

24 Resumen

25 El presente trabajo aborda el problema de análisis de redes de coautoría dentro de una revista
26 científica local en un periodo de 20 años desde su primera edición, que al momento de
27 escritura del trabajo no figuraba en la red scopus o web of science. Se extrajeron los datos vía
28 'minado web' (web mining) de la página oficial de la revista, se consolidó una base de datos y
29 finalmente se desarrolló una herramienta de software para poder realizar análisis de coautoría
30 con dichos datos. La herramienta es una aplicación web que consta de tres secciones
31 principales 'análisis preliminar de los datos' que permite explorar de forma resumida los datos
32 con los que se va a realizar el análisis; un 'análisis de periodos' donde se pueden seleccionar
33 uno o varios períodos a analizar, que incluye visualización con navegación interactiva de la red,
34 análisis de estructura del grafo, análisis de mundo pequeño, análisis de libre escala, búsqueda
35 de comunidades, breve análisis de títulos vía minado de texto (text mining); y una última
36 sección de 'análisis temporal' que tiene foco en mostrar las variaciones de la red de forma
37 gráfica e interactiva entre los períodos de análisis. Finalmente se realiza el análisis de la
38 revista, utilizando la herramienta para contestar las hipótesis planteadas en el capítulo inicial.

39 Palabras clave: redes de coautoría; minado web; minado de texto; análisis de componentes;
40 análisis de comunidades; aplicación web interactiva; análisis temporal; mundo pequeño; libre
41 escala; desambiguación de autores; R; shiny; open source.

42

44 Abstract

45 This work tackles coauthorship networks in a local scientific journal in a time lapse of 20 years
46 since its first edition, at writing time the network does not appear on scopus or web of science.
47 Data was gathered via web mining of the journal official website, making transformations to
48 consolidate a database and finally develop a software tool to do the coauthorship network
49 analysis over that data. The software is a web application composed by 3 sections: an
50 'exploratory data analysis' to know which data is used to make the coauthorship network; a
51 'yearly analysis' where you can select one or several years to aggregate data; this includes
52 interactive network visualization, graph structure analysis, small world analysis, free scale
53 analysis, identifying communities, brief title analysis via text mining ; and last section of temporal
54 analysis which focuses on variations of network over time that can be explored in an interactive
55 way. Finally we use the web application for the journal analysis and answer the hypothesis and
56 some questions.

57 Keywords: coauthorship networks; web mining; text mining; component analysis; community
58 analysis; interactive web application; temporal analysis; small world; free scale; author
59 disambiguation; R; Shiny; opensource

62 Agradecimientos

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

65 Indice

66	
67	Resumen 3
68	Abstract 4
69	Agradecimientos 5
70	Indice 6
71	Capítulo 1: Introducción 10
72	Introducción 10
73	Organización de la tesis 11
74	Características de la Antropología Biológica 12
75	Características de la RAAB 13
76	Conceptos clave 14
77	Redes y Grafos 14
78	Redes de Coautoría 15
79	Fuerza de colaboración 16
80	Grado 17
81	Intermediación 17
82	Componente 18
83	Mundo Pequeño 19
84	Antecedentes sobre análisis de coautorías 19
85	Objetivos 20
86	Objetivo 1: Desarrollar una herramienta para el análisis de redes de coautoría 20
87	Objetivo 2: Analizar redes de coautorías de la RAAB 20
88	Hipótesis 21
89	Resumen del capítulo 22
90	Capítulo 2: Generación del conjunto de datos 23
91	Introducción 23
92	Materiales y métodos 26
93	Herramientas de trabajo 26
94	Introducción al proceso 26
95	Análisis de la estructura del sitio 27
96	Descripción de los datos de interés 28
97	Adquisición de los datos 29
98	Parseo de los archivos 30
99	Parseo de los archivos HTML 30
100	Parseo de los archivos bibTex 30
101	Parseo de los archivos PDF 30

102	Generación de la Base	31
103	Resultados intermedios	31
104	Generación de la base relacional	31
105	Estructura de la base final	31
106	Generación de los grafos	33
107	Descripción de las propiedades de los nodos y aristas	34
108	Propiedades de los Nodos	34
109	Propiedades de las Aristas	35
110	Resumen del capítulo	36
111	Capítulo 3: Creación de la Herramienta para análisis	37
112	Introducción	37
113	Materiales y métodos	37
114	Estructura Interna de la aplicación	37
115	Herramientas de trabajo	37
116	Respecto de la estructura de la aplicación	38
117	Disponibilidad local, en la red, en internet	39
118	Estructura de la interfaz gráfica	40
119	Sección 1: Consideraciones	40
120	Datos Crudos	41
121	Composición	42
122	Resumen	43
123	Preguntas	44
124	Sección 2: Análisis Estático	46
125	Configuración	46
126	Visualización	48
127	Artículos asociados	50
128	Estructura de la red	51
129	Métricas de la red	51
130	Métricas de los nodos	53
131	Detalles de los componentes	56
132	Comparación de modelos	59
133	Validación de Mundo Pequeño	63
134	Validación de Libre Escala	64
135	Comunidades	66
136	Sección 3: Análisis temporal	70
137	Animación del grafo de manera acumulada	71
138	Medidas de cada periodo	72
139	Medidas Acumuladas a cada periodo	72
140	Ranking de autores en el tiempo	73
141	Ranking de autores en el tiempo para medidas acumuladas	74
142	Resumen del capítulo	74

143	Capítulo 4: Análisis de la Red	75
144	Introducción	75
145	Hipótesis 1: La cantidad de autores aumenta en el tiempo debido al crecimiento de la	
146	disciplina en el país, lo que se refleja en un aumento en la cantidad de nodos en la red	78
147	Autores en artículos por año	83
148	Hipótesis 2: Los autores fundadores se refuerzan en el tiempo, lo cual puede ser medido	
149	por grado, intermediación y fuerza de colaboración	85
150	Hipótesis 3: Los componentes más grandes absorben a los más pequeños a lo largo del	
151	tiempo, dado que autores periféricos pasan a trabajar en conjunto con autores principales	90
152	Verificación de absorción de componentes en la red	92
153	Relación con actividades de investigación	95
154	Búsqueda de comunidades	97
155	Explicación de la métrica de evaluación de algoritmos: Modularidad	98
156	Explicación del algoritmo seleccionado: Optimización de modularidad multi-nivel	99
157	Búsqueda de sentido en los grupos obtenidos	100
158	Explicación de criterio de búsqueda: verificación de comunidades según títulos	
159	publicados.	101
160	Hipótesis 4: La red agregada total (desde 1996 a 2016) es de mundo pequeño puesto que	
161	es una característica presente en la mayoría de las redes de coautorías científicas	107
162	Mundo pequeño	107
163	Libre Escala	108
164	Caso aplicado	108
165	Mundo pequeño para la red agregada total	108
166	Libre escala para la red total	110
167	Mundo pequeño para el periodo 1996 - 2001	111
168	Libre escala 1996 - 2001	113
169	Mundo pequeño para el periodo 2002 - 2006	114
170	Libre escala 2002 - 2006	116
171	Mundo pequeño para el periodo 2007 - 2011	117
172	Libre escala 2007 - 2011	118
173	Mundo pequeño para el periodo 2012 - 2016	120
174	Libre escala 2012 - 2016	122
175	Resumen de análisis de las propiedades de mundo pequeño y libre escala	123
176	Comparación con otras redes de coautoría	124
177	Resumen de capítulo	125
178	Capítulo 5: Conclusiones	127
179	Resumen	127
180	Productos asociados	127
181	Consideraciones sobre el conjunto de datos y la variación de estructura del sitio en el	
182	tiempo	128
183	Trabajos futuros	128
184	Actualización de la base de datos	128

185	Aspectos técnicos	129
186	Proyección sobre artículos	129
187	Análisis de textos y tópicos en el tiempo	129
188	Análisis de conectividad / puentes	129
189	Bibliografía	131
190	ANEXO: software y librerías utilizadas	135
191	ANEXO: Detalle de algoritmos para búsqueda de comunidades	139
192	ANEXO: Detalle procesamiento bibTEX	140
193	Problemas de transformación y encoding bibtex	140
194	Problemas autores	141
195	Problemas con el volumen de la revista	141
196	Problema Abstracts Vacíos en bibtex	142
197	ANEXO: Detalle consideraciones PDF	144
198	Obtención de los pdf	144
199	Parseo de los pdf	144
200	ANEXO: Detalle estandarización autores	145
201	Revisión y validación con expertos	146
202	Iteraciones de clustering	146
203	Resultado de Revisión y validación con expertos	152
204		
205		

207 Capítulo 1: Introducción

208 Introducción

209 Las redes de coautoría son una herramienta útil para analizar patrones de colaboración
210 científica. Estos patrones pueden ser estudiados dentro de una disciplina, región geográfica o
211 revista científica, así como en forma comparativa ([Kumar 2015](#)). Al analizar redes de coautoría
212 en una disciplina particular, las características relacionadas al grado de cohesión o
213 fragmentación, el tamaño de los nodos y componentes o el patrón de conexión entre nodos,
214 entre otras, brindan información sobre el diálogo colaborativo entre los distintos grupos de
215 investigación que la conforman y la “importancia” relativa de cada uno en el contexto particular.
216

217 En esta tesis se analizó una red de coautorías de la Revista Argentina de Antropología
218 Biológica (RAAB), como ejemplo de análisis dentro de una disciplina particular y a nivel local
219 (Argentina). La RAAB es el órgano oficial de comunicación científica de la Asociación de
220 Antropología Biológica Argentina (AABA), por lo que el análisis de coautorías brindó una
221 oportunidad para el estudio de las dinámicas de colaboración dentro de esta disciplina a nivel
222 nacional. En particular, resultó de interés analizar la estructura de la red como reflejo de las
223 interacciones entre distintos grupos de investigación y obtener una aproximación al estudio del
224 crecimiento y desarrollo de una disciplina de origen relativamente reciente en Argentina.

225 Para ello, en esta tesis se propuso el desarrollo de una herramienta de visualización y análisis
226 de la red, a fin de facilitar el estudio de este caso en particular, así como también su aplicación
227 a otros similares. Siguiendo lo propuesto por [Morel, Serruya, et al. \(2009\)](#), se analizó la red a
228 fin de evaluar el incremento en el número de publicaciones, grupos de investigación y
229 colaboraciones entre los grupos, así como el cambio en estas dinámicas a lo largo del tiempo.
230
231
232

234 Organización de la tesis

235

236 La presente tesis está organizada en cinco capítulos que abarcan las siguientes temáticas:
 237 introducción (Capítulo 1), metodología (Capítulos 2 y 3), resultados y análisis (Capítulo 4), y
 238 discusión y perspectivas (Capítulo 5) (ver [Fig. 1.1](#)).

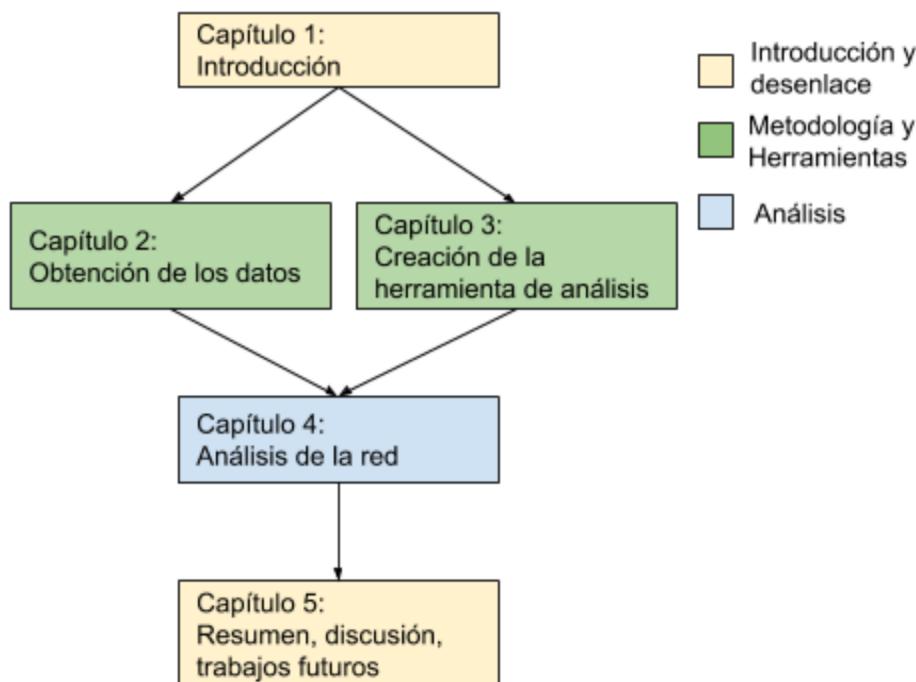


Fig. 1.1: Organización general de la presente tesis

239

240 En el Capítulo 1 tratamos el contexto general de la presente tesis y una breve justificación de lo
 241 realizado en los capítulos siguientes. Se presentan los objetivos e hipótesis subyacentes a este
 242 trabajo.

