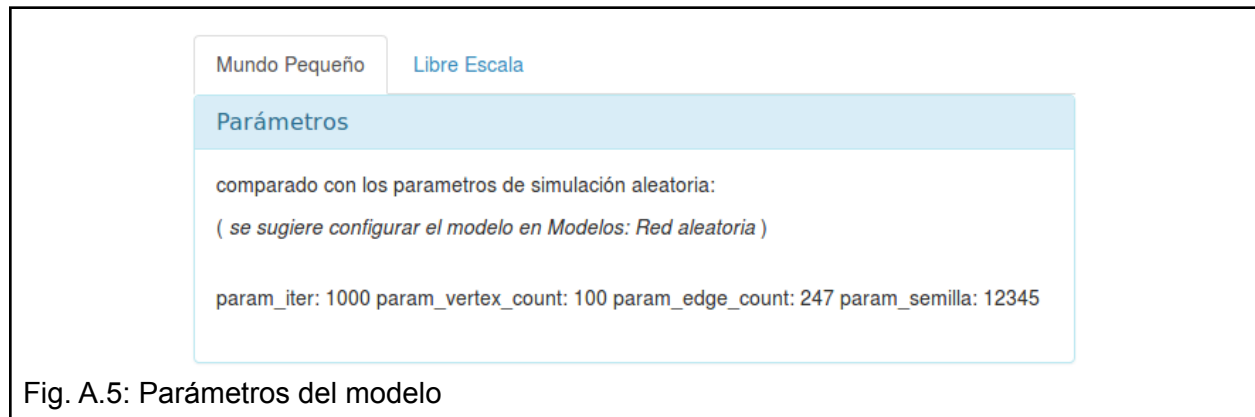
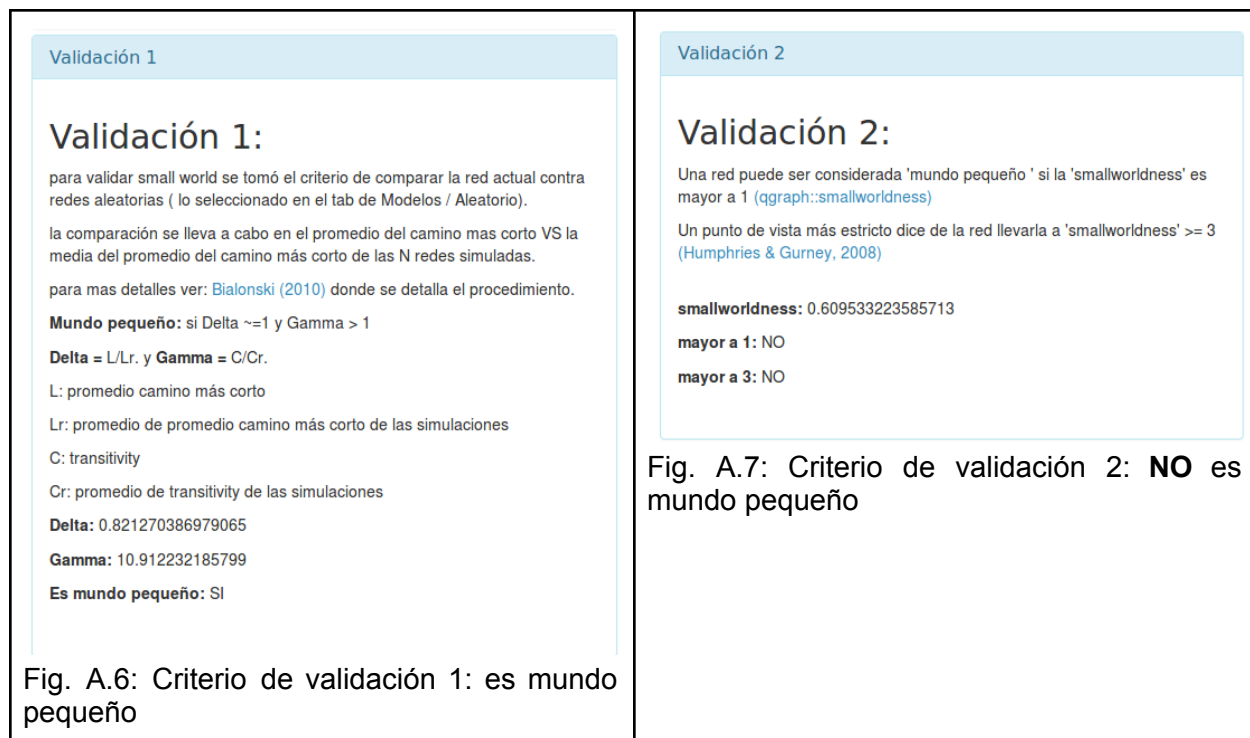
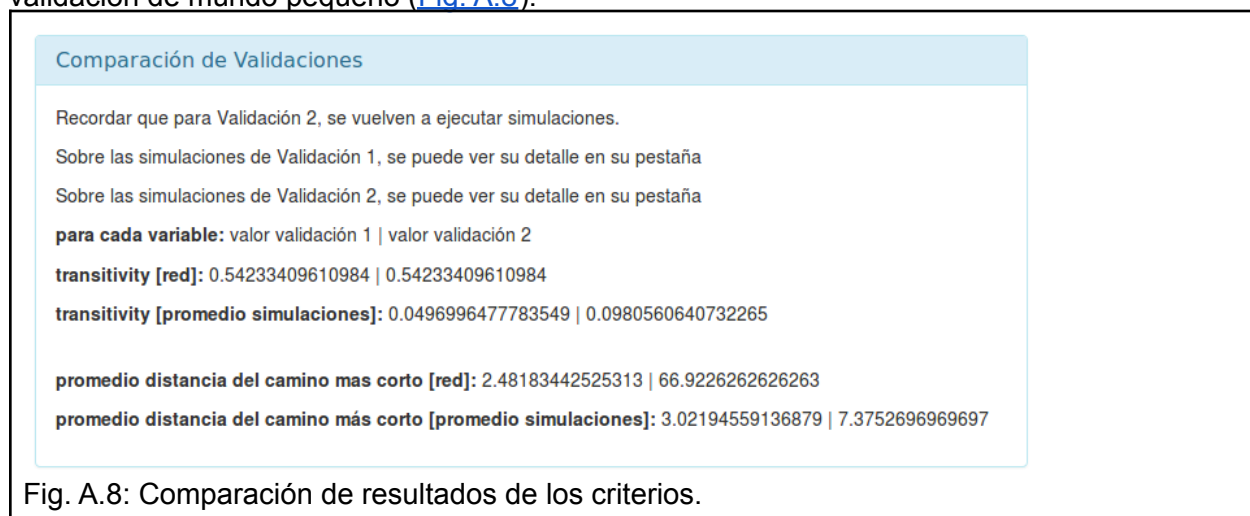


Fig. A.5: Parámetros del modelo

Los resultados determinan que la red es de mundo pequeño según el criterio 1, sin embargo para el criterio 2 no, considerando tanto el umbral laxo como el estricto ([Fig. A.6](#) y [Fig. A.7](#)).



Como se mencionó anteriormente, puede evaluarse la diferencia entre ambos enfoques de validación de mundo pequeño ([Fig. A.8](#)).



## Libre escala 1996 - 2001

Podemos ver que para el periodo 1996-2001, la red es de libre escala ([Fig. A.9](#)). Además, vemos que hay mayor probabilidad en grados menores a 10 y menor en grados mayores a 10 ([Fig. A.10](#)).

## Validaciones

Mundo Pequeño

Libre Escala

Una red Libre escala es una red en la cual su distribución de grado sigue la ley de potencia.

[Barabasi \(2015\)](#)

**Verificamos por la ley de potencia:**

Luego de ajustar los grados de la red a la ley de potencia, tenemos que verificar:

**El parámetro alfa de la función es mayor que 2?**

Exponente alfa: 3.4329176798469

El exponente es mayor a 2: TRUE

Los datos originales pueden ser obtenidos dada la distribución de ley de potencia obtenida.

P-Valores pequeños ( Menores a 0.05 ) Indican que el test rechaza la hipótesis

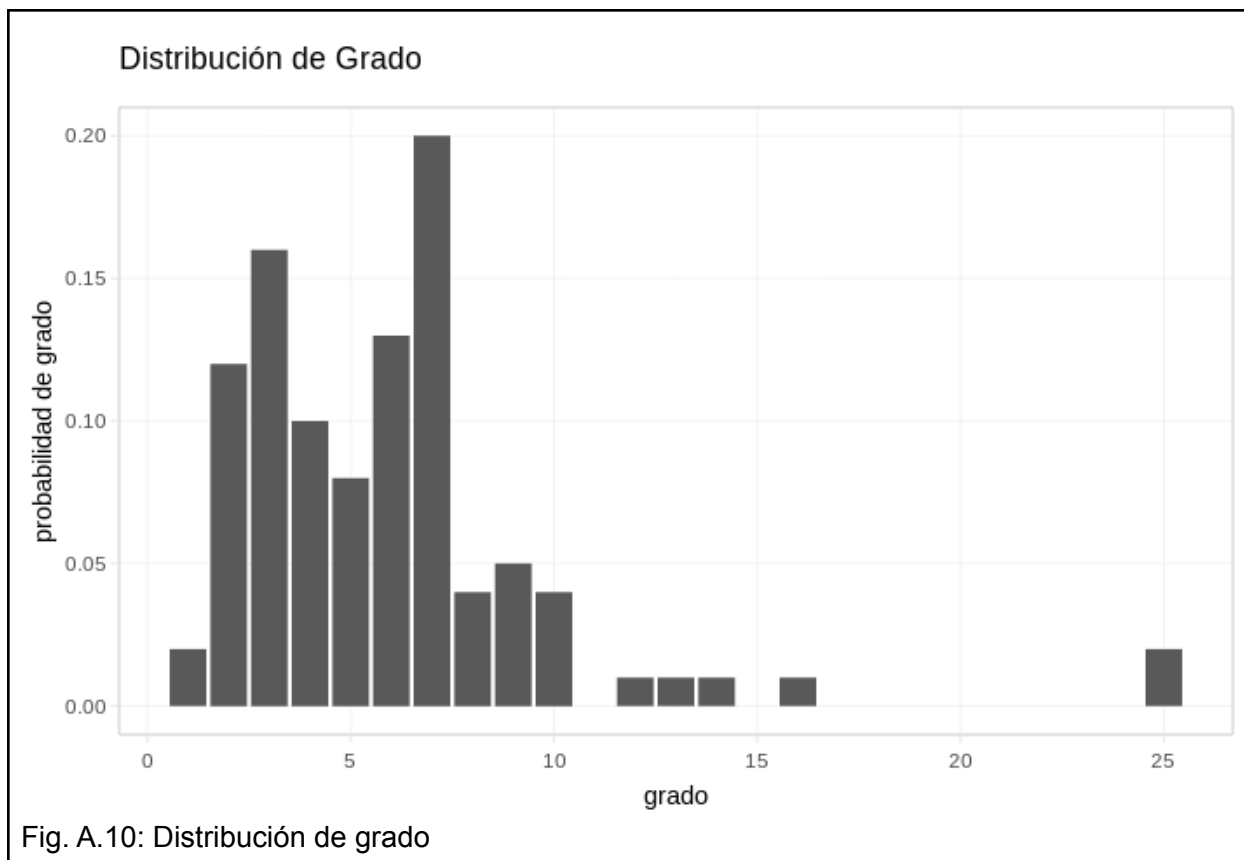
**El test de ajuste de Kolmogorov-Smirnov significativo?**

p-value test KS: 0.999544718619725

menor a 0.05: FALSE

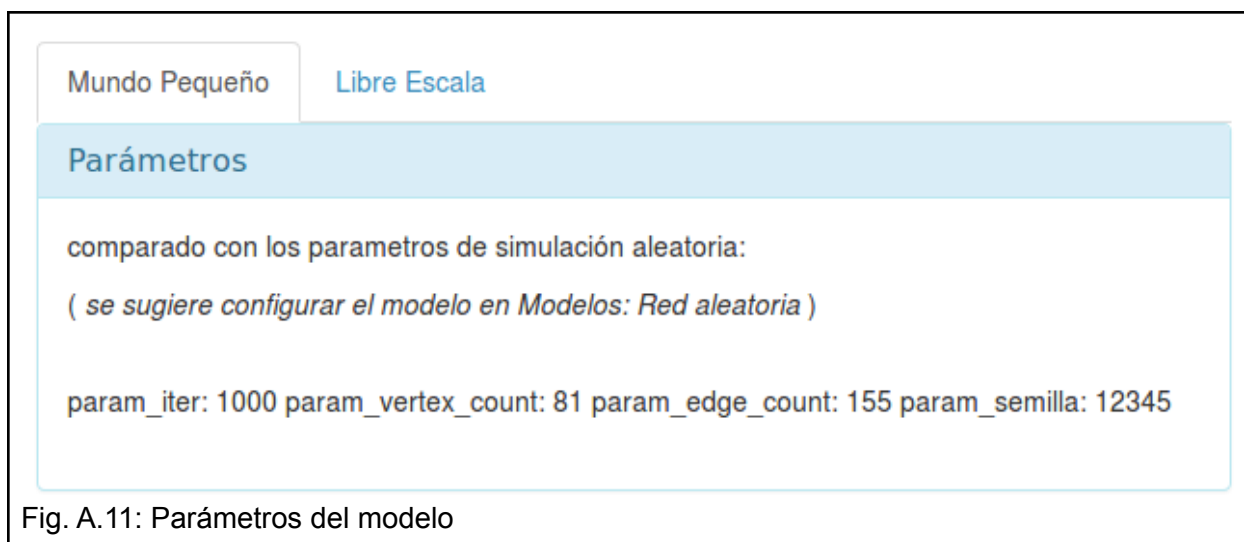
Es de Libre Escala.

Fig. A.9: resultado de comparación con modelos de libre escala. Es de libre escala.

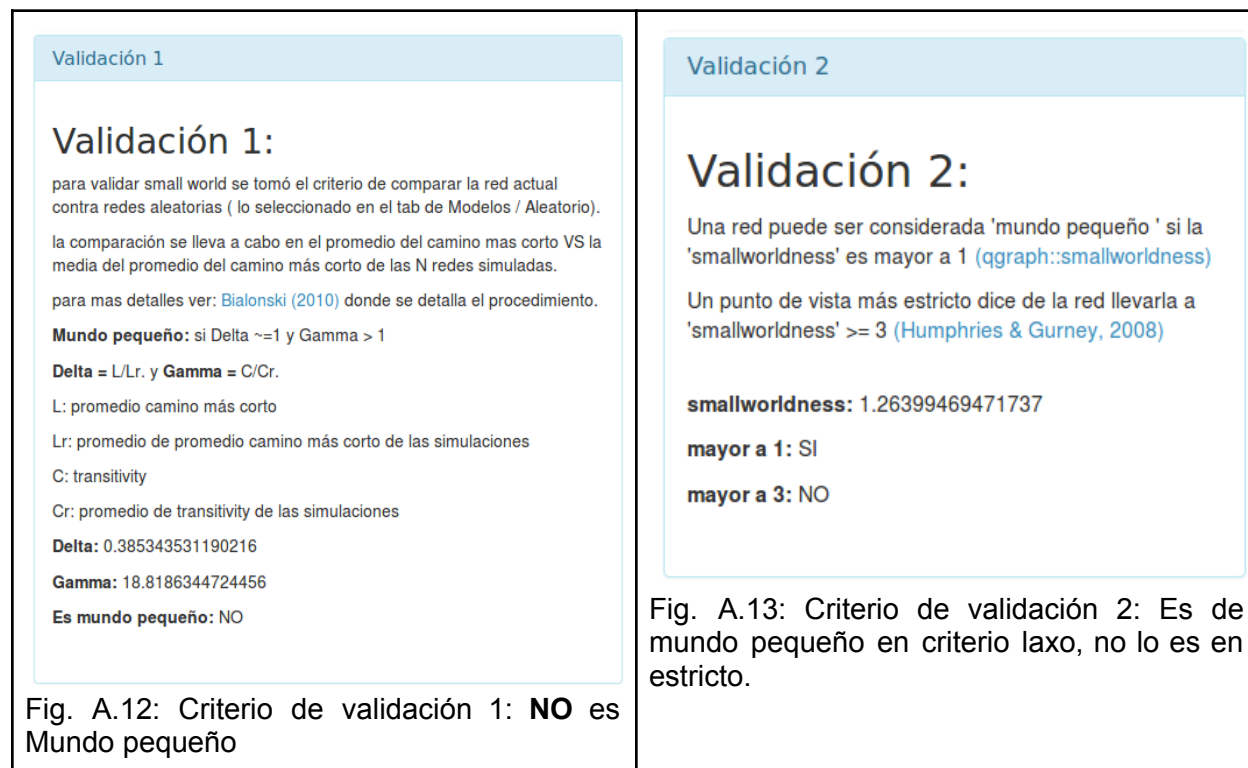


Mundo pequeño para el periodo 2002 - 2006

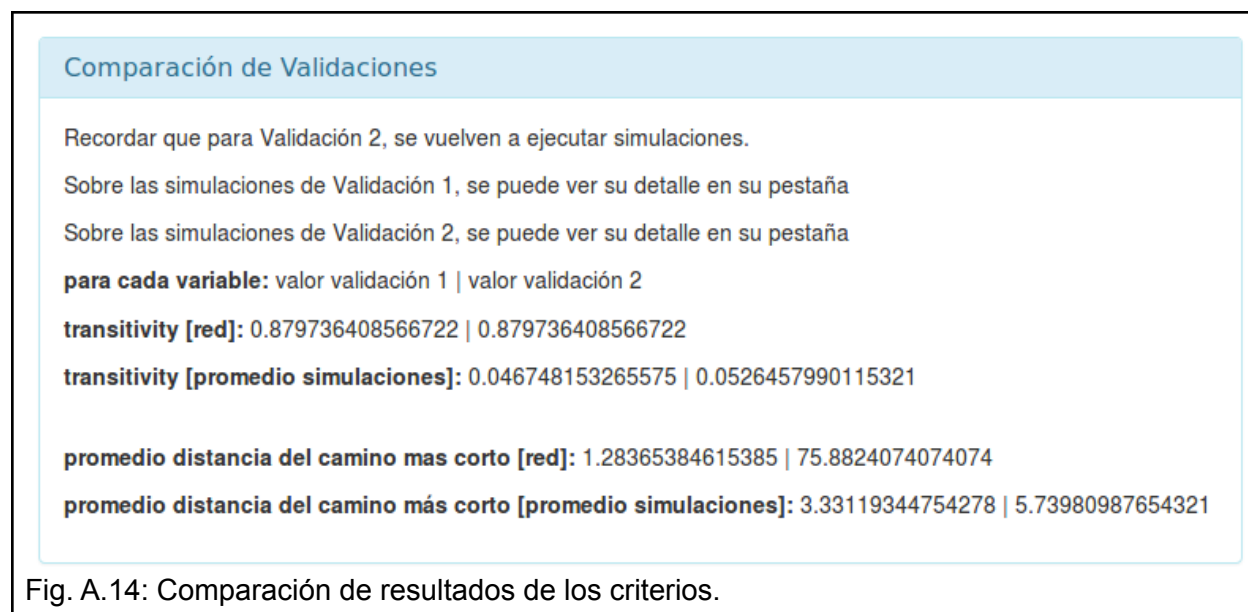
Los parámetros utilizados para esta ejecución se pueden verificar en la [Fig. A.11](#).



Podemos ver que para el periodo 2002-2006, la red no es de mundo pequeño según el criterio 1 (Fig. A.12). Sin embargo, para el criterio 2 sí cumple con esta característica bajo el criterio laxo pero no bajo el estricto (Fig. A.13).



La diferencia entre ambos enfoques de validación de mundo pequeño se observa en la [Fig. A.14](#).



## Libre escala 2002 - 2006

Podemos ver que para el periodo 2002-2006, la red es de libre escala ([Fig. A.15](#)), hay mayor probabilidad de grados bajos y bajas probabilidades de grados elevados. Sin embargo, en la distribución vemos que el grado 7 se destaca bastante ([Fig. A.16](#)).

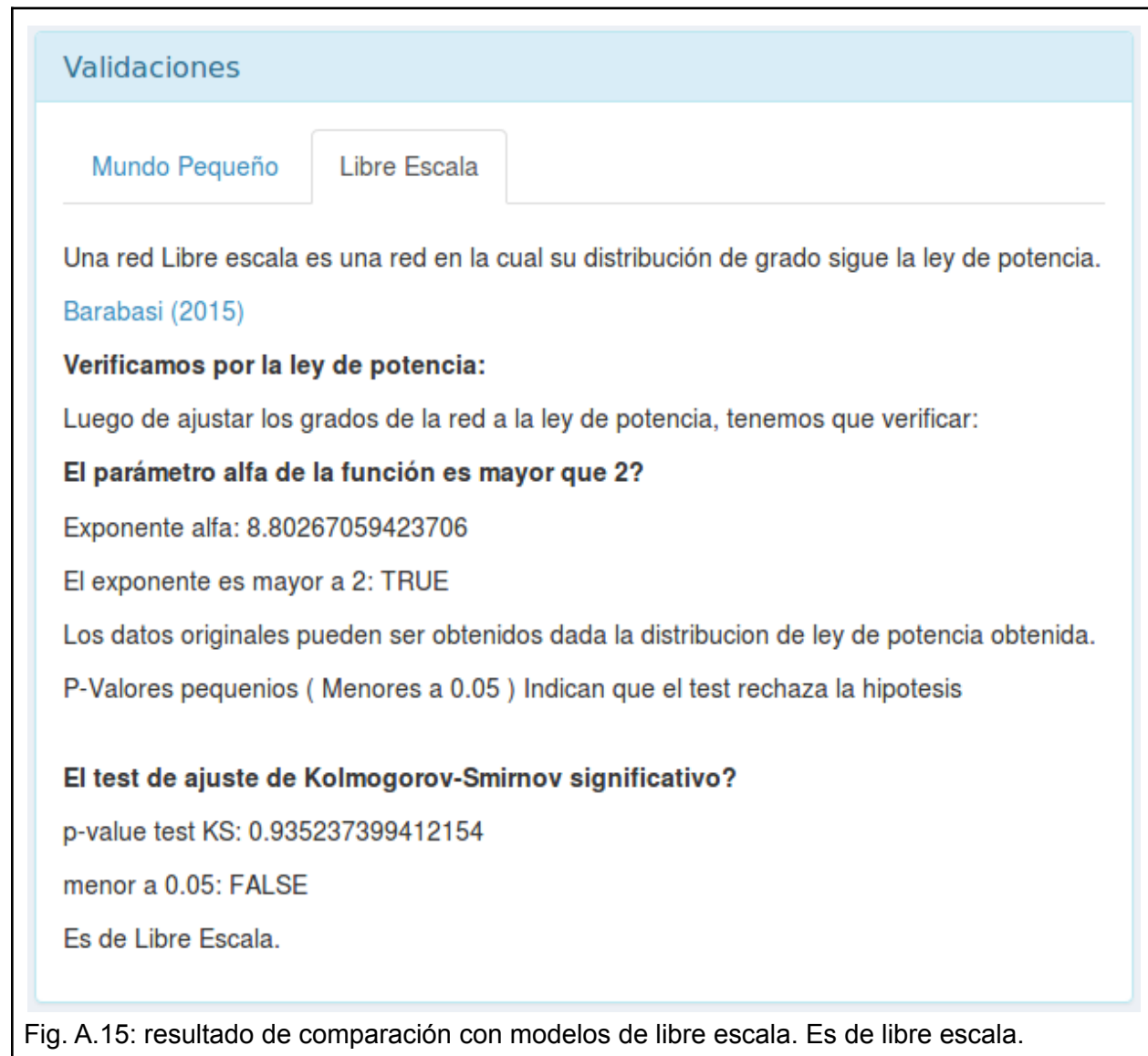
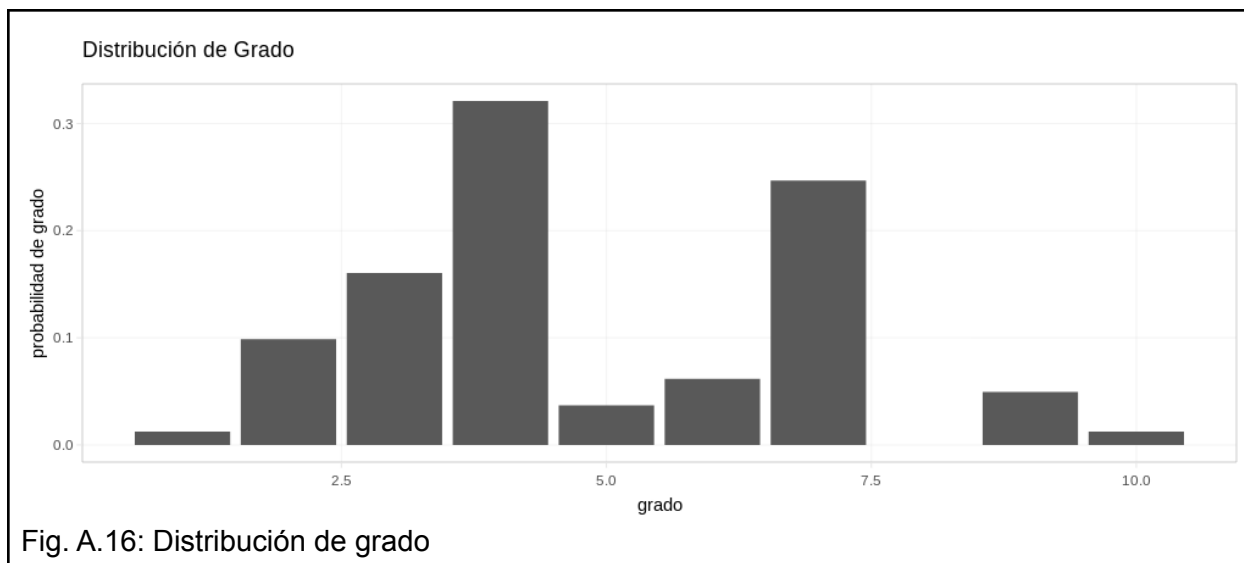
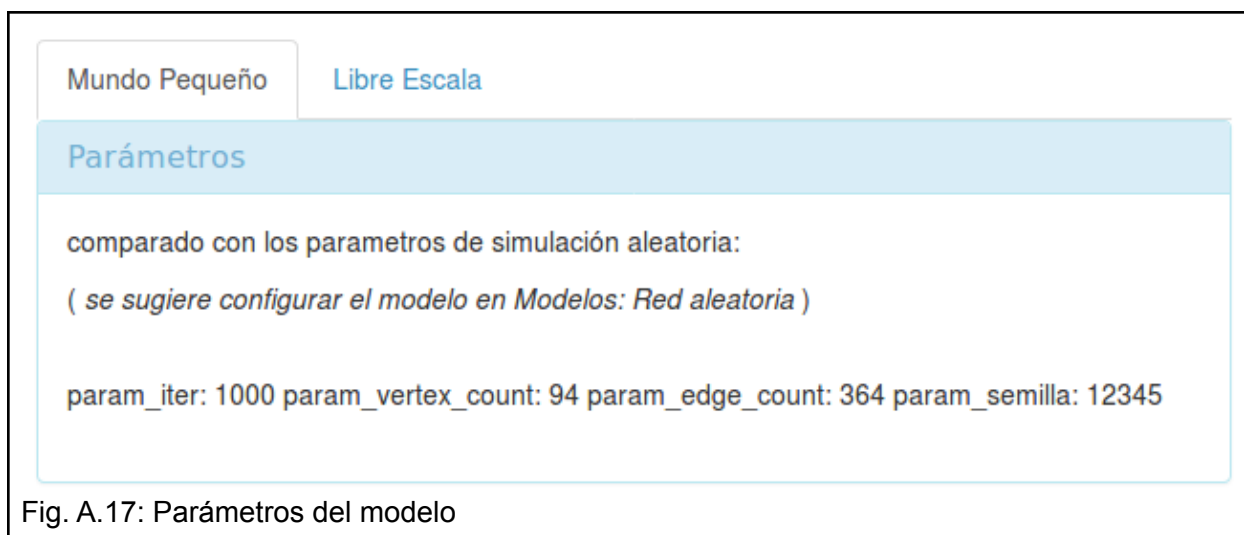