243

244 En el Capítulo 2 se especifica cómo se obtuvieron los datos, qué transformaciones realizamos
 245 sobre ellos, y finalmente cómo dimos origen a la base de datos que se utilizó para el desarrollo
 246 de la aplicación.

247

248 El tercer capítulo cuenta la estructura de la aplicación a nivel *software*, las secciones que
 249 componen la aplicación, así como el fundamento de estas funcionalidades que componen cada
 250 sección.

251

252 En el Capítulo 4 realizamos el análisis del caso específico de la RAAB usando la base de datos
 253 creada en el Capítulo 2, la herramienta creada en el Capítulo 3, a fin de contrastar las hipótesis
 254 planteadas inicialmente. Además, se ejemplifica cómo con la herramienta desarrollada se
 255 pueden realizar con facilidad varias tareas comunes en el análisis de redes de coautoría.

256

257 Finalmente en el Capítulo 5, último capítulo, resumimos las tareas realizadas, y se discute el

258 proceso, los resultados obtenidos y los potenciales trabajos futuros.
 259

260 Características de la Antropología Biológica

261 La Antropología Biológica puede definirse como “el estudio de la evolución y variabilidad
 262 biológica de los humanos en el presente y en el pasado” ([Madrigal y González-José, 2016:5](#)).
 263 Dentro de esta disciplina se estudian procesos y fenómenos de las poblaciones humanas
 264 desde tres ejes o puntos de vista principales: evolutivo, trans-cultural y biocultural ([Madrigal y](#)
 265 [González-José, 2016](#)). Como ejemplo de estos enfoques, [Madrigal y González-José \(2016:6\)](#)
 266 presentan el caso particular del parto humano:

267
 268 *Por lo tanto, aún si estamos estudiando algo tan clínico como, por ejemplo, el parto, lo vemos*
 269 *desde estos tres puntos de vista:*

270

271 1. **Evolutivo:** *¿Cuál es el papel de la pelvis en la locomoción y en el parto? ¿Cuándo en*
 272 *nuestra historia evolutiva los humanos empezamos a caminar de manera bípeda, así afectando*
 273 *la pelvis? ¿Cómo paren los primates no humanos?*

274

275 2. **Transcultural:** *¿De qué manera las familias y los asistentes médicos ayudan a las madres*
 276 *que están pariendo en culturas alrededor del mundo? ¿Es posible que la manera de parir típica*
 277 *de la biomedicina no sea la mejor para la mujer?*

278

279 3. **Biocultural:** *¿Cómo es que el estrés de estar pariendo en un hospital afecta el éxito del*
 280 *parto para mujeres que están aisladas de su familia? ¿Podríamos mejorar la salud de madre y*
 281 *bebé incorporando prácticas culturales de la mujer que está pariendo?*

282

283 En otras palabras, los antropólogos biológicos estudian un hecho biológico como una
 284 enfermedad o un gen, incorporando información acerca del pasado evolutivo de dicha
 285 enfermedad o gen, incorporando información acerca de la manifestación biológica de la
 286 enfermedad o gen en diferentes culturas-ecosistemas, y estudiando cómo la enfermedad o gen
 287 afecta la cultura donde se manifiesta, y cómo la cultura afecta la enfermedad o gen.

288

289 La Antropología Biológica como disciplina constituye un campo interesante para analizar redes
 290 de colaboración científica puesto que, como se desprende del ejemplo anterior, la articulación
 291 de diferentes enfoques puede dar lugar al trabajo transversal entre distintas líneas de
 292 investigación. En este marco, el análisis de coautorías en la Revista Argentina de Antropología
 293 Biológica (RAAB) brindaría una primera aproximación al estudio de las colaboraciones dentro
 294 de esta disciplina a nivel nacional.

296 Características de la RAAB

298 La Asociación de Antropología Biológica Argentina (AABA) comienza a gestarse hacia fines de
 299 la década de 1980. Una de las necesidades existentes era crear una revista que documentara
 300 la producción científica dentro de la disciplina en nuestro país. De esta forma, uno de los
 301 objetivos de la AABA fue la creación de la "Revista Argentina de Antropología Biológica"
 302 (RAAB) la cual constituye el órgano oficial de difusión de la Asociación ([Cesani, Sardi, et al.
 303 2013](#))

305 En la [página web de la RAAB](#) puede encontrarse una descripción de sus características
 306 generales: [<https://revistas.unlp.edu.ar/raab/about>]

Enfoque y alcance

La Revista Argentina de Antropología Biológica publica estudios originales y de significación en los campos de la antropología biológica: evolución y variación humana, adaptación, genética, crecimiento y desarrollo, antropología forense, experimentación, morfología y primatología. Además, publica temas de antropología social, arqueología, psicología evolutiva, geografía humana, demografía, ecología, epidemiología, museología, conservación, historia y filosofía de la ciencia, medicina, entre otros, siempre que los contenidos tengan una vinculación explícita con algún área en particular de la antropología biológica.

Sobre La RAAB

Es propiedad de la AABA. Se comenzó a publicar en 1996 con el objetivo de difundir estudios originales en los campos de la Antropología Biológica y temas afines. Actualmente se publican dos números por año que incluyen artículos científicos y de revisión, comentarios bibliográficos, noticias breves, resúmenes de tesis de postgrado, cartas al editor, reportes técnicos y comunicaciones breves sobre diversos temas de interés en la especialidad. Los idiomas de publicación son español, portugués e inglés. La recepción de manuscritos está abierta todo el año y se publican en orden de aceptación. El proceso de evaluación de los manuscritos sigue la modalidad de revisión por pares.

Indizada en

Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal LATINDEX (Folio N° 7579, versión papel; Folio N° 20444, versión on-line)

Servicio de Difusión de la Creación Intelectual de la UNLP (SeDiCI)

HOLLIS Catalog, Harvard University (012566584)

Núcleo Básico de Revistas Científicas Argentinas (Resolución N° 1855/13 CONICET)

SciELO Argentina <http://www.scielo.org.ar/>

Redalyc. Red de Revistas Científicas de América Latina y el Caribe, España y Portugal

308

309 Respecto de la estructura de la revista, generalmente su publicación es en uno o varios
310 volúmenes anuales. A estos volúmenes a lo largo de la tesis también los llamamos "ediciones".
311 Cada volumen tiene artículos publicados, estos pueden ir en diferentes categorías. El foco de
312 esta tesis está centrado en los de tipo "Trabajo Original"

313 Conceptos clave

314 Redes y Grafos

315

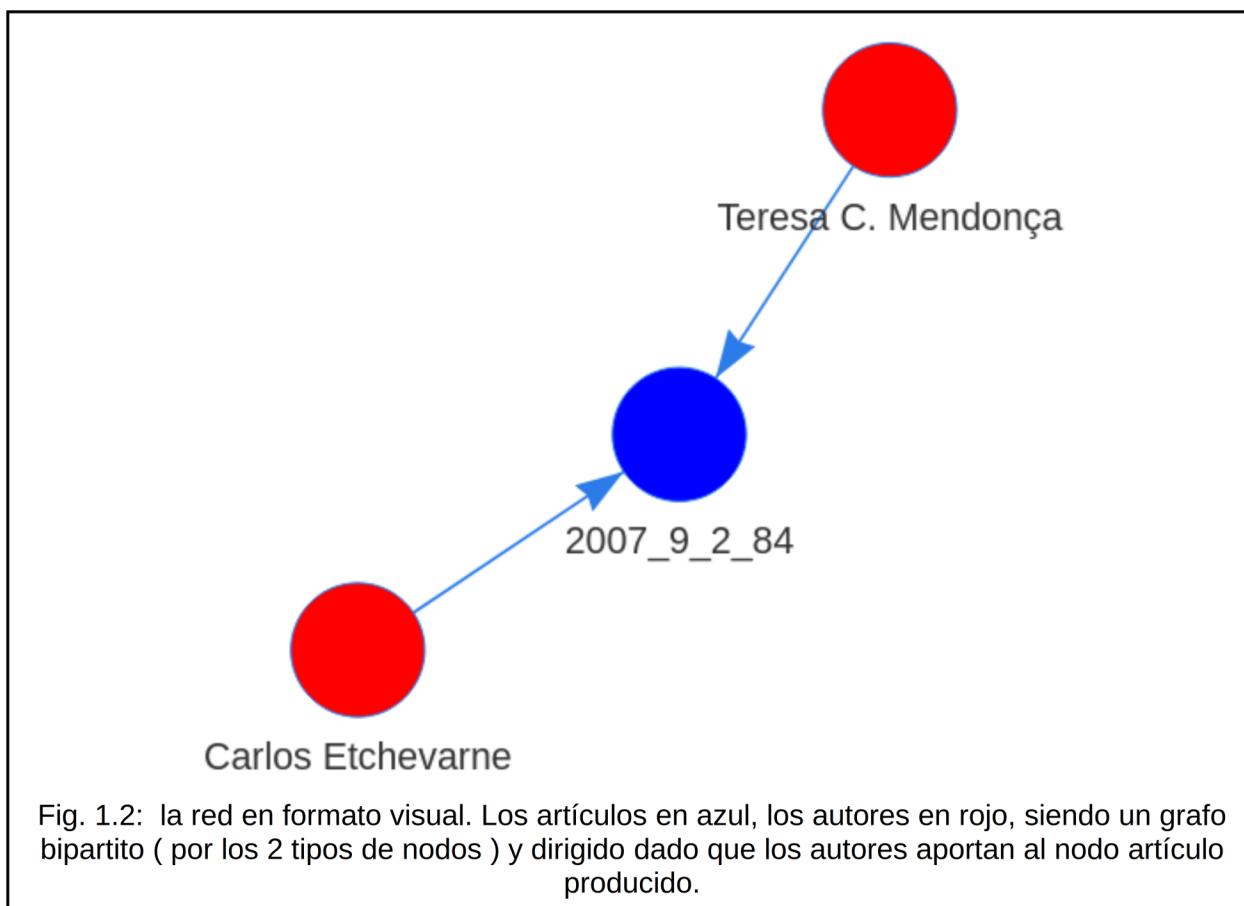
316 En palabras de [Barabási \(2016\)](#), traducidas y adaptadas por el autor de esta tesis:

317

318 *Si queremos entender sistemas complejos, primero necesitamos saber cómo sus componentes
319 interactúan entre ellos. En otras palabras, necesitamos un mapa de cómo están conectados.
320 Una red es un catálogo de los componentes de un sistema, comúnmente llamados nodos o
321 vértices, y las interacciones directas entre ellos, llamadas conexiones o aristas. Este tipo de
322 representación de redes ofrece un lenguaje común para estudiar sistemas que pueden diferir
323 enormemente en naturaleza, apariencia o alcance.*

324

325 En nuestro caso la red está compuesta de "nodos autores", "nodos artículos" y las conexiones
326 representan "autor escribió artículo"; este tipo de redes que tienen 2 tipos de nodos se llaman
327 bipartitas. A lo largo de la tesis entraremos en detalles respecto de los grafos. Un ejemplo de la
328 naturaleza de la red en formato visual se puede ver en la [Fig. 1.2](#).
329



330 Redes de Coautoría

331 Extraído del resumen de [Kumar \(2015\)](#) que a su vez cita a [Newman \(2001\)](#). Traducido y
332 adaptado por el autor de la tesis:

333

334 *Una red social es una representación de dos entidades o más conectadas entre ellas a través
335 de algún tipo de relación. Si uno acota las redes sociales humanas para las cuales existen
336 datos cuantitativos, quedan pocas: una de estas son las redes de los actores de películas en
337 las que los nodos actores trabajan juntos en una película; y otra es la red de investigadores,
338 donde los investigadores escriben un trabajo científico en conjunto ([Newman 2001](#)) De estas
339 redes la que probablemente tiene más función social es la red de investigadores. A diferencia
340 de la red de películas y actores, los investigadores eligen con quien hacer investigación y luego
341 escriben los resultados del trabajo en la forma de un artículo científico con coautores. Estas
342 colaboraciones dejan huellas digitales en la forma de bibliografía, que luego puede ser seguida
343 y evaluada.*

344

345 El texto de Newman fue escrito antes de que internet fuera adoptada masivamente. Hoy en día
346 abundan muchos casos de redes, sin embargo queda investigar cuales existen con el nivel de
347 calidad que tiene una red de escritura científica.