Fig. A.15: resultado de comparación con modelos de libre escala. Es de libre escala.

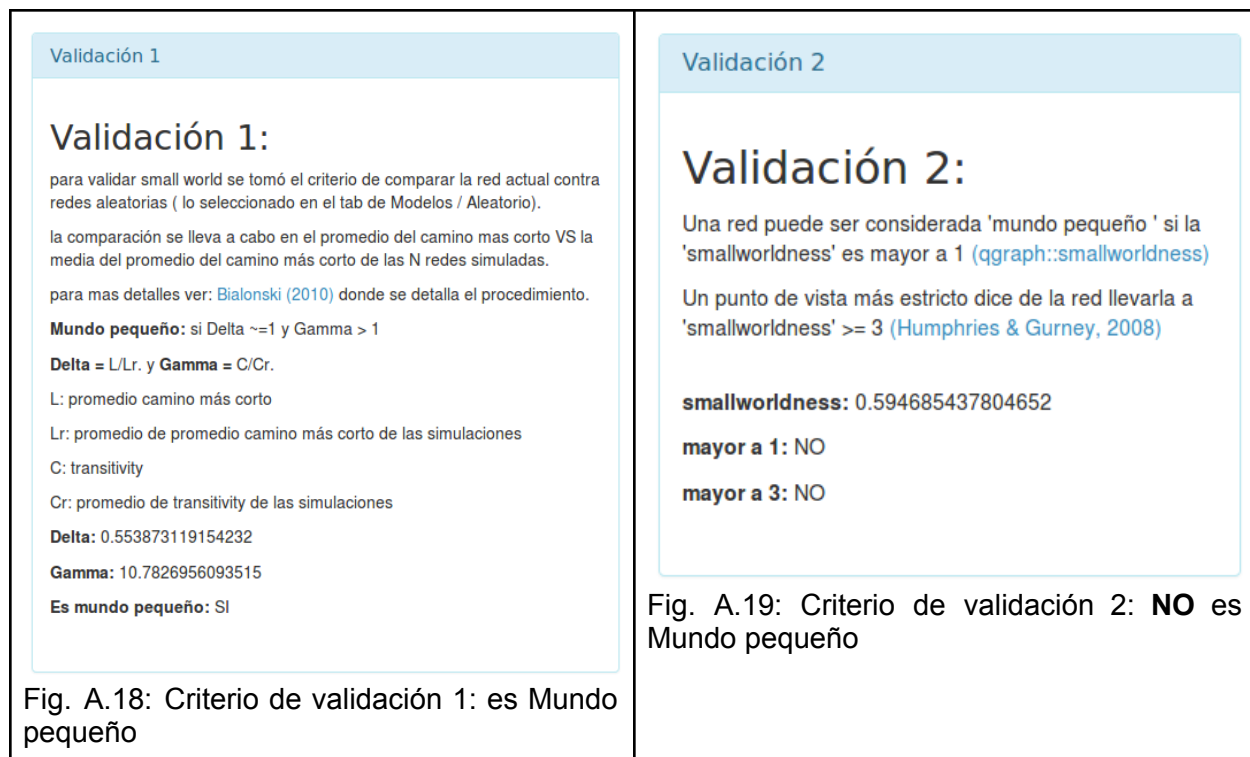


## Mundo pequeño para el periodo 2007 - 2011

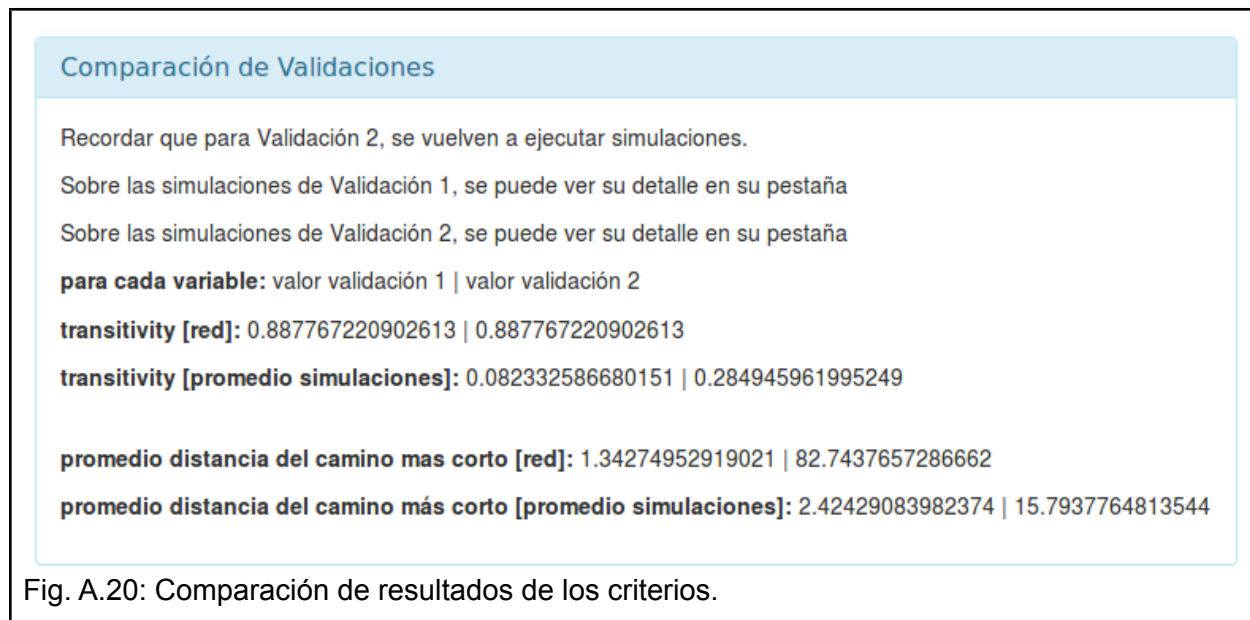
Los parámetros utilizados para esta ejecución se pueden verificar en la [Fig. A.17](#).



Podemos ver que para el periodo 2007-2011, la red es de mundo pequeño para el criterio 1, sin embargo para el criterio 2 no, considerando tanto el umbral laxo como el estricto ([Fig. A.18](#) y [Fig. A.19](#)).



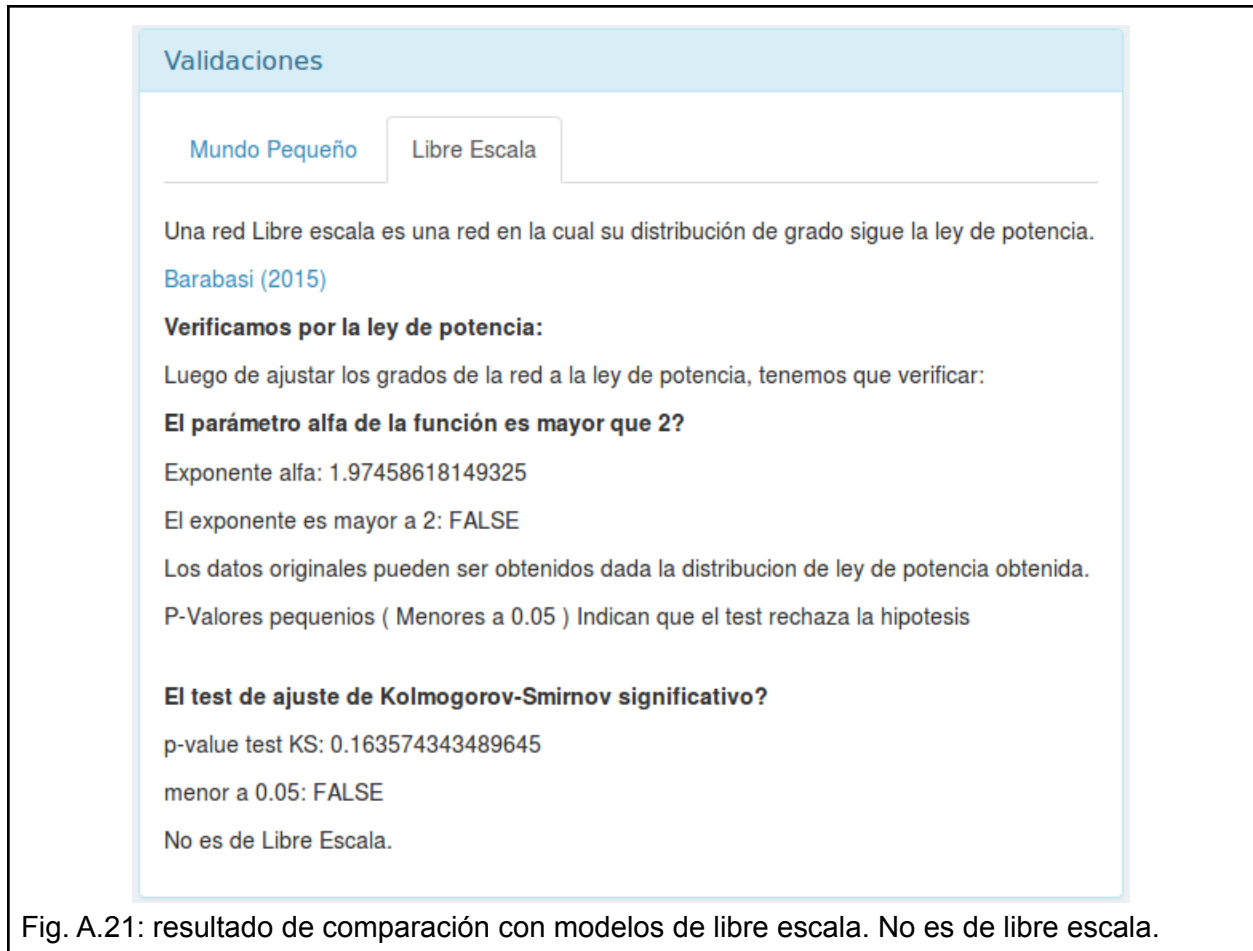
La diferencia entre ambos enfoques de validación de mundo pequeño se observa en la [Fig. A.20](#).



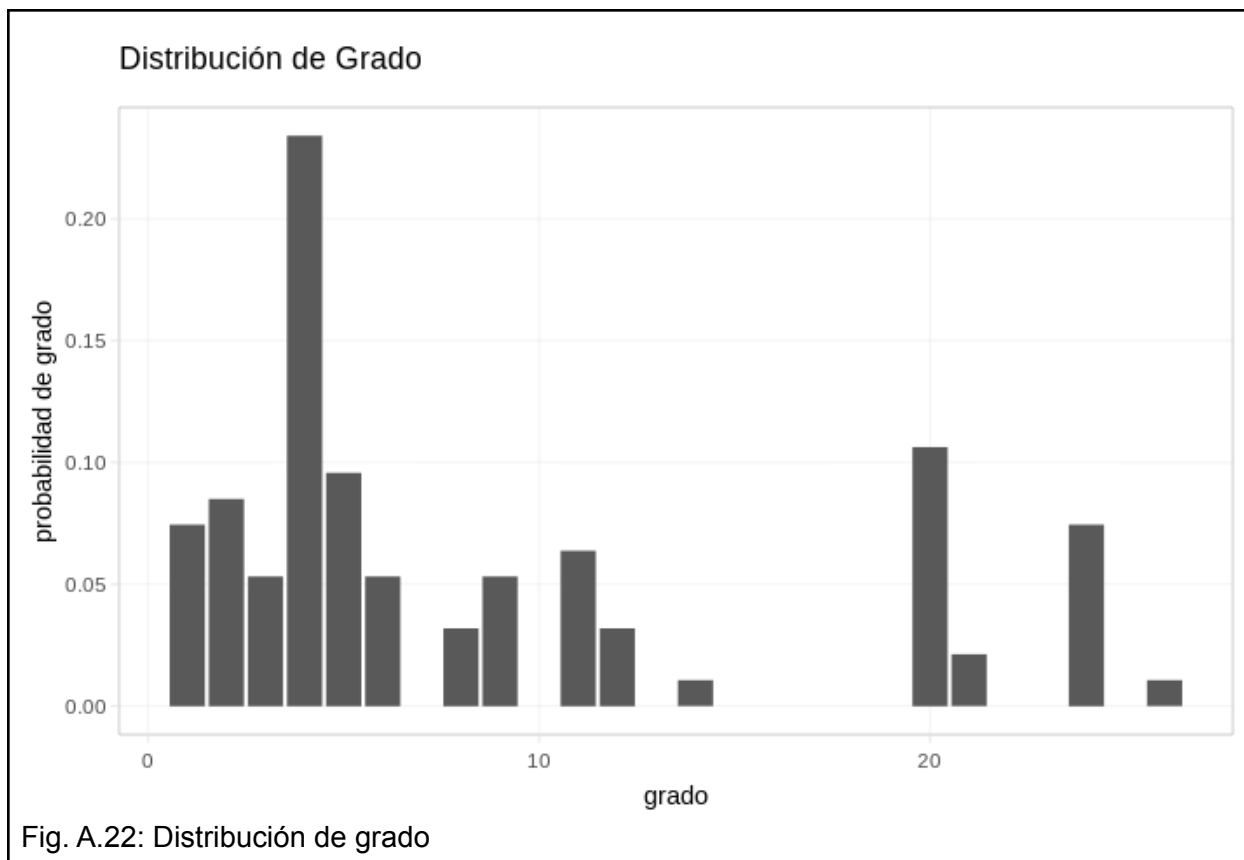
Libre escala 2007 - 2011

Para el periodo 2007 - 2011 vemos que la red no es de libre escala ([Fig. A.21](#)).



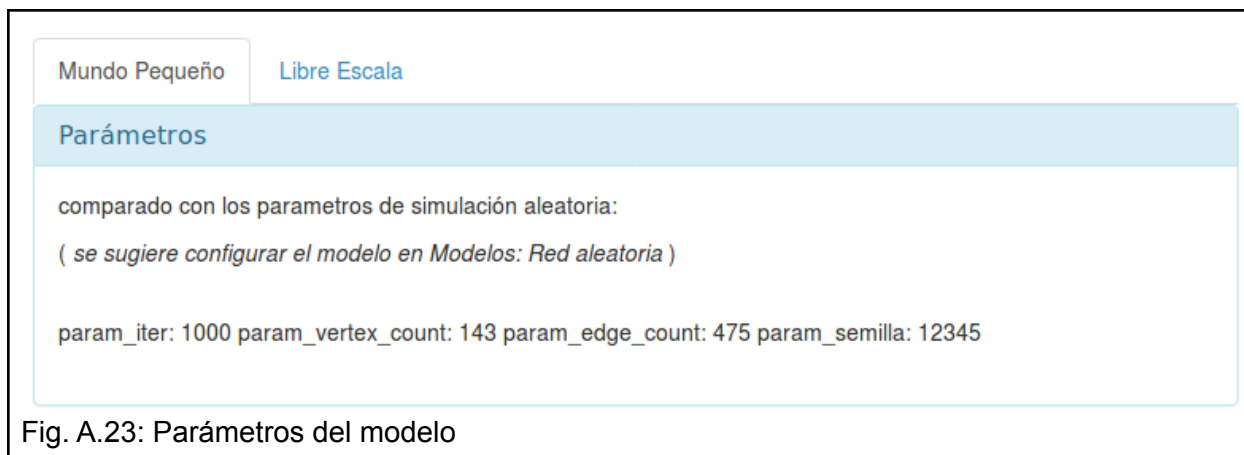


Teniendo en cuenta la distribución, esto lo vemos dado que en altos grados hay una probabilidad elevada, por ejemplo, como se observa en los grados 20 y 24 ([Fig. A.22](#)).

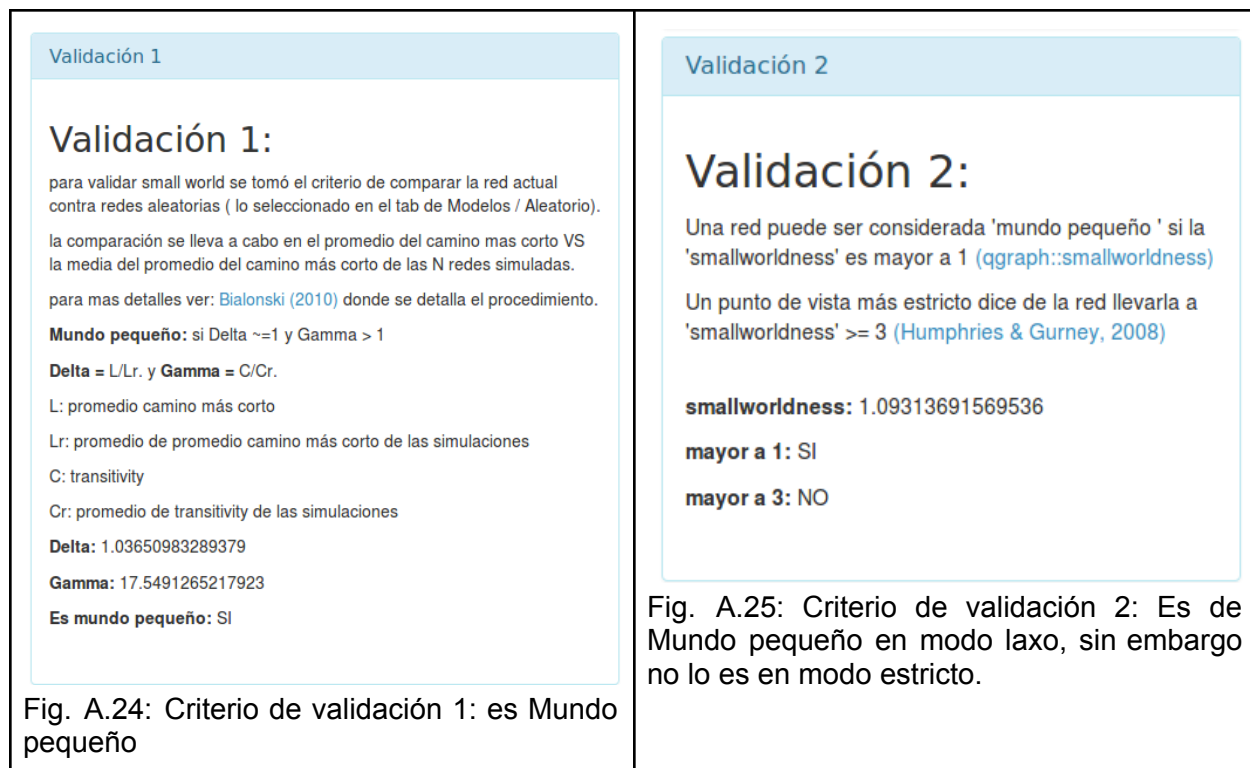


## Mundo pequeño para el periodo 2012 - 2016

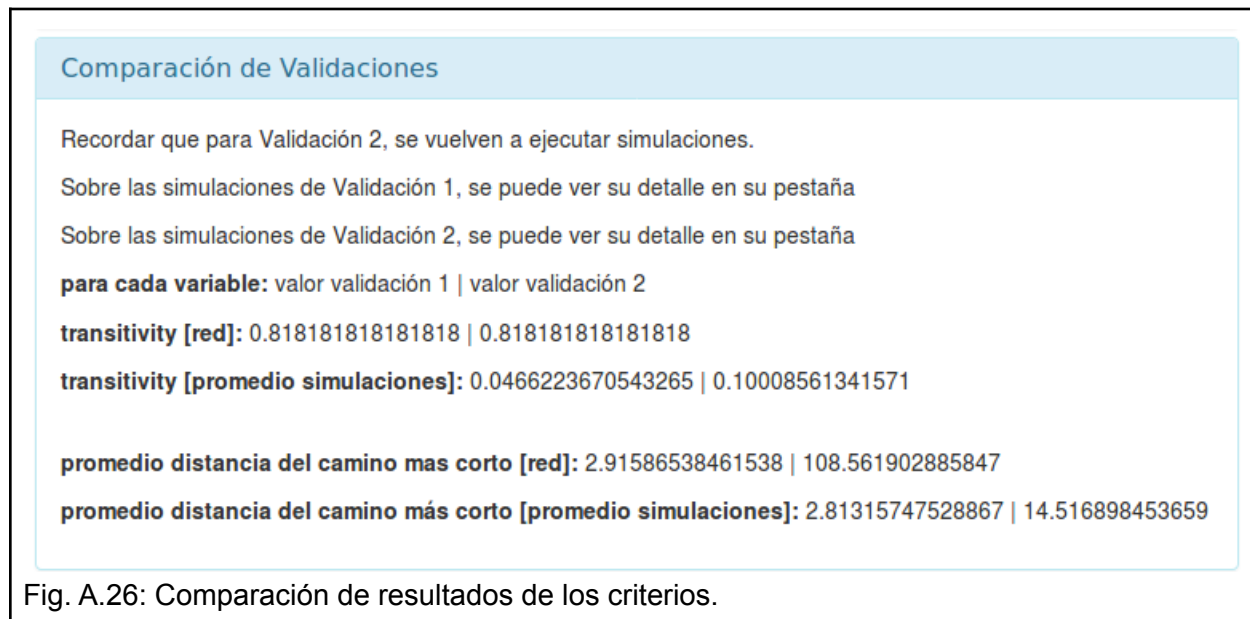
Los parámetros utilizados para esta ejecución se pueden verificar en la [Fig. A.23](#).



Podemos ver que para el periodo 2012-2016, la red es de mundo pequeño para el criterio 1 ([Fig. A.24](#)). Sin embargo, para el criterio 2 sí cumple esta característica para el umbral laxo pero no para el estricto ([Fig. A.25](#)).