348

349 Las redes de coautoría se presentan al realizar la proyección de la red sobre los nodos autores,
350 quedando los autores como nodos, y los nodos artículos son reemplazados para formar las
351 relaciones, a las cuales se les da más peso según la cantidad de trabajos realizados en

352 conjunto.

353

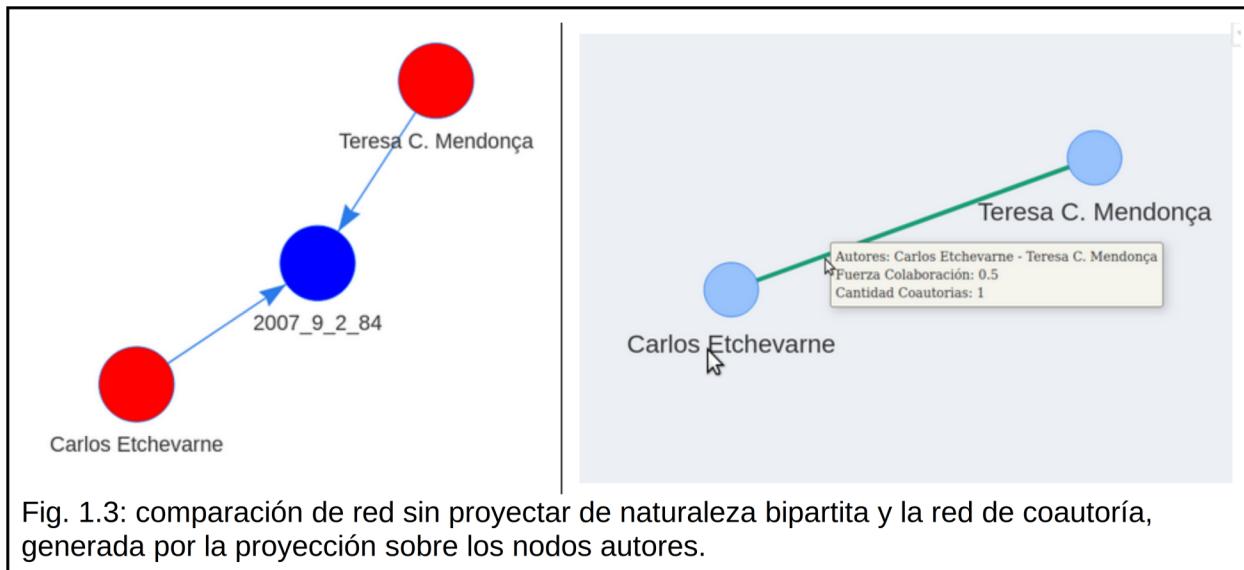


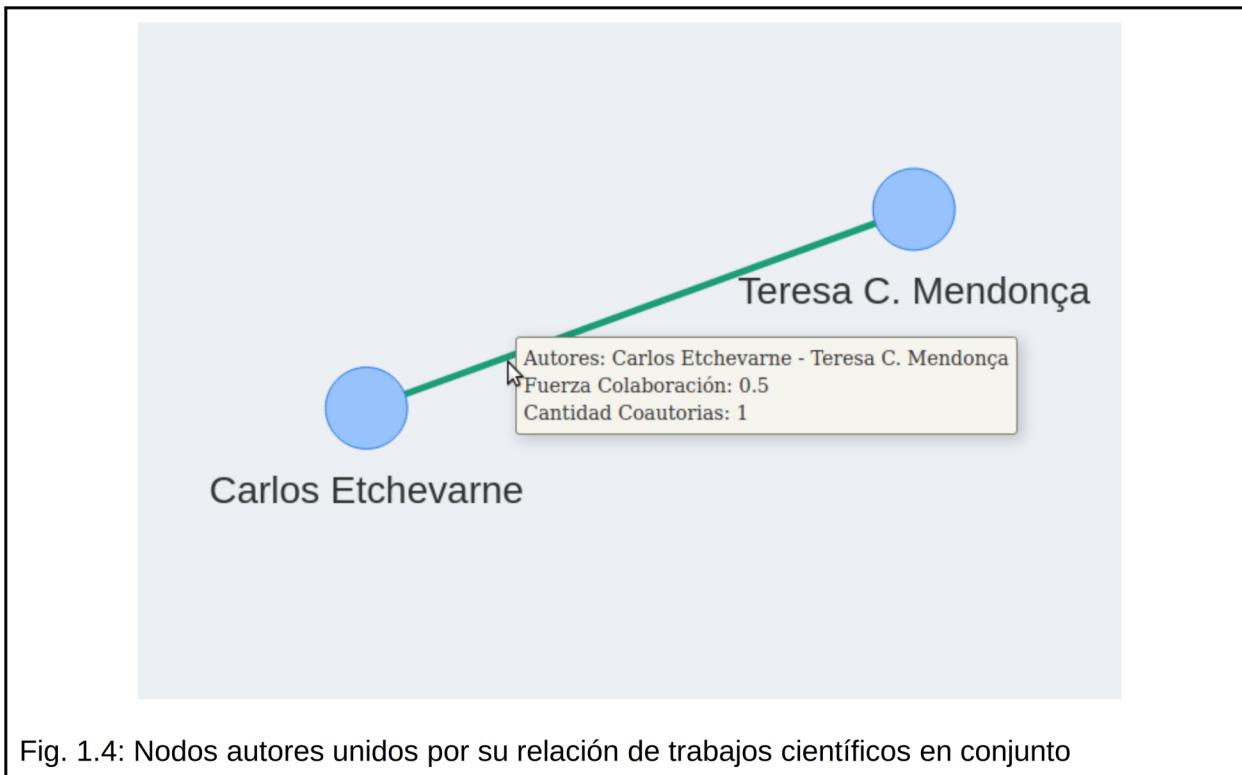
Fig. 1.3: comparación de red sin proyectar de naturaleza bipartita y la red de coautoría, generada por la proyección sobre los nodos autores.

354

355 Para entender la proyección podemos ver en paralelo la red sin proyectar y la red proyectada
356 sobre los autores en la [Fig. 1.3](#)

357 Fuerza de colaboración

358 Es una Métrica descripta en el paper de [Newman \(2004\)](#), representa el artículo creado dividido
359 la cantidad de autores que lo escriben, esta métrica plasma cómo se reparte el esfuerzo de la
360 generación de un artículo realizando una cuenta simple como dividir el esfuerzo para realizar
361 un artículo por su cantidad de autores. En la [Fig. 1.4](#) vemos que realizaron 1 trabajo en
362 conjunto entre 2 autores, por lo tanto $1 \text{ trabajo} / 2 \text{ autores} = 0.5$ de fuerza de colaboración.
363



366 En la [Fig. 4](#) se observan 2 nodos autores unidos por su relación. Cuando vemos los detalles de
 367 esta relación observamos que se tuvo en cuenta la medida llamada Fuerza de colaboración, del
 368 inglés *collaboration strength*.

369 Grado

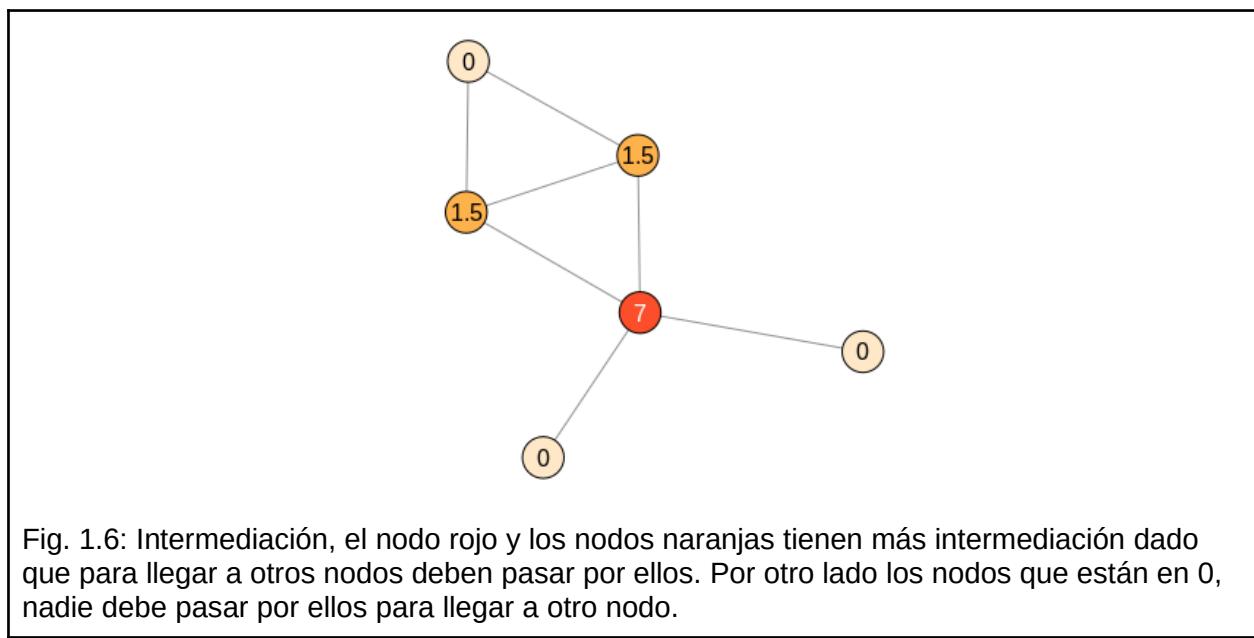
370 El grado del nodo es la cantidad de conexiones directas del nodo. Como ejemplo se puede ver
 371 la [Fig. 1.5](#).
 372



Fig. 1.5: nodos y sus grados. Sin conexiones grado 0 , 1 conexión grado 1 .. 3 conexiones grado 3.

373 Intermediación

374 La intermediación representa la importancia de un nodo según su participación en caminos
 375 para llegar de un nodo a otro. Vemos un ejemplo de esta métrica en la [Fig. 1.6](#).



376 Componente

377 Un componente es una porción de red en donde todos los nodos están conectados directa o
 378 indirectamente por lo menos por una arista. El análisis de componentes da un vistazo general
 379 de la estructura de la red (Fig. 1.7) y provee información valiosa en relación a su nivel de
 380 fragmentación ([Morel, et al. 2009](#)).
 381

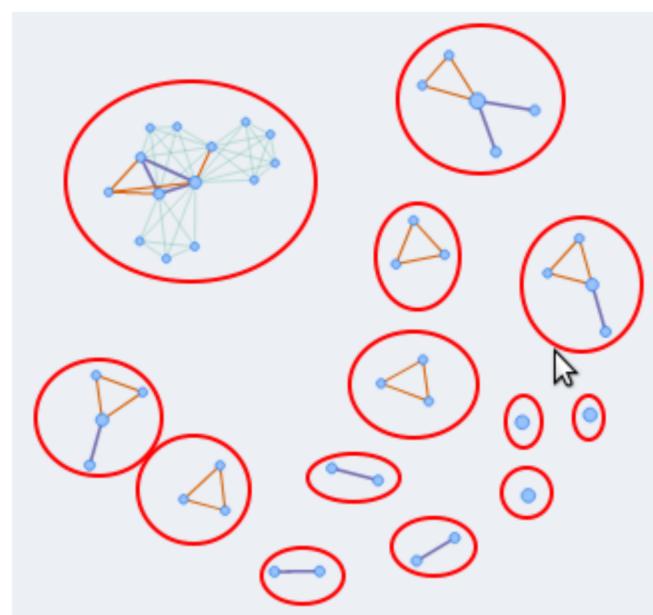


Fig. 1.7: Ejemplo de gráfico de componentes obtenido para la red de la RAAB correspondiente al año 1996. Los componentes se resaltan en rojo.

384

Mundo Pequeño

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

388

389 Para comprender un poco mejor el estudio de “Small world”, surge de la frase en inglés ‘My it's
 390 a small world’ / ‘Que mundo pequeño’ cuando dos personas conocen a otra y, por lo tanto,
 391 estarían interconectados a través de una tercera. Este estudio de [Milgram \(1967\)](#) consistió en
 392 enviar mensajes desde una persona en un punto A hasta otra en un punto B bajo ciertas reglas,
 393 entre ellas, que solo podían mandar el mensaje a personas conocidas inmediatas.

394

395 Para ganar una intuición gráfica podemos revisar la [Fig. 1.8](#) donde el autor en el círculo rojo
 396 desea contactar con el autor en el círculo azul, recordemos que las aristas que unen son
 397 trabajos realizados en conjunto.

398

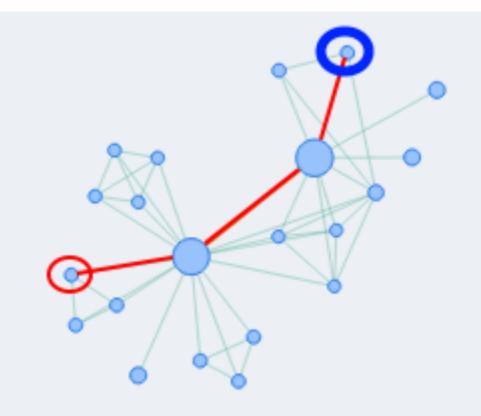


Fig. 1.8: Si hablase el nodo del círculo rojo que desea contactarse con el nodo azul “*El coautor de mi trabajo original, tiene otro coautor que trabajó con nodo en círculo azul que nos podría presentar*”

399

400 Nos interesa revisar esta propiedad para la red de la RAAB dado que es una característica
 401 frecuente en el contexto de redes científicas según [Kumar \(2015\)](#).

402

403

Antecedentes sobre análisis de coautorías

404

405 El análisis de redes sociales es un tópico que existe desde aproximadamente la década de
 406 1950, lo que se denominaba sociometría. Se hizo popular en la década de 1960 por el estudio
 407 de mundo pequeño de [Milgram \(1967\)](#), y luego volvió a tener popularidad durante la década de
 408 1990 con el juego “6 grados de Kevin Bacon” que medía qué tan lejos alguien del círculo de
 409 actuación en Hollywood distaba del actor Kevin Bacon medido en interacciones por compartir
 410 películas o series.

411

412 En bibliometría, las redes no son algo nuevo, generalmente se estudiaron las redes de citación,
 413 en las que los nodos son los papers y las citas las aristas. Sin embargo, las redes de coautoría,
 414 presentan otro enfoque cercano a la comunidad científica más que al conocimiento, tal vez por
 415 eso tuvieron menos atención anteriormente. A pesar de esto nos puede brindar mucha

416 información ([Newman 2004](#))

417

418 En este trabajo utilizamos varios de los estudios mencionados en la revisión de [Kumar \(2015\)](#);
419 como por ejemplo: desambiguación de nombres de autor, el estudio de mundo pequeño, la
420 búsqueda de comunidades, la visualización de la red, tamaño del componente más grande,
421 entre otros. Cada uno de estos es profundizado a lo largo del estudio.

422

423 Un enfoque deseado para esta tesis fue el punto de vista temporal dado que exploramos el
424 periodo de 20 años desde el inicio de la revista, para ello nos basamos en parte del estudio
425 realizado por [Koseoglu \(2016\)](#); que revisa temas como evolución temporal de una revista
426 específica, evaluando métricas, visualizaciones y comparaciones.

427 Objetivos

428 El **objetivo general** de la presente tesis es realizar una aporte a la comprensión del desarrollo
429 de la Antropología Biológica en nuestro país a través del análisis de una red de colaboración
430 científica. Para ello, se plantearon dos objetivos específicos.

431 **Objetivo 1:** Desarrollar una herramienta para el análisis de redes de
432 coautoría

433 Se planteó la creación de una herramienta para facilitar el análisis de redes de coautoría
434 científica mediante la simplificación de tareas repetitivas, como por ejemplo: variación de
435 parámetros para aplicar filtros y/o algoritmos, selección de diferentes nodos/aristas/grupos,
436 aplicación de filtros a los datos, visualización general de la red desde diferentes puntos de
437 vista, obtención de detalles de nodos y aristas, obtención de detalles de los datos, entre otras.
438 Dentro de este objetivo se buscó, además, facilitar la exploración de los datos mediante
439 interacciones intuitivas brindadas a través de una interfaz web de acceso abierto para
440 diferentes personas que puedan realizar el o los análisis. Los tres focos de la herramienta son
441 un breve análisis de los datos que dan origen a la red, el análisis de la red para ciertos períodos
442 seleccionados, y finalmente el análisis temporal de la red.

443 **Objetivo 2:** Analizar redes de coautorías de la RAAB

444 Utilizando la herramienta desarrollada, se propuso analizar el caso particular de la Revista
445 Argentina de Antropología Biológica (RAAB) y nos interesarnos principalmente en cómo fue el
446 comportamiento de la red en el tiempo.

447

448 Hipótesis

449 En congruencia con lo planteado en el objetivo 2, nos planteamos las siguientes hipótesis:

450

- 451 - Hipótesis 1: La cantidad de autores aumenta en el tiempo debido al crecimiento de la
- 452 disciplina en el país, lo que se refleja en un aumento en la cantidad de nodos en la red.
- 453 - Hipótesis 2: Los autores fundadores se refuerzan en el tiempo, lo cual puede ser
- 454 medido por grado, intermediación y fuerza de colaboración.
- 455 - Hipótesis 3: Los componentes más grandes absorben a los más pequeños a lo largo del
- 456 tiempo, dado que autores periféricos pasan a trabajar en conjunto con autores
- 457 principales.
- 458 - Hipótesis 4: La red agregada total (desde 1996 a 2016) es de mundo pequeño puesto
- 459 que es una característica presente en la mayoría de las redes de coautorías científicas.

460

461 Estas hipótesis tienen sus fundamentos en:

462

463 Hipótesis 1: "*La cantidad de autores aumenta en el tiempo debido al crecimiento de la*
464 *disciplina en el país, lo que se refleja en un aumento en la cantidad de nodos en la red*" : nos
465 interesa verificar si la disciplina de Antropología Biológica crece en el país analizando la
466 cantidad de autores a lo largo del tiempo en la Revista Argentina de Antropología Biológica.
467 Acotamos el estudio a dicha publicación dado que buscamos representación de la materia a
468 nivel local, entendemos que esta hipótesis puede ser expandida en publicaciones
469 internacionales por autores locales y en otras revistas, lo cual queda fuera del alcance de la
470 presente tesis.

471

472 Hipótesis 2: "*Los autores fundadores se refuerzan en el tiempo, lo cual puede ser medido por*
473 *grado, intermediación y fuerza de colaboración.*": nos interesa analizar si el avance de la revista
474 depende de los autores fundadores, o si está impulsada por nuevos autores y proponemos
475 métricas con las cuales poder evaluar esta actividad.

476

477 Hipótesis 3: "*Los componentes más grandes absorben a los más pequeños a lo largo del*
478 *tiempo, dado que autores periféricos pasan a trabajar en conjunto con autores principales*" nos
479 interesa para tratar de entender en mayor profundidad el comportamiento interno de la red
480 revisando cómo se comportan los componentes. Para lograr esto analizamos la evolución
481 temporal del tamaño del componente principal identificando momentos específicos de
482 absorción de componentes.

483

484

485 Hipótesis 4: “*La red agregada total (desde 1996 a 2016) es de mundo pequeño puesto que es*
486 *una característica presente en la mayoría de las redes de coautorías científicas*” nos interesa
487 para comprar con puntos de vista externos. Realizamos validación de propiedades comunes
488 que se dan en redes del mundo real como Mundo pequeño y Libre escala. De forma
489 complementaria para complementar la vista hacia afuera, contrastamos con redes de coautoría
490 existentes en trabajos previos.

491 Resumen del capítulo

492 Se puso en contexto el trabajo realizado, la estructura, se detallaron objetivos e hipótesis a
493 tratar en esta tesis. En los siguientes capítulos desarrollaremos en profundidad lo planteado en
494 los objetivos y verificaremos las hipótesis.

495

496

497

498 Capítulo 2: Generación del conjunto de datos

499 Introducción

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

506

507 Para realizar la extracción de datos se usó *web-scraping*. Para comprender mejor qué significa
 508 esta técnica a continuación se presenta un fragmento de [Saurkar et al.\(2018\)](#), adaptado al
 509 español y contextualizado para este caso por el autor de esta tesis:

510

511 *La extracción de los datos de una página web de forma manual es una tarea exhaustiva y
 512 consumidora de tiempo. Debido a esa limitación, se crean las técnicas de web-scraping. Al
 513 usar técnicas de web-scraping el usuario puede extraer información disponible en múltiples
 514 sitios web en una sola base de datos u hoja de cálculo. La técnica de web-scraping es
 515 subdisciplina de la tecnología de minado de la red mining.*

516 *Web mining se encuentra entre las técnicas de Recuperación de la información, Extracción de
 517 la información y Minado de datos. Tanto la Recuperación como la Extracción de la información
 518 cumplen roles importantes en el seguimiento y obtención de información valiosa a partir de
 519 datos no estructurados, antes de que esa información esté en condiciones de ser procesada
 520 por herramientas y técnicas de Minado de datos.*

521

522 Podemos decir que son una serie de pasos necesarios para poder disponer de los datos de
 523 forma adecuada para un posterior uso en gran parte automatizados. Para este caso específico
 524 se busca generar dos estructuras resultantes: una estructura de base de datos relacionales y
 525 otra en formato de grafos.

526

527 Acorde a [Bourgeois \(2014\)](#) una base de datos relacional es un sistema que organiza los datos
 528 en forma de tablas que están relacionadas. Cada tabla está compuesta por columnas o
 529 campos que definen la naturaleza de los datos guardados en esa tabla. Luego, un registro
 530 corresponde a una instancia particular de un conjunto de campos de una tabla, es decir, son los
 531 datos de cada elemento ([Fig. 2.1](#))

532

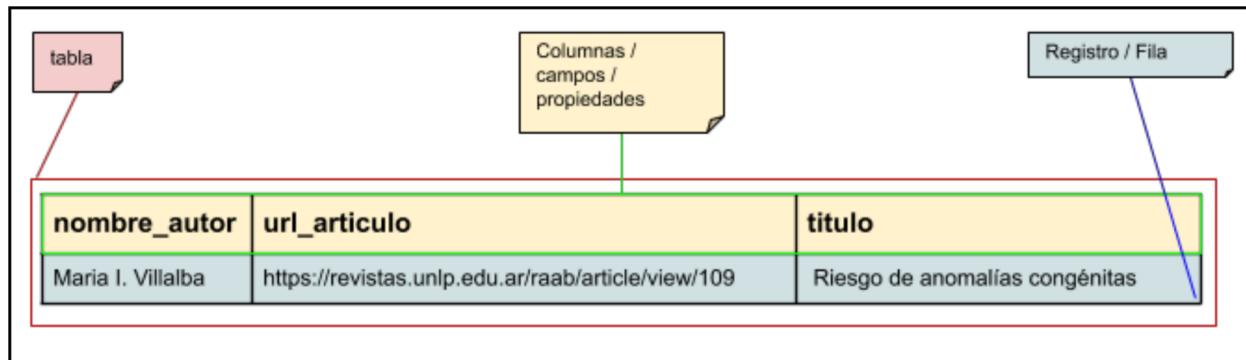


Fig. 2.1: Ejemplo de una tabla en una base de datos relacional.