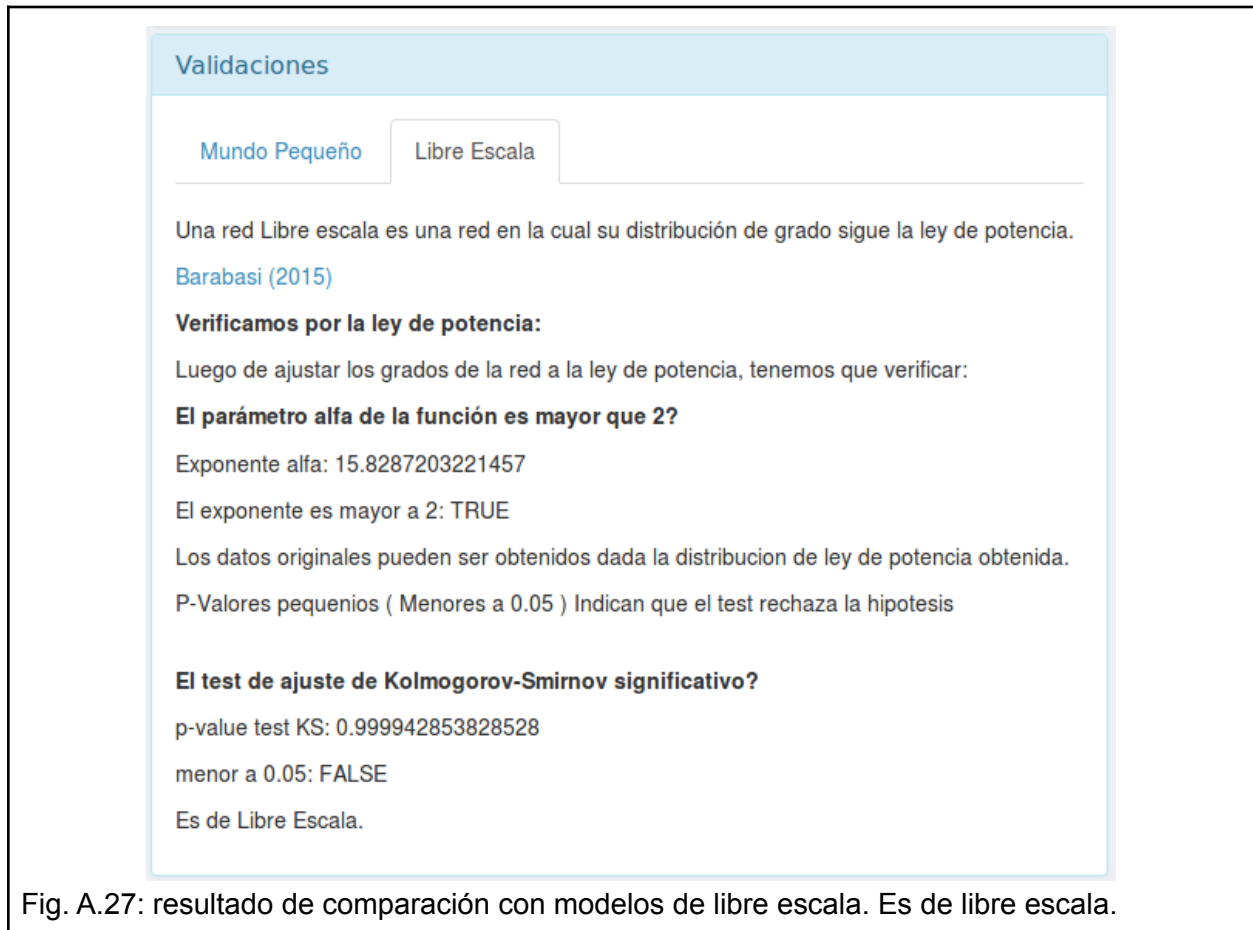
Nuevamente, la diferencia entre ambos enfoques de validación de mundo pequeño se pueden observar en la [Fig. A.26](#).

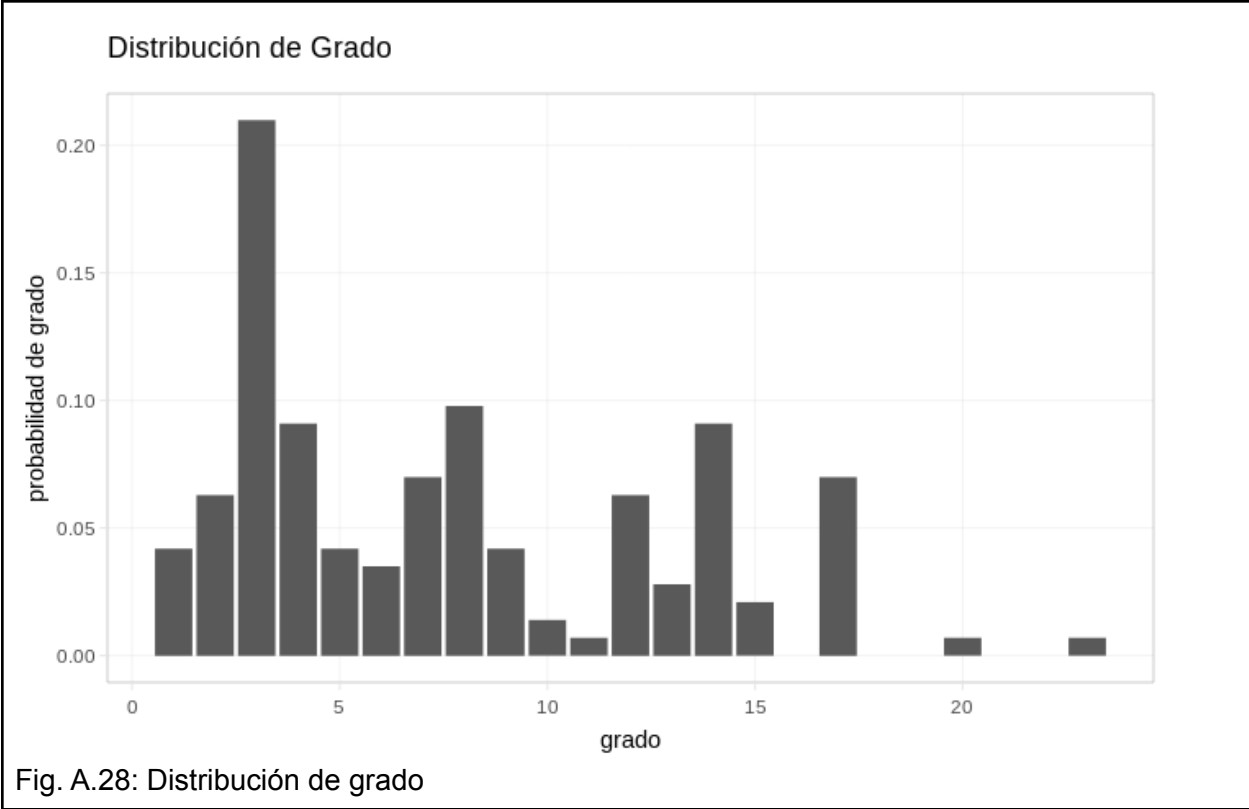


## Libre escala 2012 - 2016

En el periodo de 2012-2016, la red es de libre escala ([Fig. A.27](#)). Sin embargo, en la distribución se observa que en los grados 14 y 17 hay probabilidad alta ([Fig. A.28](#)).

Si vemos la referencia es de 0.05, y el máximo manejado para este periodo es de 0.20.





### Anexo: Cantidad de autores en artículos por año

Tabla A.14: cantidad de artículos con cantidad de autores por año.

año		1996	1999	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
# autores por artículo	1	3	2	3	2	3	2	45	4	5	1	4	5	2	6	3	1	5	4	3	0
	2 a 4	14	14	4	8	6	4	61	7	4	3	7	3	8	14	0	8	9	13	8	3
	5 a 7	3	8	4	1	0	2	22	0	1	1	1	1	0	0	1	2	1	1	6	1
	8 o +	0	0	0	0	0	0	2	0	2	1	1	0	0	1	4	0	2	1	0	2

## Anexo: software y librerías utilizadas

Dado que las librerías son parte esencial de cualquier sistema desarrollado hoy en día, se especifica qué librerías se utilizaron para cada sección de la aplicación y con qué propósito. Si como también sus citas correspondientes.

libreria / software	sección	proposito	citas
R	Generación y aplicación	Lenguaje de programación	R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <a href="https://www.R-project.org/">https://www.R-project.org/</a> . Version: 4.1.2
tidyr	Generación y aplicación	Transformación datos	Hadley Wickham and Lionel Henry (2019). tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions. R package version 0.8.3. <a href="https://CRAN.R-project.org/package=tidyr">https://CRAN.R-project.org/package=tidyr</a>
dplyr	Generación y aplicación	Transformación datos	Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2019). dplyr: A Grammar of Data Manipulation. R package version 0.8.1. <a href="https://CRAN.R-project.org/package=dplyr">https://CRAN.R-project.org/package=dplyr</a>
purrr	Generación y aplicación	Transformación datos	Lionel Henry and Hadley Wickham (2019). purrr: Functional Programming Tools. R package version 0.3.2. <a href="https://CRAN.R-project.org/package=purrr">https://CRAN.R-project.org/package=purrr</a>
stringr	Generación y aplicación	Transformación datos	Hadley Wickham (2019). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. <a href="https://CRAN.R-project.org/package=stringr">https://CRAN.R-project.org/package=stringr</a>
stringi	Generación y aplicación	Transformación datos	Gagolewski M. and others (2019). R package stringi: Character string processing facilities. <a href="http://www.gagolewski.com/software/stringi/">http://www.gagolewski.com/software/stringi/</a> .
lubridate	Generación y aplicación	Transformación datos	Garrett Grolemond, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL <a href="http://www.jstatsoft.org/v40/i03/">http://www.jstatsoft.org/v40/i03/</a> .
RSQLite	Generación y aplicación	Operaciones Base de datos	Kirill Müller, Hadley Wickham, David A. James and Seth Falcon (2018). RSQLite: 'SQLite' Interface for R. R package version 2.1.1. <a href="https://CRAN.R-project.org/package=RSQLite">https://CRAN.R-project.org/package=RSQLite</a>
DBI	Generación y aplicación	Operaciones Base de datos	R Special Interest Group on Databases (R-SIG-DB), Hadley Wickham and Kirill Müller (2018). DBI: R Database Interface. R package version 1.0.0. <a href="https://CRAN.R-project.org/package=DBI">https://CRAN.R-project.org/package=DBI</a>
igraph	Aplicación	Operaciones grafos	Csardi G, Nepusz T: The igraph software package for complex network research, InterJournal, Complex Systems 1695. 2006. <a href="http://igraph.org">http://igraph.org</a>
ggplot2	Aplicación	Graficar	H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
qgraph	Aplicación	Operaciones grafos	Sacha Epskamp, Angelique O. J. Cramer, Lourens J. Waldorp, Verena D. Schmittmann, Denny Borsboom (2012). qgraph: Network Visualizations of Relationships in Psychometric Data. Journal of Statistical Software, 48(4), 1-18. URL <a href="http://www.jstatsoft.org/v48/i04/">http://www.jstatsoft.org/v48/i04/</a> .
RColorBrewer	Aplicación	Selección colores	Erich Neuirth (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. <a href="https://CRAN.R-project.org/package=RColorBrewer">https://CRAN.R-project.org/package=RColorBrewer</a>
visNetwork	Aplicación	Visualización grafos	Almende B.V., Benoit Thieurmél and Titouan Robert (2019). visNetwork: Network Visualization using 'vis.js' Library. R package version 2.0.6.

			<a href="https://CRAN.R-project.org/package=visNetwork">https://CRAN.R-project.org/package=visNetwork</a>
shiny	Aplicación	Funcionamiento aplicación	Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie and Jonathan McPherson (2019). shiny: Web Application Framework for R. R package version 1.3.2. <a href="https://CRAN.R-project.org/package=shiny">https://CRAN.R-project.org/package=shiny</a>
DT	Aplicación	Visualización y funcionalidades tablas interactivas	Yihui Xie, Joe Cheng and Xianying Tan (2019). DT: A Wrapper of the JavaScript Library 'DataTables'. R package version 0.6. <a href="https://CRAN.R-project.org/package=DT">https://CRAN.R-project.org/package=DT</a>
shinyWidgets	Aplicación	Funcionamiento Aplicación	Victor Perrier, Fanny Meyer and David Granjon (2019). shinyWidgets: Custom Inputs Widgets for Shiny. R package version 0.4.8. <a href="https://CRAN.R-project.org/package=shinyWidgets">https://CRAN.R-project.org/package=shinyWidgets</a>
bsplus	Aplicación	Funcionamiento Aplicación	Ian Lyttle (2018). bsplus: Adds Functionality to the R Markdown + Shiny Bootstrap Framework. R package version 0.1.1. <a href="https://CRAN.R-project.org/package=bsplus">https://CRAN.R-project.org/package=bsplus</a>
plotly	Aplicación	Gráficos interactivos	Carson Sievert (2018) plotly for R. <a href="https://plotly-r.com">https://plotly-r.com</a>
shinydashboard	Aplicación	Funcionamiento Aplicación	Winston Chang and Barbara Borges Ribeiro (2018). shinydashboard: Create Dashboards with 'Shiny'. R package version 0.7.1. <a href="https://CRAN.R-project.org/package=shinydashboard">https://CRAN.R-project.org/package=shinydashboard</a>
shinydashboardPlus	Aplicación	Funcionamiento Aplicación	David Granjon (2019). shinydashboardPlus: Add More 'AdminLTE2' Components to 'shinydashboard'. R package version 0.7.0. <a href="https://CRAN.R-project.org/package=shinydashboardPlus">https://CRAN.R-project.org/package=shinydashboardPlus</a>
ndtv	Aplicación	Animaciones grafos	Skye Bender-deMoll (2018). ndtv: Network Dynamic Temporal Visualizations. R package version 0.12.3. <a href="https://CRAN.R-project.org/package=ndtv">https://CRAN.R-project.org/package=ndtv</a>
intergraph	Aplicación	Transformación grafos	Bojanowski, Michal (2015) intergraph: Coercion Routines for Network Data Objects. R package version 2.0-2. <a href="http://mbojan.github.io/intergraph">http://mbojan.github.io/intergraph</a>
assertive	Aplicación	Pruebas de funciones	Richard Cotton (2016). assertive: Readable Check Functions to Ensure Code Integrity. R package version 0.3-5. <a href="https://CRAN.R-project.org/package=assertive">https://CRAN.R-project.org/package=assertive</a>
testthat	Aplicación	Pruebas de funciones	Hadley Wickham. testthat: Get Started with Testing. The R Journal, vol. 3, no. 1, pp. 5--10, 2011
igraphdata	Aplicación	Pruebas de funciones	Gabor Csardi (2015). igraphdata: A Collection of Network Data Sets for the 'igraph' Package. R package version 1.0.1. <a href="https://CRAN.R-project.org/package=igraphdata">https://CRAN.R-project.org/package=igraphdata</a>
forcats	Aplicación	Transformación datos	Hadley Wickham (2019). forcats: Tools for Working with Categorical Variables (Factors). R package version 0.4.0. <a href="https://CRAN.R-project.org/package=forcats">https://CRAN.R-project.org/package=forcats</a>
visdat	Aplicación	Gráficos	Tierney N (2017). "visdat: Visualising Whole Data Frames." _JOSS_, *2*(16), 355. doi:10.21105/joss.00355 (URL: <a href="https://doi.org/10.21105/joss.00355">https://doi.org/10.21105/joss.00355</a> ), <URL: <a href="http://dx.doi.org/10.21105/joss.00355">http://dx.doi.org/10.21105/joss.00355</a> >.
ggridges	Aplicación	Gráficos	Claus O. Wilke (2018). ggridges: Ridgeline Plots in 'ggplot2'. R package version 0.5.1. <a href="https://CRAN.R-project.org/package=ggridges">https://CRAN.R-project.org/package=ggridges</a>
skimr	Aplicación	Cálculo, visualización estadísticas	Amelia McNamara, Eduardo Arino de la Rubia, Hao Zhu, Shannon Ellis and Michael Quinn (2019). skimr: Compact and Flexible Summaries of Data. R package version 1.0.5. <a href="https://CRAN.R-project.org/package=skimr">https://CRAN.R-project.org/package=skimr</a>

ggrepel	Aplicación	graficos	Kamil Slowikowski (2019). ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.8.1. <a href="https://CRAN.R-project.org/package=ggrepel">https://CRAN.R-project.org/package=ggrepel</a>
Rcpp	compilacion	Compilación código	Dirk Eddebuettel and Romain Francois (2011). Rcpp: Seamless R and C++ Integration. Journal of Statistical Software, 40(8), 1-18. URL <a href="http://www.jstatsoft.org/v40/i08/">http://www.jstatsoft.org/v40/i08/</a> .  Dirk Eddebuettel, Dirk (2013) Seamless R and C++ Integration with Rcpp. Springer, New York. ISBN 978-1-4614-6867-7.  Dirk Eddebuettel and James Joseph Balamuta (2017). Extending R with C++: A Brief Introduction to Rcpp. PeerJ Preprints 5:e3188v1. URL <a href="https://doi.org/10.7287/peerj.preprints.3188v1">https://doi.org/10.7287/peerj.preprints.3188v1</a> .
devtools	compilacion	Obtención código	Hadley Wickham, Jim Hester and Winston Chang (2019). devtools: Tools to Make Developing R Packages Easier. R package version 2.0.2. <a href="https://CRAN.R-project.org/package=devtools">https://CRAN.R-project.org/package=devtools</a>
htmlwidgets	Aplicación	Funcionamiento aplicación	Ramnath Vaidyanathan, Yihui Xie, JJ Allaire, Joe Cheng and Kenton Russell (2018). htmlwidgets: HTML Widgets for R. R package version 1.3. <a href="https://CRAN.R-project.org/package=htmlwidgets">https://CRAN.R-project.org/package=htmlwidgets</a>
jsonlite	Aplicación	Funcionamiento aplicación	Jeroen Ooms (2014). The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. arXiv:1403.2805 [stat.CO] URL <a href="https://arxiv.org/abs/1403.2805">https://arxiv.org/abs/1403.2805</a> .
RBioFabric	Aplicación	Visualización	William Longabaugh (2016). RBioFabric: BioFabric network visualization tool. R package version 0.4. <a href="http://www.BioFabric.org/">http://www.BioFabric.org/</a>
OpenRefine	Generación	Limpieza datos autores.	OpenRefine. A free, open source, power tool for working with messy data. <a href="http://openrefine.org/">http://openrefine.org/</a> version: OpenRefine 2.8
here	Generación y Aplicación	Acceso a archivos	Kirill Müller (2017). here: A Simpler Way to Find Your Files. R package version 0.1. <a href="https://CRAN.R-project.org/package=here">https://CRAN.R-project.org/package=here</a>
Rstudio	Generación y Aplicación	Programación	RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <a href="http://www.rstudio.com/">http://www.rstudio.com/</a> .
Git	Generación y Aplicación	Versionado	Git is a free and open source distributed version control system designed to handle everything from small to very large projects with speed and efficiency. <a href="https://git-scm.com/">https://git-scm.com/</a>
RefManageR	Generación	Extracción citas	McLean MW (2017). "RefManageR: Import and Manage BibTeX and BibLaTeX References in R." _The Journal of Open Source Software_. doi: 10.21105/joss.00338 (URL: <a href="https://doi.org/10.21105/joss.00338">https://doi.org/10.21105/joss.00338</a> ).  McLean MW (2014). _Straightforward Bibliography Management in R Using the RefManager Package_. arXiv: 1403.2036 [cs.DL], <URL: <a href="https://arxiv.org/abs/1403.2036">https://arxiv.org/abs/1403.2036</a> >.
rvest	Generación	Extracción HTML	Hadley Wickham (2019). rvest: Easily Harvest (Scrape) Web Pages. R package version 0.3.4. <a href="https://CRAN.R-project.org/package=rvest">https://CRAN.R-project.org/package=rvest</a>
pdftools	Generación	Extracción PDF	Jeroen Ooms (2019). pdftools: Text Extraction, Rendering and Converting of PDF Documents. R package version 2.2. <a href="https://CRAN.R-project.org/package=pdftools">https://CRAN.R-project.org/package=pdftools</a>
tesseract	Generación	Extracción OCR	Jeroen Ooms (2018). tesseract: Open Source OCR Engine. R package version 4.0. <a href="https://CRAN.R-project.org/package=tesseract">https://CRAN.R-project.org/package=tesseract</a>
network	Aplicación	grafos	Butts C (2015). _network: Classes for Relational Data_. The Statnet Project (<URL: <a href="http://www.statnet.org">http://www.statnet.org</a> >). R package version 1.13.0.1,



			<URL: <a href="https://CRAN.R-project.org/package=network">https://CRAN.R-project.org/package=network</a> >. Butts C (2008). "network: a Package for Managing Relational Data in R." _Journal of Statistical Software_, *24*(2). <URL: <a href="http://www.jstatsoft.org/v24/i02/paper">http://www.jstatsoft.org/v24/i02/paper</a> >.
networkDynamic	Aplicación	grafos	Carter T. Butts, Ayn Leslie-Cook, Pavel N. Krivitsky and Skye Bender-deMoll (2019). networkDynamic: Dynamic Extensions for Network Objects. R package version 0.10.0. <a href="https://CRAN.R-project.org/package=networkDynamic">https://CRAN.R-project.org/package=networkDynamic</a>
tidytext	Aplicación	text mining	Silge J, Robinson D (2016). "tidytext: Text Mining and Analysis Using Tidy Data Principles in R." JOSS, 1(3). doi: 10.21105/joss.00037, <a href="http://dx.doi.org/10.21105/joss.00037">http://dx.doi.org/10.21105/joss.00037</a> .
modifiedmk	resultados	Test estadístico	Sandeep Kumar Patakamuri and Nicole O'Brien (2021). modifiedmk: Modified Versions of Mann Kendall and Spearman's Rho Trend Tests. R package version 1.6. <a href="https://CRAN.R-project.org/package=modifiedmk">https://CRAN.R-project.org/package=modifiedmk</a>

## Anexo: Detalle de algoritmos para búsqueda de comunidades

Cada uno de los algoritmos disponibles representa diferentes formas de considerar la agrupación y su interpretación depende de que representan los nodos y las aristas. Los algoritmos utilizados están disponibles a través de la [librería igraph](#). Un breve resumen se puede ver en [Tabla A.15](#).

Tabla A.15: Resumen algoritmos de búsqueda de comunidades.

Función en Herramienta	Tipo Comunidad	Referencia de implementación	Descripción algoritmo
Cluster_edge_betweenness	Edge Betweenness	<a href="http://igraph.org/r/doc/cluster_edge_betweenness.html">http://igraph.org/r/doc/cluster_edge_betweenness.html</a>	Identifica la intermediación de las aristas, y luego las va quitando a las que mayor valor tienen, para ver cómo se van formando nuevos subgrafos. Para detalles ver <a href="#">Newman-Girvan (2004)</a>
Cluster_label_prop	Propagating Labels	<a href="http://igraph.org/r/doc/cluster_label_prop.html">http://igraph.org/r/doc/cluster_label_prop.html</a>	Cada nodo arranca con una etiqueta única, en el siguiente paso cada nodo adopta la etiqueta que la mayoría de sus vecinos tiene. Para más detalles ver <a href="#">Raghavan,Albert. et al. (2007)</a>
Cluster_leading_eigen	Leading Eigenvector	<a href="http://igraph.org/r/doc/cluster_leading_eigen.html">http://igraph.org/r/doc/cluster_leading_eigen.html</a>	Utiliza la matriz de modularidad, y luego utiliza los autovectores y autovalores de la misma. Para más detalles ver <a href="#">Newman (2006)</a>
Cluster_louvain	Multi-level Optimization of Modularity	<a href="http://igraph.org/r/doc/cluster_louvain.html">http://igraph.org/r/doc/cluster_louvain.html</a>	Mezcla una forma jerárquica con la modularidad. Optimiza la modularidad primero localmente, cuando no quedan más vértices. Vuelve a arrancar tomando cada comunidad como un vértice. Termina cuando solo queda un vértice. Para más detalles ver <a href="#">Blondel Guillaume. et al.(2008)</a>
Cluster_walktrap	Short Random Walks	<a href="http://igraph.org/r/doc/cluster_walktrap.html">http://igraph.org/r/doc/cluster_walktrap.html</a>	La idea es que caminatas aleatorias cortas se quedan en la misma comunidad. Para más detalles ver <a href="#">Pons-Latapy (2005)</a>
Cluster_infomap	Infomap community finding	<a href="http://igraph.org/r/doc/cluster_infomap.html">http://igraph.org/r/doc/cluster_infomap.html</a>	Encuentra estructuras de comunidad que minimizan el largo de la descripción de la trayectoria de un caminador aleatorio. <a href="#">Rosvall, Axelsson. et al.(2009)</a>

## Anexo: Detalle procesamiento bibTEX

Los archivos bibtex fueron generados de la siguiente manera:

- 1) Se accedió al link de javascript generado por revista
- 2) Se guardó la página HTML que tenía un elemento textarea con el texto bibtex correspondiente
- 3) Se extrajo el texto bibtex del elemento textarea
- 4) Se guardó en un archivo bibtex
- 5) Se trató de parsear el archivo bibtex con la [librería RefManageR](#)
- 6) Se encontraron problemas generalmente relacionados a mala escritura del formato bibtex
  - a) Tema de escribir autor1 Y autor2, mientras que en bibtex es autor1 **AND** autor2.
- 7) Se encontraron problemas relacionados con el encoding de los archivos.
  - a) ASCII vs UTF8 (acentos, ñ, etc )
- 8) Se resolvieron ambos problemas
  - a) El primero identificando los casos de forma manual
  - b) El 2do utilizando sistema operativo Linux para realizar la transformación y guardar en UTF-8
- 9) Se encontraron inconsistencias en los datos presentados en el formato bibtex y lo que se encontraba presentado en los HTML de la revista.
  - a) Falta de dato volumen
  - b) Falta de abstract
- 10) Se terminó extrayendo la información que se consideraba complementaria y se almacenó en la base de datos de la revista.

### Problemas de transformación y encoding bibtex

Detalles técnicos de los sistemas operativos utilizados:

Linux: Linux Mint 17.3 Rosa ( basado en Ubuntu 14.04 )

Windows: Windows 10 y Windows 7

Luego de resolver estos temas en Linux ( con caracteres no ASCII), se pudo generar el archivo de texto y luego levantarlo sin problemas.

Sin embargo cuando utilizo una computadora con Sistema operativo Windows no se pudo levantar el archivo generado en Linux por problemas de BOM ya que en Windows por default toma los archivos con cierta configuración de R. Mientras el default de configuración Linux es UTF-8, por lo tanto hace más compatible los acentos y caracteres no ASCII

De esta forma se optó por hacer las tareas de Interpretación de archivos en Linux, y luego de tener generados los archivos bases, se pudo seguir trabajando en ambos sistemas operativos.