533

534 Las relaciones entre las tablas se realizan a través de identificadores que permiten conectarlas
 535 (Fig. 2.2).

536

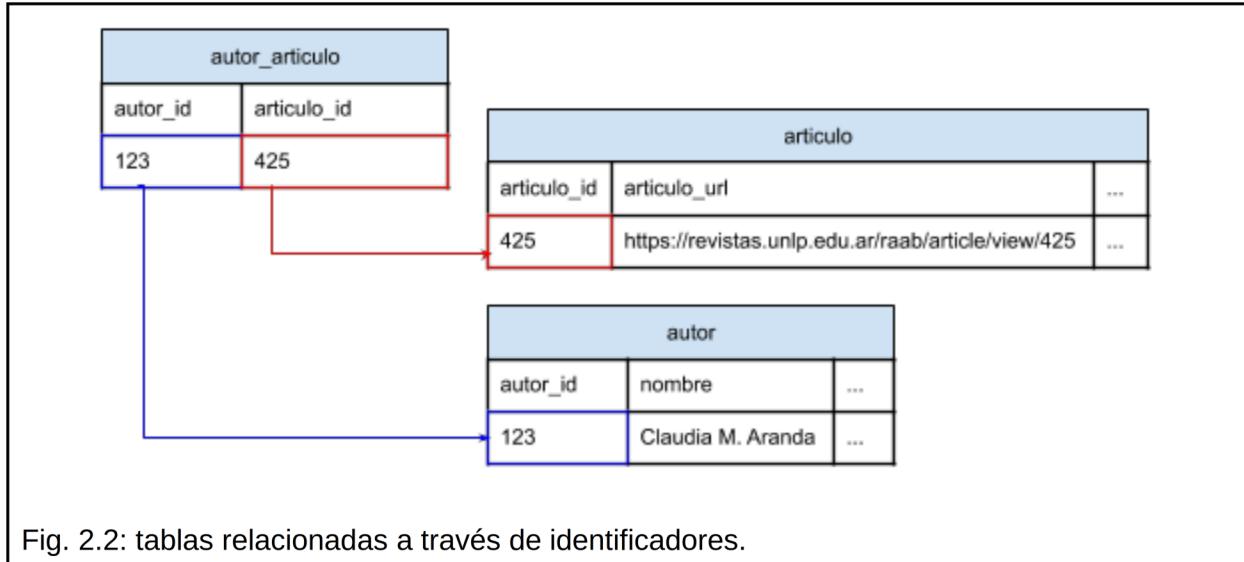


Fig. 2.2: tablas relacionadas a través de identificadores.

537

538 En esta tesis uno de los principales objetivos fue encontrar cómo están conectados los autores
 539 participantes de la revista, a través de los artículos publicados en ella. Para disponer de una
 540 estructura que ayude a comprender este sistema se utilizaron grafos. [Barabási \(2016\)](#) brinda
 541 una breve explicación al respecto, adaptada al español por el autor de esta tesis:

542

543 *Si queremos entender sistemas complejos, primero necesitamos saber cómo sus componentes
 544 interactúan entre ellos. En otras palabras, necesitamos un mapa de cómo están conectados.
 545 Una red es un catálogo de los componentes de un sistema, comúnmente llamados nodos o
 546 vértices, y las interacciones directas entre ellos, llamadas conexiones o aristas. Este tipo de
 547 representación de redes ofrece un lenguaje común para estudiar sistemas que pueden diferir
 548 enormemente en naturaleza, apariencia o alcance.*

549

550

551 Este mismo autor ofrece una consideración interesante respecto de la terminología utilizada en
 552 cuanto a redes o grafos [Barabási \(2016\)](#):

553

¿Redes o Grafos?

En la literatura científica se intercambia frecuentemente la terminología entre redes y grafos:

<i>Redes</i>	<i>Grafos</i>
<i>Nodos</i>	<i>Vertices</i>
<i>Conexiones</i>	<i>Aristas</i>

Sin embargo, hay una pequeña diferencia entre estas terminologías, el conjunto {red, nodo, conexión} se refiere a sistemas de la vida real, mientras que los términos {grafo, vértice, arista} se utilizan para la discusión de representaciones matemáticas de esas redes. Finalmente, esta distinción raramente es tenida en cuenta y se usan como sinónimos.

554

555 Por la naturaleza de las redes de autoría, usualmente tenemos dos tipos de nodos: autores y
 556 artículos. Sus conexiones son dirigidas y muestran la relación “autor escribe artículo” ([Fig. 2.3](#)).

557

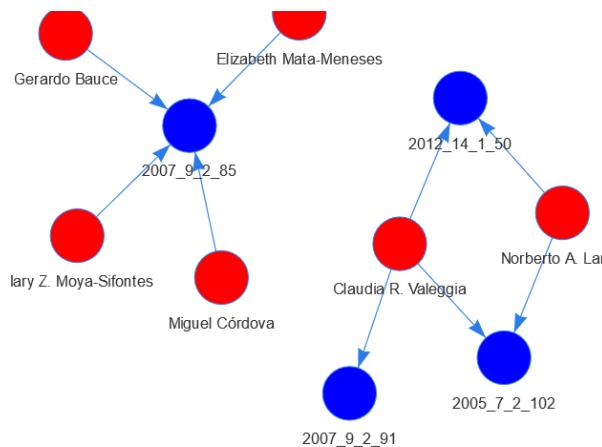


Fig. 2.3: ejemplo de red de autoría con dos tipos de nodos: autores y artículos, y su relación “autor escribe artículo”; Siendo los autores en rojo y los artículos en azul.

558

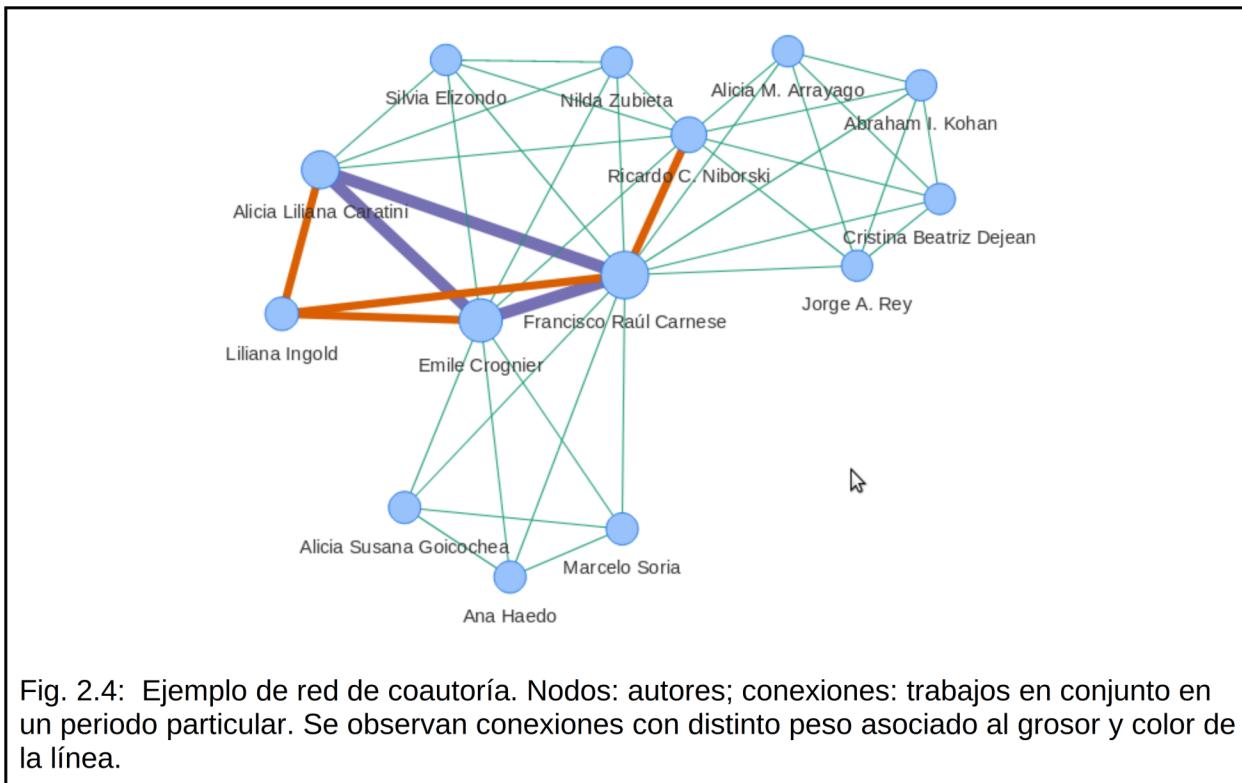
559 Al transformar dicha red en una de coautoría, como menciona [Newman \(2004\)](#): los nodos en la
 560 red representan científicos, y la línea entre ellos indica que fueron coautores de por lo menos
 561 un artículo durante el periodo analizado.

562 Para transformar una red de autoría en una de coautoría se realiza la proyección de la red
 563 sobre los nodos autores, transformando los nodos artículos en conexiones, que representan
 564 cuánto han trabajado en conjunto los autores conectados, y pasando de ser una red dirigida a
 565 una red no dirigida, cuyas conexiones tienen un peso acorde a la cantidad de artículos creados
 566 por ambos autores ([Fig. 2.4](#)).

567

568

569



572 Materiales y métodos

573 Herramientas de trabajo

574 La generación del conjunto de datos fue armada casi en su totalidad en *scripts* en Lenguaje de
575 programación R con múltiples librerías. Para la sección del procesamiento de autores se utilizó
576 la herramienta Openrefine.

577

578 El detalle de herramientas de software utilizadas se destacan en el anexo: [librerías utilizadas](#),
579 donde se puede ver qué librerías se usaron para qué secciones y su correspondiente cita. Los
580 detalles de uso de las herramientas se ven a lo largo de este capítulo.

581 Introducción al proceso

582 Se estudió la estructura del sitio web de la RAAB para identificar la información que pudiera ser
583 de utilidad en el presente estudio. Una vez determinados qué datos se iban a extraer, se
584 armaron los *scripts* para realizar *web-scraping*, luego para el procesamiento de los mismos, y
585 finalmente para el almacenamiento en un formato más conveniente.

586

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

591

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

595

596 Finalmente se transformó el resultado de la base relacional a un formato de grafos, para su
 597 posterior uso en la herramienta de análisis de coautoría facilitando la exploración de los datos.

598

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

600

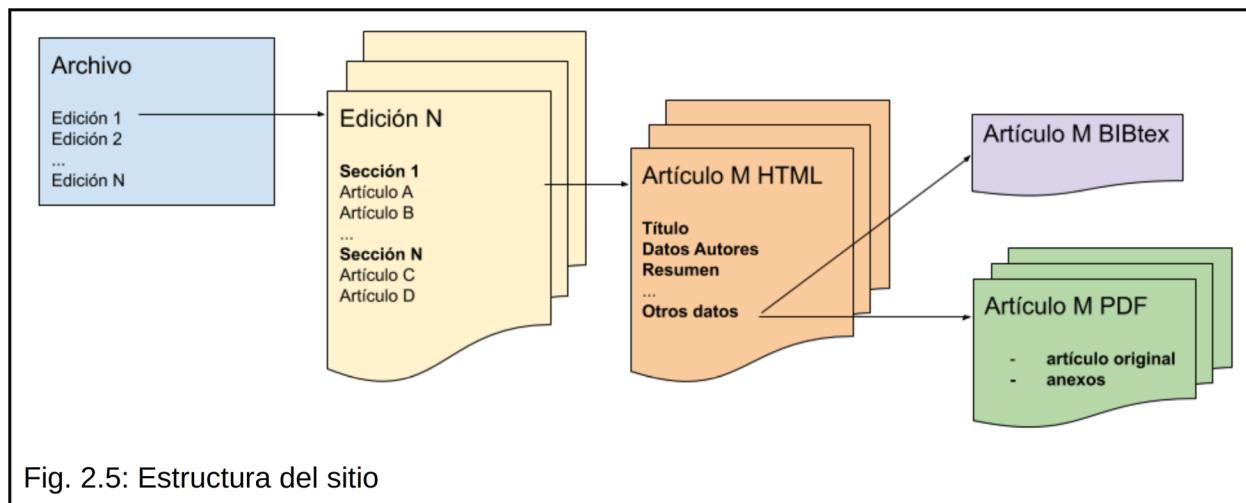
1. Análisis de estructura del sitio (código fuente de las páginas, navegación del sitio)
2. Adquisición de los datos
 - a. Descarga del archivo de ediciones
 - i. Generación de un índice de ediciones (tabla intermedia)
 - b. Descarga de cada edición
 - i. Generación de un índice de artículos (tabla intermedia)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

601 Análisis de la estructura del sitio

602 Se exploró la página de la revista para ver cómo es la estructura. Se identificó que existe una
 603 página “archivo” en la cual se tiene un listado de todas las ediciones de la revista. Para cada
 604 edición de la revista, existe una página en la que se publican datos de la edición y las
 605 secciones con los *links* a los artículos correspondientes, así como sus archivos PDF
 606 relacionados.

607 Para cada artículo existe una página en la que se publican datos del artículo junto con un
 608 resumen del mismo, los PDF asociados y las citas a los artículos con un *link* para la
 609 generación de citas en diferentes formatos ([Fig. 2.5](#)).