Tabla A.16 Problemas con archivos bibtex

Link bibtex	problema
<a href="https://revistas.unlp.edu.ar/raab/article/cite/229/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/229/BibtexCitationPlugin</a>	Apellidos con apóstrofe.Solución: reemplazar el apóstrofe por comilla simple. caso: Gabriela L'Heureux

<a href="https://revistas.unlp.edu.ar/raab/article/cite/1513/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1513/BibtexCitationPlugin</a>	Apellidos con apóstrofe.Solución: reemplazar el apóstrofe por comilla simple. caso: Manuel D. D'Angelo del Campo
<a href="https://revistas.unlp.edu.ar/raab/article/cite/612/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/612/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/425/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/425/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/2541/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/2541/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/554/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/554/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/2505/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/2505/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/458/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/458/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/1172/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1172/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/258/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/258/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/2506/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/2506/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/259/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/259/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/574/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/574/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/2514/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/2514/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/260/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/260/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/561/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/561/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/2516/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/2516/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/261/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/261/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/557/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/557/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/2515/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/2515/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/265/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/265/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/579/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/579/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/2511/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/2511/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/268/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/268/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/2507/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/2507/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/269/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/269/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/2177/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/2177/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/270/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/270/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/2513/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/2513/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/271/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/271/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/272/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/272/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/273/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/273/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/274/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/274/BibtexCitationPlugin</a>	Falta campo volumen
<a href="https://revistas.unlp.edu.ar/raab/article/cite/275/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/275/BibtexCitationPlugin</a>	Falta campo volumen



<a href="https://revistas.unlp.edu.ar/raab/article/cite/1745/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1745/BibtexCitationPlugin</a>	Abstract Vacio
<a href="https://revistas.unlp.edu.ar/raab/article/cite/1746/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1746/BibtexCitationPlugin</a>	Abstract Vacio
<a href="https://revistas.unlp.edu.ar/raab/article/cite/1747/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1747/BibtexCitationPlugin</a>	Abstract Vacio
<a href="https://revistas.unlp.edu.ar/raab/article/cite/1748/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1748/BibtexCitationPlugin</a>	Abstract Vacio
<a href="https://revistas.unlp.edu.ar/raab/article/cite/1813/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1813/BibtexCitationPlugin</a>	Abstract Vacio
<a href="https://revistas.unlp.edu.ar/raab/article/cite/1818/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1818/BibtexCitationPlugin</a>	Abstract Vacio
<a href="https://revistas.unlp.edu.ar/raab/article/cite/1826/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1826/BibtexCitationPlugin</a>	Abstract Vacio
<a href="https://revistas.unlp.edu.ar/raab/article/cite/1829/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1829/BibtexCitationPlugin</a>	Abstract Vacio
<a href="https://revistas.unlp.edu.ar/raab/article/cite/1831/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1831/BibtexCitationPlugin</a>	Abstract Vacio
<a href="https://revistas.unlp.edu.ar/raab/article/cite/1837/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1837/BibtexCitationPlugin</a>	Abstract Vacio
<a href="https://revistas.unlp.edu.ar/raab/article/cite/1839/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1839/BibtexCitationPlugin</a>	Abstract Vacio
<a href="https://revistas.unlp.edu.ar/raab/article/cite/1840/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1840/BibtexCitationPlugin</a>	Abstract Vacio
<a href="https://revistas.unlp.edu.ar/raab/article/cite/1842/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1842/BibtexCitationPlugin</a>	Abstract Vacio
<a href="https://revistas.unlp.edu.ar/raab/article/cite/1852/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1852/BibtexCitationPlugin</a>	Abstract Vacio
<a href="https://revistas.unlp.edu.ar/raab/article/cite/1861/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1861/BibtexCitationPlugin</a>	Abstract Vacio
<a href="https://revistas.unlp.edu.ar/raab/article/cite/1804/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1804/BibtexCitationPlugin</a>	Abstract Vacio
<a href="https://revistas.unlp.edu.ar/raab/article/cite/1869/BibtexCitationPlugin">https://revistas.unlp.edu.ar/raab/article/cite/1869/BibtexCitationPlugin</a>	Abstract Vacio

## Anexo: Detalle consideraciones PDF

Los archivos PDF al final no fueron considerados. Se eliminó la sección de text mining para tópicos, quedando para potenciales trabajos futuros. Sin embargo existieron las siguientes lecciones aprendidas:

### Obtención de los pdf

Se pudieron bajar sin problema, se agregó una demora entre archivo y archivo, para no saturar el servidor de la revista.

### Interpretación de los pdf

Conversión de los pdf a texto; Existían diferentes formatos de los pdf;

Se probó:

- Bajar un artículo de la edición más antigua: es un escaneo
- Bajar un artículo de una edición intermedia: es un escaneo
- Bajar un artículo de una edición moderna: es un texto

Potenciales soluciones:

- OCR (*optical character recognition*) para los escaneos
- Extracción de texto para los pdf más modernos.

Extracción de texto ( Se utilizó el lenguaje R , con la [librería pdftools](#) )

- Errores con formato de más de 1 columna.
- Errores de extracción de texto por márgenes de página.
- Se decidió convertir cada hoja a una imagen y luego pasarlas por OCR.

Aplicación de OCR ( Se utilizó el lenguaje R , con las [librerías pdftools y tesseract](#) )

- Extracción de las imágenes de cada página a través de la [librería pdftools](#)
- Procesar cada imagen por artículo ejecutando OCR a través de la [librería tesseract](#), se aplicó el uso del motor de idioma en español.
- Compaginar todos los textos en un archivo txt final que contenga todos los resultados OCR de cada una de las páginas.

Con esto se logró superar considerablemente el problema de márgenes, sin embargo seguían existiendo problemas con los textos, y extracción de los datos de los mismos.

Se decidió no incluir la parte de text mining en la tesis, dado que sino se extendería considerablemente. Además dado que hoy en día (2017-2020) hay una revolución con la interpretación de imágenes, que tal vez en breve tengamos mejores resultados pudiendo realizar de forma más fácil en trabajos futuros.

## Anexo: Detalle estandarización autores

Se juntaron los conjuntos de datos de los orígenes de HTML , y los archivos de citas BIBtex. Se exportaron los datos resultantes a un archivo en formato texto separado por tabulaciones.

Este archivo se importó en la herramienta [OpenRefine](#) la cual resulta muy práctica a la hora de buscar similitudes entre registros de cada columna. La versión utilizada para realizar este trabajo fue OpenRefine 2.8.

Una vez cargado el archivo , se procedió a repetir la columna autor , y a la nueva ponerle de nombre “autor\_original”. Esta columna se agrega para cumplir 2 funciones, una de copia de respaldo, en caso de que llegase a haber inconvenientes. Y la segunda, que es para mapear el valor original de la columna con el valor obtenido luego de procesar la columna “autor” repetidas veces. ( [Fig. A.29](#) )



Fig. A.29: openrefine resaltadas columna autor, la columna autor original, y un ejemplo de un registro donde el autor es la misma persona, sin embargo figura con diferentes nombres en los artículos.

Sobre la columna “autor” las tareas realizadas fueron para buscar similitudes a través de métodos de agrupamiento ( *clustering* ). A continuación se pueden ver los métodos que se ejecutaron, y en qué orden. Más allá de este resultado se exportó un archivo con las transformaciones realizadas; sin embargo este archivo no conserva los métodos de clustering usados, ni los parámetros utilizados; por lo cual se pasó a detallar en formato textual dentro de este anexo. ( ver [Iteraciones de clustering](#) )

## Revisión y validación con expertos

Luego de depurar empíricamente con diferentes formas de agrupar, validamos los nombres y casos con los expertos del área. Surge una lista de autores a unificar, que no aparecieron en el análisis de clustering o que había que modificar. ( ver [Tabla A.17](#) )



Se cambió el formato de excel enviado por los expertos a un formato más amigable para la automatización. Se armó una utilidad intermedia que dado un excel, transforma a el formato de importación del openrefine, que es un archivo con un formato específico. Gracias a la herramienta se pudieron distinguir errores que se pasaron, en la parte de la unificación de autores.

## Iteraciones de clustering

### Clustering - Iteración 1

#### **Método y parámetros**

Clustering > columna autor > método: keycollision > keying function: fingerprint

#### **Criterio de selección**

Empírico. Es la primera opción al seleccionar la opción de clustering en openrefine en la versión utilizada (version: 2.7).

#### **Observación**

Dado que son varios los cambios se procede a detallar reglas generales aplicadas, para ver detalles de cuáles fueron exactamente ver el archivo exportado de reglas del openrefine.

#### **Resultado**

Al realizar el primer clustering en nombres de autores se puede ver que existen varios Temas con escritura del nombre:

- Nombres que consideran el nombre intermedio (María Antonia Luis)
- Nombres que al intermedio lo ponen con una abreviación (María A. Luis)
- Abreviación del nombre intermedio con punto (María A. Luis)
- Abreviación del nombre intermedio sin punto (María A Luis)
- Nombres con acentos o caracteres especiales (María A. Luis)
- Nombres en los que se obvian los acentos o caracteres especiales. (Maria A. Luis)
- Nombres que cambian en el tiempo > ej: Cesani > Cesani Rossi

### Clustering - Iteración 2

#### **Método y parámetros**

Clustering > columna autor > método: keycollision > keying function: fingerprint

#### **Criterio de selección**

Empírico. Volver a tratar de ejecutar el mismo método para ver que no existan más duplicados con los parámetros utilizados.

#### **Resultado**

Se trata de dar un segundo clustering bajo la misma metodología y no se encuentran resultados

### Clustering - Iteración 3

#### **Método y parámetros**

Se procede a probar con otro tipo de configuración de clustering  
método: keycollision > keying function: ngram-fingerprint > ngram-size: 2

### ***Criterio de selección***

Empírico. Dentro de las opciones de parámetros dentro del método de key collision es la 2da opción. El tamaño del engrama fue seleccionado porque es el que viene por defecto.

### ***Observación***

Dados que son varios los cambios se procede a detallar reglas generales aplicadas, para ver detalles de cuáles fueron exactamente ver el archivo exportado de reglas del openrefine.

### ***Resultado***

Se encuentran temas con:

- Más de una abreviación en nombres intermedios
- Apellidos compuestos con guiones
- Múltiples nombres y apellidos

## **Clustering - Iteración 4**

### ***Método y parámetros***

Se procede a ejecutar nuevamente bajo la misma metodología.

método: keycollision > keying function: ngram-fingerprint > ngram-size: 2

### ***Criterio de selección***

Empírico. Validación de lo aplicado en Iteración 3. Se espera que no haya resultados.

### ***Observación***

Dados que son varios los cambios se procede a detallar reglas generales aplicadas, para ver detalles de cuáles fueron exactamente ver el archivo exportado de reglas del openrefine.

### ***Resultado***

No se encuentran resultados.

Clustering - Iteraciones 5-9

### ***Método y parámetros***

Se procede a ejecutar nuevamente bajo la misma metodología , variando el tamaño del engrama con los valores 3,4,5,6,7

método: keycollision > keying function: ngram-fingerprint > ngram-size: 2

### ***Criterio de selección***

Empírico. Probando expandir el tamaño del engrama base para el clustering.

### ***Resultado***

No se encuentran resultados. Para los parámetros 3,4,5,6.

En el parámetro con valor 7 si, pero no se utilizaron las sugerencias planteadas, dado que afecta los nombres y apellidos cortos.

## **Clustering - Iteración 10**

### ***Método y parámetros***

Se procede a pasar a otra estrategia. Método: keycollision , keying function: metaphone3

### ***Criterio de selección***

Empírico. Probando otra función.

### **Resultado**

Dando nuevos resultados esta vez relacionados con:

- Casos de nombres compuestos o apellidos compuestos ( Fernando Ramírez Rozzi )
  - Para la elección del adecuado, generalmente se termina buscando en google o cuit online u otras páginas para verificar existencia de la persona y asociada al área; para ver perfiles asociados a la persona, queda constatar con los expertos
- Hay casos dudosos como por ejemplo: Rolando González-José , vs Rolando González
  - Al ser apellidos tan comunes es difícil de realizar elecciones; a constatar con expertos.

Respecto de consideraciones con apellidos del estilo

- Cesani > Cesani Rossi ; que pueden asociar a cambio de estados civiles. Por lo tanto sigue siendo la misma persona, hay que determinar si estos cambios interesan , o si se desea cambiar el análisis de red.
- Hay que tener cuidado al aplicar las modificaciones del campo dado que por ejemplo: Francisco Di Fabbio Rocca, no es lo mismo que Francisco de Ávila ; hay que ser criteriosos con las elecciones.

### **Clustering - Iteración 11**

#### **Método y parámetros**

Se procede a validar la estrategia anterior. Método: keycollision , keying function: metaphone3

#### **Criterio de selección**

Empírico. Verificando método utilizado en iteración 10, para búsqueda de más resultados.