610



611 Descripción de los datos de interés

612 Se puede ver en la [Tabla 2.1](#) los datos extraídos, sus orígenes y la razón por la cual resultan de
613 interés.

614

615 Tabla 2.1: 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), <i>Link</i> 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	author title journal

	<p>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.</p>	volume number year Keywords (opcional) Abstract (opcional) issn pages url
Archivos PDF	<p>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.</p>	El texto completo del artículo

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

617 Adquisición de los datos

618 La obtención de los datos se realizó siguiendo el plan de acción detallado en la [Fig. 2.6](#). El
619 detalle de los pasos está en la [Tabla 2.2](#).

620

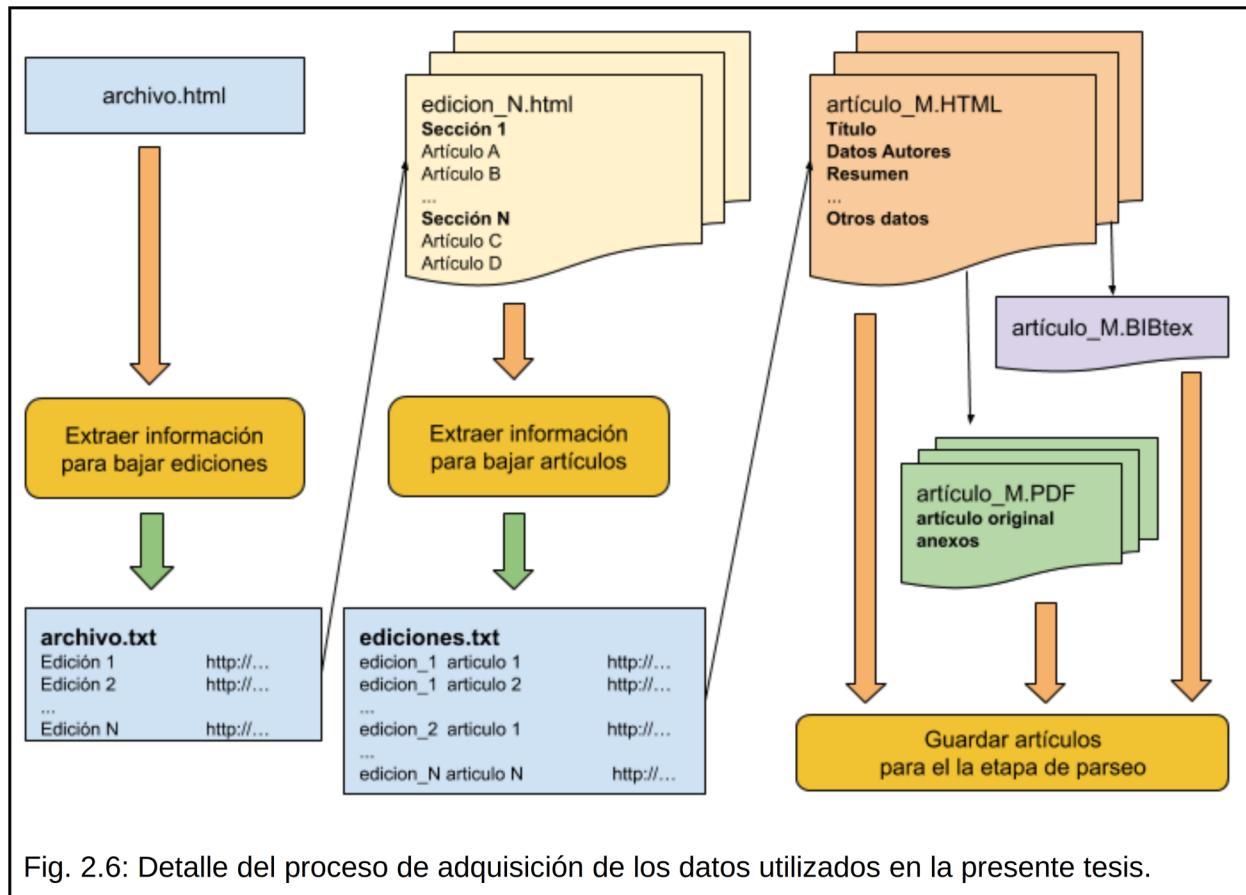


Fig. 2.6: Detalle del proceso de adquisición de los datos utilizados en la presente tesis.

621

622

623 Tabla 2.2: 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

624

625 Parseo de los archivos

626 Se habla de *parseo* 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.

629 Parseo de los archivos HTML

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

632 Parseo de los archivos bibTex

633 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](#).

636 Parseo de los archivos PDF

637 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.

640 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](#)

646 [PDF](#).

647

648 Generación de la Base

649 Resultados intermedios

650 Luego del *parseo*, los datos se guardaron en tablas intermedias en *data.frames* de R y a su
651 vez en archivos intermedios con formato de texto separado por tabulaciones. Generar archivos
652 intermedios forma parte fundamental del proceso a fin de no repetir tareas y ahorrar tiempos de
653 procesamiento en caso de querer ejecutar todo el procedimiento nuevamente.

654 Cada función intermedia fue configurada con parámetros por defecto, que pueden ser
655 sobreescritos en caso de desecharlo.

656 Generación de la base relacional

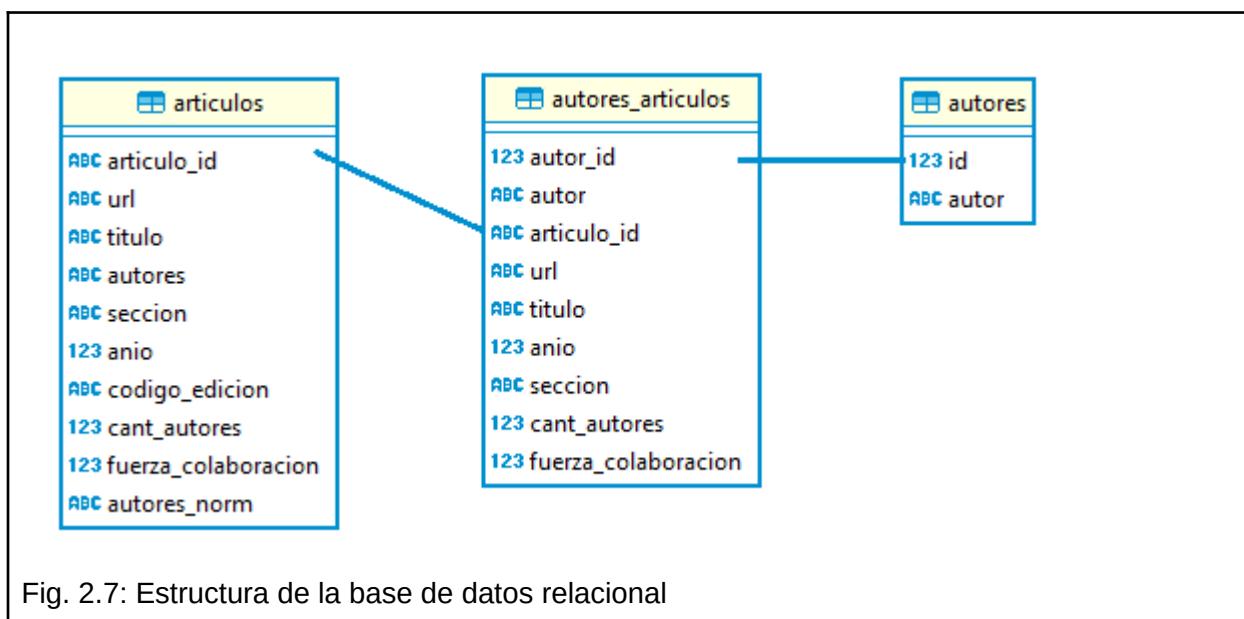
657 El armado de las tablas finales requirió de un extenso proceso de estandarización de las
658 referencias a los autores (nombres y apellidos) y sus afiliaciones institucionales . En el caso de
659 los autores se logró de manera exitosa, mientras que en el de las filiaciones surgieron
660 dificultades, como la falta de datos disponibles en la revista, que imposibilitaron la
661 estandarización y, por lo tanto, el uso de esa información. Para más detalle de este proceso se
662 puede ver el anexo [ANEXO: Detalle estandarización autores](#)

663

664 Una vez finalizado el proceso de estandarización de autores, las tablas finales se guardaron en
665 una base de datos relacional *SQLite*. Se optó por esta herramienta dado que es fácilmente
666 instalable a través de la [librería RSQLite](#) y brinda la posibilidad de tener un formato de base
667 relacional sin la necesidad de instalar programas externos a R. Asimismo, brinda la potencia
668 del almacenamiento como base de datos relacional en un formato conveniente para su
669 reproducibilidad.

670 Estructura de la base final

671 La estructura final fue una N:M, en donde N autores pueden participar en la creación de M
672 artículos. Para representar esta estructura se contó con una tabla de artículos, una tabla de
673 autores, y una tabla intermedia uniendo autores y artículos, en este caso denominada
674 “autores_articulos” ([Fig. 2.7](#)). En la tabla intermedia existieron campos replicados para facilitar
675 las consultas a la base ([Tabla 2.3](#)).
676



677

678

Tabla 2.3: Propiedades de las tablas que conforman la base de datos relacional.

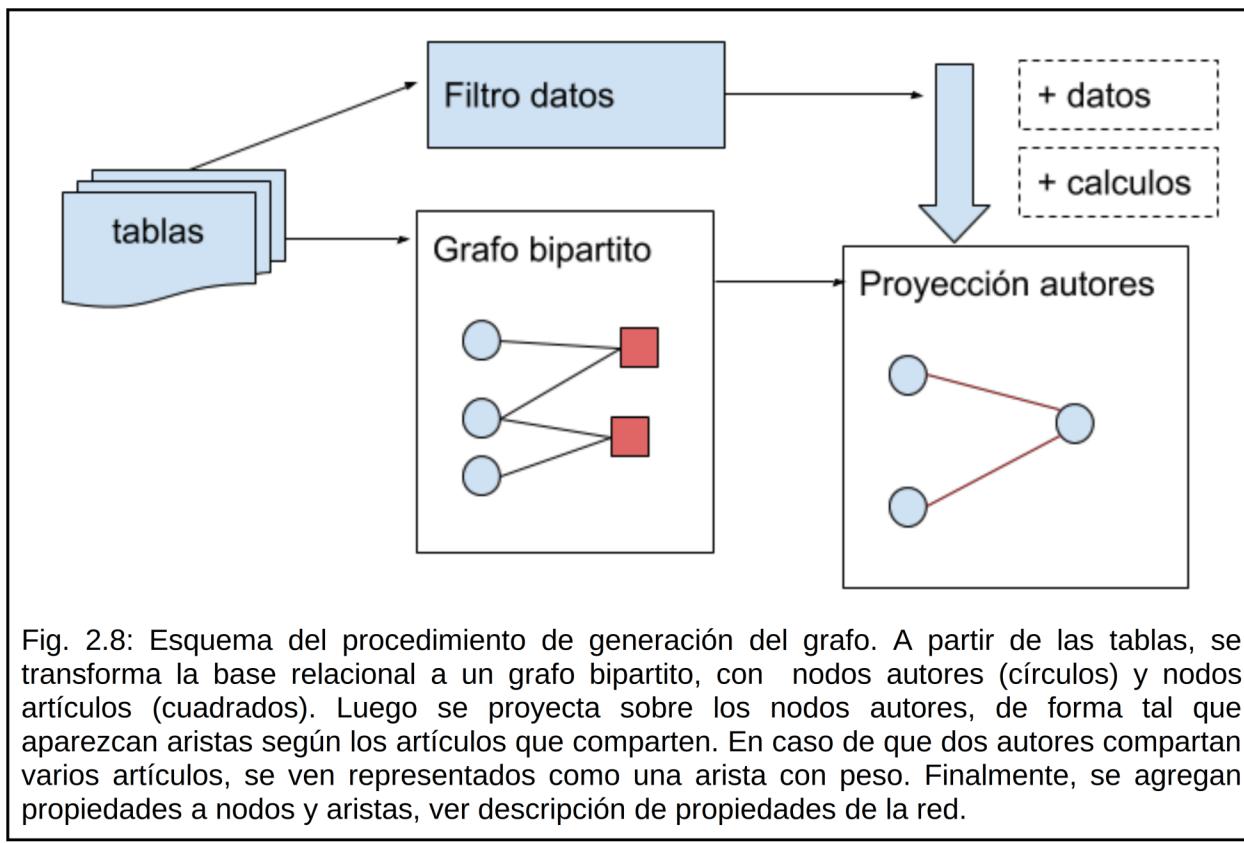
tabla	campo	descripción	Ejemplos / Observaciones
articulos	articulo_id	Identificador de artículo	Ejemplo: 1090, extraído de la página.
articulos	url	Url del artículo	Ejemplo: https://revistas.unlp.edu.ar/raab/article/view/1090
articulos	titulo	Titulo del artículo	Ejemplo: Riesgo de anomalías congénitas 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,	Ejemplo: Hebe Campaña;Jorge

		ordenados en orden alfabetico.	S. López-Camelo;Maria 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

679 Generación de los grafos

680 La transformación de la base de datos relacional a un grafo que facilite su visualización se
 681 realizó de la siguiente manera ([Fig. 2.8](#)):

- 682 1. Recepción de un parámetro de filtro para el grafo (ej.: periodos a analizar -ver [Capítulo 3](#))
- 683 2. Aplicación del filtro a la tabla “autores_articulos” para la preparación de un data.frame
 684 que representa los datos necesarios para crear el grafo.
- 685 3. Utilización de la [librería igraph](#), para la creación de un grafo bipartito de nodos autores
 686 y nodos artículos.
- 687 4. Utilización de la [librería igraph](#) para la realización de la proyección sobre los nodos
 688 autores, pasando a significar “estos autores están juntos por haber realizado un artículo
 689 en conjunto”
- 690 5. Adjudicación de propiedades a los nodos autores y las aristas relaciones de coautoría.
 691



693

694 Descripción de las propiedades de los nodos y aristas

695 En este apartado se detallan las propiedades disponibles para los nodos y las aristas.
 696 Recordemos que los nodos representan a los Autores, y las aristas representan la relación de
 697 coautoría entre los autores que unen. Cada uno de estos, tienen características propias como
 698 en el caso de autores pueden ser el nombre y apellido del autor, y en el caso de las aristas
 699 pueden ser los autores que la generan. En la [Tabla 2.4](#) podemos ver las propiedades de los
 700 autores y en la [Tabla 2.5](#) de las aristas.

701
 702 Debemos recordar que la generación del grafo se realiza luego de aplicar filtros a los datos, por
 703 lo tanto algunos valores son calculados según los filtros aplicados. Para comprender mejor
 704 acerca de los filtros ver Capítulo 3 ([Capítulo 3, Estructura de la Interfaz gráfica, Sección 2](#))

705

706 Propiedades de los Nodos

707 Estas propiedades están disponibles para cada nodo autor ([Tabla 2.4](#)).

708

709 Tabla 2.4: 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]"

711 Propiedades de las Aristas

712 Estas propiedades están disponibles para cada arista que representa la relación de coautoría
713 entre autor 1 y autor 2 ([Tabla 2.5](#)).

714

715 Tabla 2.5: 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. Se genera a partir fuerza_colaboracion en 3 separaciones. (leves , medias , fuertes)	1
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"

716 Resumen del capítulo

717 En este capítulo revisamos las diferentes tareas realizadas para generar una base de datos a
718 fin de poder utilizarla posteriormente en una herramienta de análisis de redes. Dichas tareas
719 comprendieron la extracción desde una página web alojada en Internet, el procesamiento y
720 estructuración de la información para el objetivo específico, así como el armado de
721 documentación que acompañe las tareas realizadas.

722 En el próximo capítulo veremos cómo la información generada es explotada a través de una
723 herramienta que simplifica el acceso y estudio de los datos para transformarlos en información.

724 Capítulo 3: Creación de la Herramienta para análisis

725 Introducción

726 Como parte del Objetivo 1, Desarrollo de una herramienta para análisis de redes de coautoría,
 727 el objetivo de centralizar tareas en una aplicación es disponibilizar tareas rutinarias de análisis
 728 de redes de coautoría a través de una interfaz web para facilitar su accesibilidad.

729 Algunas de las tareas contenidas en la aplicación son:

- 730
- 731 • visualización general de la red desde diferentes puntos de vista,
 - 732 • selección de diferentes nodos/aristas/grupos,
 - 733 • detalles de nodos y aristas, detalles de los datos, entre otras,
 - 734 • filtros en los datos,
 - 735 • variación de parámetros para filtros y/o algoritmos.

736 Se consideró clave facilitar la exploración de los datos mediante interacciones intuitivas
 737 brindadas a través de una interfaz web para facilitar el acceso a diferentes personas que
 738 realicen los análisis.

740

741 En este capítulo se abordarán temas de metodología del trabajo, estructura de la aplicación, y
 742 funcionalidades disponibles.

743 Materiales y métodos

744 Estructura Interna de la aplicación

745 Para disponibilizar la herramienta a más de una persona, se decidió que la mejor forma sería
 746 mediante una aplicación web. Se entiende por aplicación web a una herramienta de software
 747 informático que se encuentra alojada en un servidor, al cual se le hacen consultas a través de
 748 clientes generalmente navegadores de Internet, lo cual facilita su acceso por analistas o
 749 investigadores desde diferentes lugares en el mismo momento sin necesidad de instalar nuevo
 750 software más allá del navegador.

751 Herramientas de trabajo

752 La aplicación fue creada enteramente en Lenguaje R con múltiples librerías, que se destacan
 753 en el anexo: [librerías utilizadas](#), donde se puede ver cuáles librerías se usaron para qué
 754 secciones y su correspondiente cita. Entre ellas se destaca *Shiny*, que permite crear una
 755 aplicación web en el entorno de R, facilitando la reproducibilidad al reducir las herramientas
 756 utilizadas dado que no se utilizan otros lenguajes para poder crear la aplicación.

757

758 Para el uso de las herramientas de trabajo, y buenas prácticas, se usaron recomendaciones
 759 destacadas en [Wilson, et al. \(2017\)](#).

760

761 Para crear la aplicación se utilizó principalmente el entorno de desarrollo RStudio, dado que
 762 facilita considerablemente el desarrollo de aplicaciones y scripts con el lenguaje R.

763

764 Se armó la aplicación en forma de proyecto como un Rproject, porque ayuda a gestionar los

765 archivos de código. Esta forma de trabajo resultó útil principalmente para gestionar referencias
 766 entre archivos y facilitar el uso de algunas librerías.
 767

768 Para el versionado del código fuente se utilizó GIT. Si bien no se utilizó desde el inicio, fue
 769 necesario a partir de un momento en que el código fue demasiado extenso y se hizo
 770 inmanejable el versionado, al igual que, al cometer errores, volver para atrás podía llegar a ser
 771 prohibitivo. Para entender un poco más de que se trata GIT debemos entender mejor qué es
 772 un Sistema de control de versiones (*Version control system, VCS*), en palabras de [Blischak, et al. \(2016\)](#)
 773 traducidas por el autor de esta tesis:
 774

Un sistema de control de versión permite realizar el seguimiento de cambios iterativos que vas armando en tu código. Por lo tanto, podés experimentar con nuevas ideas y siempre tener la opción de volver atrás a una versión específica del código que usaste para generar resultados particulares.

775 Utilizar un sistema de versiones da otras ventajas además de tener versiones del código, como
 776 poseer un orden en la forma de trabajar y, en caso de compartir remotamente, tener *backups*
 777 en otras máquinas, lo cual reduce el riesgo de pérdida del trabajo realizado.
 778

779 Respecto de la estructura de la aplicación

780 La estructura del proyecto fue cambiando en el tiempo, desde ser un simple *script* pasando a
 781 varios *scripts* separando tareas, hasta consolidarse en un proyecto para la aplicación
 782 propiamente dicha. Luego se agregaron los conceptos de *Shiny modules* en base a que la
 783 aplicación creció considerablemente. La versión final de la aplicación tiene la estructura
 784 siguiente:
 785

- proyecto
 - data (Para los datos a utilizar)
 - db_raab_grafos.sqlite
 - temporal_acumulado_data.Rdata
 - temporal_acumulado_dyn_data.Rdata
 - temporal_basico_data.Rdata
 - funcs (funciones y scripts de la aplicación)
 - imports.r (script que tiene todas las librerías a utilizar)
 - globals.r (script que tiene variables globales de la aplicación)
 - funciones.r (script que tiene funciones de la aplicación)
 - ui.r (Script que tiene la Interfaz de usuario)
 - server.r (Script que tiene el lado servidor de la aplicación)
 - shiny_module_estructura_modelos.r
 - (Módulo de Shiny: tiene UI y Server para la sección de modelos)
 - shiny_module_estructura_nodos.r
 - (Módulo de Shiny: tiene UI y Server para la sección de estructura de nodos, se recicla en varios lados de la aplicación)
 - shiny_module_exploratorio_tablas.r
 - (Módulo de Shiny: tiene UI y server para la parte del análisis exploratorio de la aplicación)
 - shiny_module_subgrafos.r
 - (Módulo de Shiny: tiene UI y server para la parte de análisis de subgrafos , se utiliza para la sección de componentes, se podría reciclar en comunidades.)
 - shiny_module_comunidades.r
 - (Módulo de Shiny: tiene ui y server utilizados en la parte de Análisis de comunidades)
 - shiny_module_articulos_text_mining.r

- (Módulo de Shiny: tiene ui y server utilizados en la parte de análisis de comunidades y de módulo subgrafos)
- test (pruebas)
 - múltiples archivos de pruebas de funciones
 - Que devuelvan lo esperado
 - múltiples archivos de prueba de concepto (probar si era viable realizar las funciones / aplicaciones para incluir en la app)
 - Pruebas de concepto y generación de archivos, para probar diversas librerías antes de agregar a la aplicación en sí.
- www (css para estilos personalizados en algunos lugares de la aplicación)
- app.r
 - Es el archivo principal de la aplicación cuando se arma un proyecto Shiny; en este caso en particular se hacen las invocaciones a todos los otros scripts que forman la aplicación y finalmente creación de la app Shiny.

786

787 Disponibilidad local, en la red, en internet

788 Inicialmente se trabajó de forma local, esto significa en la misma computadora donde se crean
 789 los scripts, esta forma de trabajo es necesaria para pruebas, sin embargo no es suficiente en
 790 caso de querer disponibilizar para otras personas.

791 Cómo revisa [Wojciechowski, et al. \(2015\)](#), existen varias formas de disponibilizar las
 792 aplicaciones desarrolladas con el paquete *Shiny*:

Las aplicaciones Shiny pueden ser compartidas fácilmente compartiendo los scripts ui.R y server.R con un colega que tenga instalado R y la librería Shiny. Sin embargo, no está limitado a esta forma. Las aplicaciones Shiny pueden ser disponibilizadas en servidores web haciéndolas accesibles para aquellos que no están familiarizados con, o no tienen instalado, R.

Hay 3 formas ofrecidas por RStudio para disponibilizar una aplicación web Shiny: shinyapps.io, Shiny Server, y Shiny server Pro. Los usuarios van a ver estas opciones como más lentas que correrlos en sus propias máquinas, predominantemente dependientes en la velocidad y latencia de la conexión a internet. Cuando una aplicación Shiny se está actualizando se verán los output de la aplicación grisados, indicando qué procesamiento se está llevando a cabo.

793

794 Otras consideraciones respecto de los servicios mencionados, como shinyapps.io, es que
 795 sirven para hacer despliegues de aplicaciones *Shiny*. El servicio es pago, sin embargo brinda
 796 una categoría gratis que da cierta cantidad de horas de uso por mes.

797

798 Por este tema se consideró la opción de montar un servidor propio de *Shiny*; como plan de
 799 backup en caso de quedarse sin crédito y no poder pagar una categoría más alta.

800

801 Las 3 opciones implican cierto nivel de aprendizaje, requisitos para poder lograr el objetivo de
 802 disponibilizar las aplicaciones. Para ver más detalles acerca de sus ventajas y desventajas
 803 referirse al artículo de [Wojciechowski, et al. \(2015\)](#).

804 Estructura de la interfaz gráfica

805 La aplicación está dividida en 3 secciones principales

806

807 Sección 1: **consideraciones**

808 Sección 2: **análisis estático**

809 Sección 3: **análisis temporal**

810

811 Para poblar las pantallas de la herramienta de análisis de red de coautoría, se utilizaron las
 812 tablas finales de la base de datos relacional. El proceso de generación de grafos se realiza al
 813 aplicar ciertos filtros en la aplicación, utilizando los datos de las tablas finales. Además, se
 814 filtran estos datos para poder acceder al resultado de forma más eficiente según los criterios
 815 escogidos.

816 Para lograr armar el esqueleto de la interfaz gráfica se utilizaron las librerías [Shiny](#),
 817 [shinyDashboard](#) y [shinyDashboardPlus](#).