### **Resultado**

Se volvió a dar una 2da corrida del algoritmo luego de las correcciones, el único caso es el de Francisco Di Fabbio Rocca y Francisco de Ávila ; que lo dejamos sin corrección. Dado que son personas diferentes.

### **Clustering - Iteración 12**

#### **Método y parámetros**

Se procede a pasar a otra estrategia. Method: keycollision , keying function: cologne-phonetic

#### **Criterio de selección**

Empírico. Siguiendo opción dentro de keycollision.

### **Resultado**

- Surge el caso de : Cristina Dejean vs Cristian Dejean > se deja separado dado que tranquilamente pueden ser 2 personas diferentes por sexo , sin embargo puede ser un error de entrada de datos. ( se considera poco probable que se pase ese tipo de error hasta nivel paper ) y cristian tiene 2 papers vs cristina 13 , podría ser algún familiar. A cotejar con expertos
- Se siguen encontrando casos de apellidos compuestos
- Se siguen encontrando casos de nombres con iniciales

### **Clustering - Iteración 13**

**Método y parámetros**

Se procede a validar la estrategia anterior.

Método: keycollision , keying function: cologne-phonetic

**Criterio de selección**

Empírico. Verificando método utilizado en iteración 13, para búsqueda de más resultados.

**Resultado**

- Se volvió a ejecutar esa estrategia, y solo aparecieron los casos de dejean y otro que no cuadran los apellidos, solo los nombres.

**Clustering - Iteración 14****Método y parámetros**

Se procede con otra estrategia:

Method: nearest neighbour , distance function: levenshtein , Radius: 1.0, Block Chars: 6

**Criterio de selección**

Empírico. Siguiendo combinación dentro de la herramienta.

**Resultado**

- En los casos de :
- Emma E. Alfaro(20 rows) / Emma L. Alfaro(14 rows)
- María C. Barboza(11 rows) / María J. Barboza(7 rows)
- No se si es la misma personas, hay que ver con los expertos.
- No se hace ese cambio

**Clustering - Iteración 15****Método y parámetros**

Se procede con otra estrategia: aumentamos el radio a 2

Method: nearest neighbour , distance function: levenshtein , Radius: 2.0, Block Chars: 6

**Criterio de selección**

Empírico. Variando los parámetros del algoritmo probado en iteración anterior.

**Resultado**

- Generico, Aparecieron casos que eran de segundos nombres.

**Clustering - Iteraciones 16,17,18,19,20,21,22****Método y parámetros**

Se procede con otras estrategias de variar el block chars: entre 3 y 10

Method: nearest neighbour, distance function: levenshtein , Radius: 2.0, Block Chars: 3, 10.

**Criterio de selección**

Empírico. Variando los parámetros del algoritmo probado en iteración anterior.

**Resultado**

- Sin resultados favorables.

## Clustering - Iteraciones 23

### **Método y parámetros**

Se procede con otra estrategia de aumentar el radio nuevamente por 1 punto.

Method: nearest neighbour , distance function: levenshtein , Radius: 3.0, Block Chars: 6

### **Criterio de selección**

Empírico. Variando los parámetros del algoritmo probado en iteración anterior.

### **Resultado**

- Resultados genericos, Muchos resultados favorables.
- Se siguen encontrando casos de apellidos compuestos
- Se siguen encontrando casos de nombres con iniciales
- Hay que revisar muy cuidadosamente los resultados dado que hay mucha similitud entre algunos y son diferentes. Por ejemplo:
  - Jorge A. Suby vs Jorge A. Rey
  - María C. Barboza vs María J. Barboza
  - Queda validar con los expertos.

## Clustering - Iteración 24

### **Método y parámetros**

Se procede a repetir la estrategia

Method: nearest neighbour , distance function: levenshtein , Radius: 3.0, Block Chars: 6

### **Criterio de selección**

Empírico. Repetimos los parámetros anteriores para ver cómo afecta.

### **Resultado**

- Se encontraron algunos casos como por ejemplo: Mariano C. Del Papa vs Mariano Del Papa
- Siguen existiendo casos con los que tener mucho cuidado dada la poca variabilidad pero evidente diferencia
  - Vanina fernandez, Vanesa fernandez
  - Queda validar con los expertos.

## Clustering - Iteración 25

### **Método y parámetros**

Se procede con otra estrategia: para reducir los casos problemáticos y ver si se encuentran otros mejores, bajamos los Block Chars a 3.

Method: nearest neighbour , distance function: levenshtein , Radius: 3.0, Block Chars: 3

### **Criterio de selección**

Empírico. Variando los parámetros del algoritmo probado en iteración anterior.

### **Resultado**

- Resultados genericos.
- Se encontraron casos como el de Ana M. Inda vs Ana Inda
- Siguen existiendo casos con los que tener mucho cuidado dada la poca variabilidad pero evidente diferencia.

- Queda validar con los expertos.

## Clustering - Iteración 27

### **Método y parámetros**

Se procede con otra estrategia: para reducir los casos problemáticos y ver si se encuentran otros mejores, bajamos los Block Chars a 2.

Method: nearest neighbour , distance function: levenshtein , Radius: 3.0, Block Chars: 2

### **Criterio de selección**

Empírico. Variando los parámetros del algoritmo probado en iteración anterior.

### **Resultado**

- Siguen existiendo casos con los que tener mucho cuidado dada la poca variabilidad pero evidente diferencia.
- Queda validar con los expertos.

## Clustering - Iteración 28

### **Método y parámetros**

Se procede con otra estrategia, cambiar la función de distancia.

Method: nearest neighbour , distance function: PPM , Radius: 3.0, Block Chars: 3

### **Criterio de selección**

Empírico. Variando la función de distancia.

### **Resultado**

- Se encuentran muchos resultados sin embargo sin valor dado que no toma en cuenta los apellidos, agrupa casos de nombres muy comunes como por ejemplo: los nombres que comienzan con 'María X yyyyy'
- Sirve mucho para identificar los segundos nombres de varias personas, como por ejemplo > cristina b dejean > cristina beatriz dejean

## Clustering - Iteración 29

### **Método y parámetros**

Se procede nuevamente con la estrategia

Method: nearest neighbour , distance function: PPM , Radius: 3.0, Block Chars: 3

### **Criterio de selección**

Empírico. Validando la estrategia de iteración anterior.

### **Resultado**

- Mismas observaciones anteriores, un caso a revisar específico:
  - Héctor Hugo Varela(14 rows) / Hugo H. Varela(8 rows)

## Resultado de Revisión y validación con expertos

La [Tabla A.17](#) tiene el mapeo de los autores revisados por expertos y su valor original.

Tabla A.17: mapeo de nombres finales - nombres originales.

autor_por_expertos	autor_original
Alberto S. Slepoy	Alberto S. Slepoy

Alberto S. Slepoy	Alberto Slepoy
Alejandro C. Tobisch	Alejandro C. Tobisch
Alejandro C. Tobisch	Alejandro Tobisch
Alfredo I. Prieto	Alfredo I. Prieto
Alfredo I. Prieto	Alfredo Prieto
Christine Keyser	Christine Keyser-Tracqui
Cristina Beatriz Dejean	Cristian Dejean
Cristian Marcelo Crespo	Cristian Crespo
Cristian Marcelo Crespo	Cristian Marcelo Crespo
Emma L. Alfaro	Emma Alfaro
Emma L. Alfaro	Emma E. Alfaro
Emma L. Alfaro	Emma L. Alfaro
Emma L. Alfaro	Emma L. Alfaro Gómez
Fernando V. Ramírez-Rozzi	Fernando Ramírez Rossi
Fernando V. Ramírez-Rozzi	Fernando V. Ramírez Rossi
Fernando V. Ramírez-Rozzi	Fernando Ramirez Rozzi
Fernando V. Ramírez-Rozzi	Fernando Ramírez-Rozzi
Fernando V. Ramírez-Rozzi	Fernando Ramírez Rozzi
Francisco A. Zangrando	Francisco A. Zangrando
Francisco A. Zangrando	A. Francisco Sangrando
Francisco Raúl Carnese	Francisco R. Carnesse
Francisco Raúl Carnese	Francisco Carnese
Francisco Raúl Carnese	Francisco Carnesse
Francisco Raúl Carnese	Francisco R Carnesse
Francisco Raúl Carnese	Francisco R. Carnese
Francisco Raúl Carnese	Francisco Raúl Carnese
Jimena Barbeito-Andrés	Gimena Barbeito
Jimena Barbeito-Andrés	Jimena Barbeito-Andrés
Héctor Mario Pucciarelli	Héctor Mario Pucciarelli
Héctor Mario Pucciarelli	Hector M. Pucciarelli
Héctor Mario Pucciarelli	Héctor M. Pucciarelli
Héctor Mario Pucciarelli	Hector Pucciarelli
Héctor Mario Pucciarelli	Héctor Pucciarelli
Héctor Hugo Varela	Hugo H. Varela
Héctor Hugo Varela	Héctor H. Varela
José A. Castro Ocón	José A. Castro
José A. Castro Ocón	José A. Castro Ocón
Leandro Hernán Luna	Leandro Hernán Luna
Leandro Hernán Luna	Leandro H. Luna
Leandro Hernán Luna	Leandro Luna

Gabriela Lorena L'Heureux	Gabriela L. L'Heureux
Gabriela Lorena L'Heureux	Lorena L'Heureux
Luis Edurado Castro	Luis E. Castro
Luis Edurado Castro	Luis Castro
Luis Edurado Castro	Luis Edurado Castro
María Antonia Luis	María A. Luis
María Antonia Luis	María Luis
María Antonia Luis	María Antonia Luis
María Laura Bergel Sanchís	María Bergel Sanchís
María Laura Bergel Sanchís	María Laura Bergel Sanchís
María Laura Bergel Sanchís	María Laura Bergel
María Laura Bergel Sanchís	Maria L. Bergel
María Laura Bergel Sanchís	Maria Bergel
María Cristina Muñe	Maria C. Muñe
María Cristina Muñe	María C. Muñe
María Cristina Muñe	María Cristina Muñe
María Cristina Muñe	Maria Muñe
María Cristina Muñe	María Muñe
María Dolores Marrodán	Maria D. Madorran
María Dolores Marrodán	Maria Madorran
María Dolores Marrodán	María Dolores Marrodán
María Dolores Marrodán	María Marrodán
María Dolores Marrodán	M. Dolores Marrodán
María Dolores Marrodán	María D. Marrodán
María E. Díaz Sánchez	María Díaz
María E. Díaz Sánchez	María E. Díaz Sánchez
María Eugenia Luna	Maria E. Luna
María Eugenia Luna	María Eugenia Luna
María Florencia Cesani	Maria Cesani
María Florencia Cesani	María Cesani
María Florencia Cesani	Maria Cesani Rossi
María Florencia Cesani	María Cesani Rossi
María Florencia Cesani	María F. Cesani
María Florencia Cesani	María F Cesani
María Florencia Cesani	María F. Cesani
María Florencia Cesani	Maria F. Cesani Rossi
María Florencia Cesani	María F. Cesani Rossi
María Florencia Cesani	María Florencia Cesani
Marta Ruiz	Marta Ruiz
María Fernanda Torres	María F. Torres
María Fernanda Torres	María Torres



María Fernanda Torres	María Fernanda Torres
María Fucini	María C. Fucini
María Fucini	María Fucini
María Isabel González	María González
María Isabel González	María Isabel González
María Laura Fuchs	María I. Fuchs
María Laura Fuchs	María Laura Fuchs
María Lorena Zonta	María L. Zonta
María Lorena Zonta	Lorena Zonta
María Lorena Zonta	María L Zonta
María Lorena Zonta	Maria L. Zonta
María Lorena Zonta	María Zonta
María Lorena Zonta	Maria Zonta
María M. Ramón	María M. Ramón
María M. Ramón	María Ramon
María M. Reca	María M. Reca
María M. Reca	María Reca
María Soledad Mesa	María Mesa
María Soledad Mesa	María S. Mesa
María Soledad Mesa	María Soledad Mesa
Pedro Hidalgo Calcines	Pedro C. Hidalgo
Pedro Hidalgo Calcines	Pedro Hidalgo Calcines
Sergio Ivan Perez	Sergio I. Perez
Sergio Ivan Perez	Sergio Ivan Perez
Sergio Ivan Perez	Sergio Perez
Sergio Ivan Perez	S. Ivan Perez
Sergio Ivan Perez	S. Perez
Susana Blanca Ocampo	Susana B. Ocampo
Susana Blanca Ocampo	Susana Ocampo
Susana Blanca Ocampo	Susana Blanca Ocampo
María C. Barboza	María J. Barboza
María C. Barboza	María C. Barboza
María E. Díaz	María E. Díaz Sánchez
María E. Díaz	Maria E. Díaz
María Asunción Bordach	María A. Bordach
María Asunción Bordach	M. Asunción Bordach