818 Sección 1: Consideraciones

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

822

823 Las consideraciones a tener en cuenta a la hora de usar la aplicación corresponden
 824 específicamente a que se tomaron los trabajos originales publicados por la RAAB durante el
 825 periodo entre 1996 y 2016. [Fig. 3.1](#).

826

Fig. 3.1: Consideraciones Generales.

827

828 En la sección de **análisis exploratorio** se encuentra un “checkbox” ([Fig. 3.2](#) y [Fig. 3.3](#)) donde
 829 es posible seleccionar si se desea ver los datos disponibles considerando o no este
 830 acotamiento temporal y con respecto al tipo de trabajo propio de la presente tesis. Esta opción
 831 fue agregada para que quien realice un análisis utilizando la herramienta sepa que el conjunto
 832 de datos completo de la revista contiene más información. La utilidad principal se ve a la hora
 833 de evaluar la cantidad de artículos en otras secciones (este punto en particular se abordará
 834 más adelante).

835

Fig. 3.2: Análisis Exploratorio acotado al objetivo de la presente tesis

Fig. 3.3: Análisis Exploratorio expandido a la totalidad de la base de datos obtenida (es decir, hasta marzo 2018 y abarcando todas las secciones)

836

837 El análisis exploratorio que se puede ver en la [Fig. 3.4](#) se focaliza en el acceso a los datos
 838 crudos, la composición de los datos, un resumen estadístico y algunas preguntas básicas de
 839 interés. Analizaremos estos valores en el próximo capítulo, por el momento solo se muestran
 840 imágenes para familiarizarse con la estructura de la interfaz gráfica.

841 Datos Crudos

842 Refiere a la visualización de los datos en forma de tabla interactiva en la cual se pueden aplicar
 843 distintos filtros. Para lograr esta utilidad se utilizó la [librería DT](#) (por *data tables*). La herramienta
 844 permite bajar los datos, buscar por palabras que figuren en cualquier columna, entre otras.
 845

Datos Crudos										
Composición Resumen Preguntas										
Bajar datos Buscar: paleoparasitología										
articulo_id	url	título	autores	sección	año	codigo_edicion	cant_autores	fuerza_colaboracion	autores_norm	
239	https://revistas.unlp.edu.ar/raab/article/view/239	Nuevas fuentes de evidencia para la paleoparasitología y la antropología biológica en Patagonia: resultados preliminares	Martin H. Fugassa;Armando Cicchino;Norma H. Sandella;Ricardo A. Guichón;Guillermo M. Denegri;Adauto Araújo	Trabajos Originales	2007	2007_9_2	6	0.1666666666666667	Ad auto Araújo;Armando Cicchino;Guillermo M. Denegri;Martin H. Fugassa;Norma H. Sandella;Ricardo A. Guichón	
281	https://revistas.unlp.edu.ar/raab/article/view/281	Nuevos aportes a la paleoparasitología del sitio arqueológico (Nombre de Jesús) (s.XVI), Cabo Vírgenes, Argentina	Martin H Fugassa;Ricardo A. Guichón	Trabajos Originales	2006	2006_8_1	2	0.5	Martin H. Fugassa;Ricardo A. Guichón	

Fig. 3.4: Visualización de la sección “Datos crudos”. A modo de ejemplo, se muestra la aplicación de un filtro por palabra.

846

847 Composición

848 Esta opción muestra un análisis visual de cómo es la composición de las tablas, en cuanto a
 849 tipos de datos y cantidad de observaciones. Para lograr esta visualización se utilizó la librería
 850 [visdat](#) (Fig. 3.5).

851



Fig. 3.5: Vista de la composición de los tipos de datos por columnas de una tabla específica.

853 Resumen

854 Permite ver un resumen estadístico de la distribución de los datos en la tabla. Para lograr esta
 855 funcionalidad se utilizó la librería [skimr](#). Mediante esta opción se pueden ver las estadísticas
 856 que se utilizan para cada tipo de dato (Fig. 3.6).

857

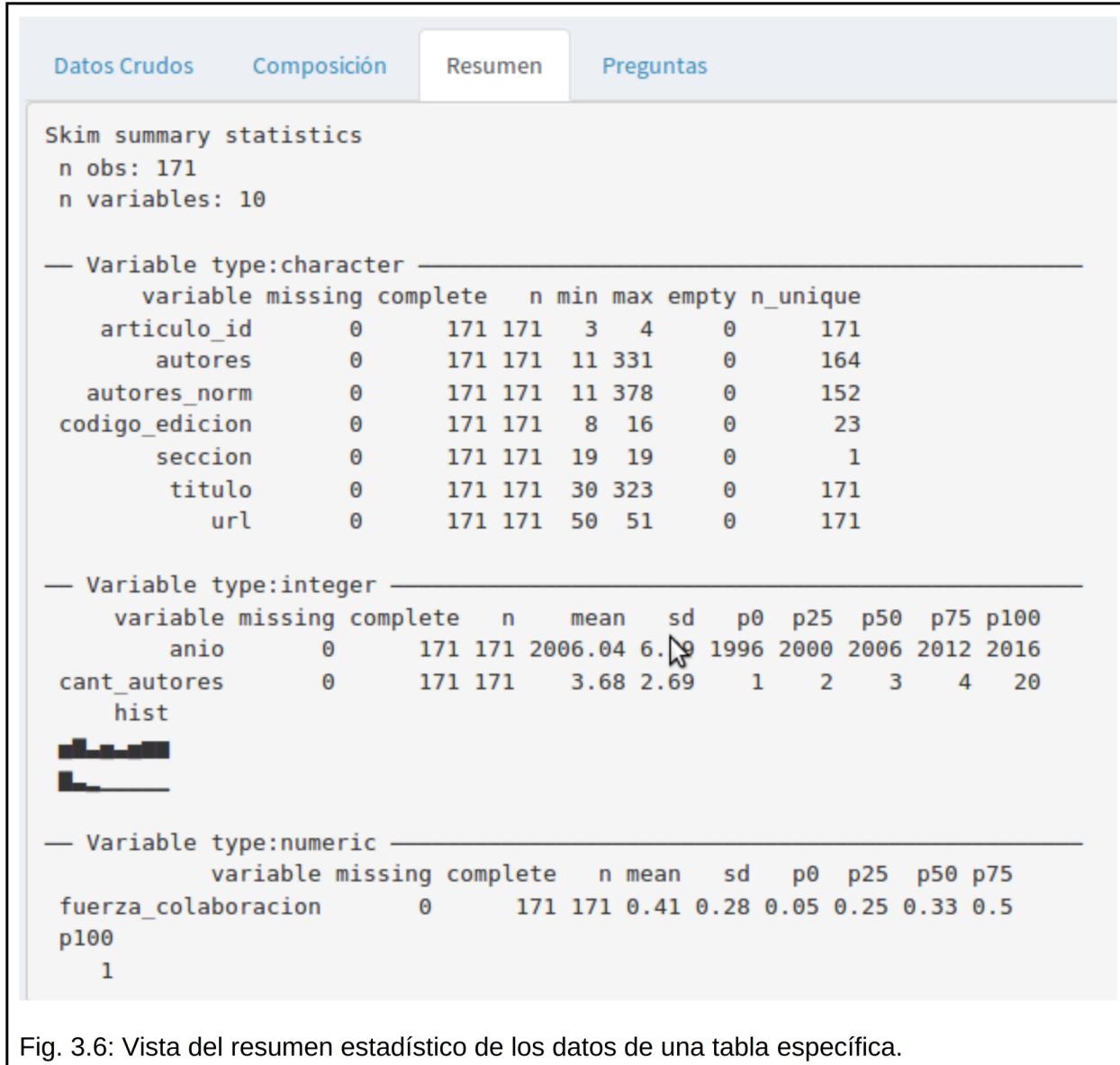


Fig. 3.6: Vista del resumen estadístico de los datos de una tabla específica.

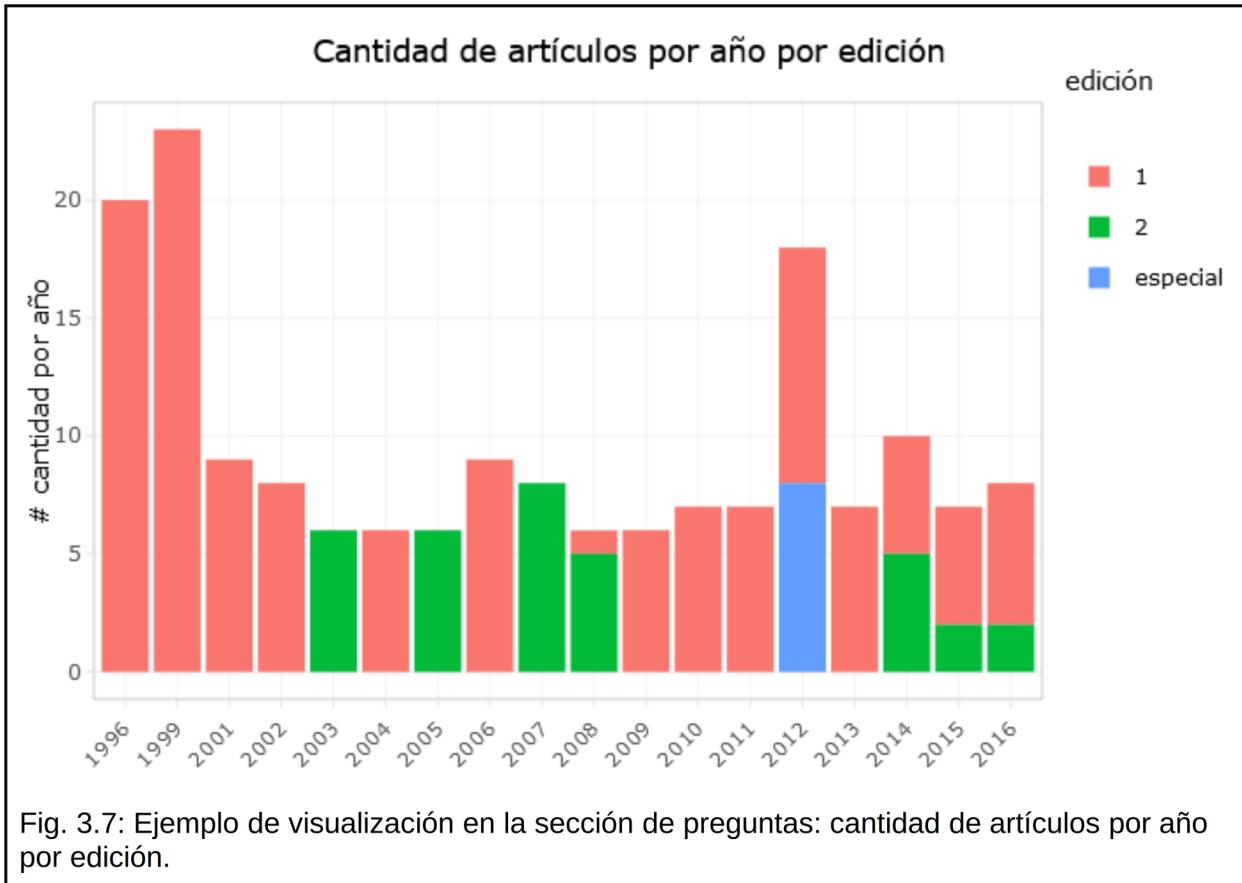
858

859

860 Preguntas

861 Este ítem refiere a preguntas básicas que salieron al ver los datos de cada tabla. Estas
 862 preguntas son utilizadas a lo largo de la tesis, dado que se usan para familiarizarse con los
 863 datos. Para poder hacer los gráficos interactivos en la sección se utilizó la librería plotly ([Fig.
 864 3.7](#)).

865



866

867

868 Las preguntas que se hicieron en este análisis están en la [Tabla 3.1](#).

869

870 Tabla 3.1: 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 +)

871

872

873 Con esto concluimos la sección 1 de **consideraciones**, el objetivo de esta sección es
874 relacionarse inicialmente con los datos, ver cómo están estructurados, y una primera
875 aproximación a cuáles son preguntas que se pueden contestar.

876

877 Luego de estar familiarizados con los datos se puede proceder tanto a la sección 2 como la 3
878 dependiendo de si se desea estudiar uno o varios períodos en particular, o analizar la
879 evolución temporal de los datos, respectivamente.

880 Un ejemplo es explorar la evolución temporal, para luego focalizarse en algún periodo
881 particular.

882

883 Sección 2: Análisis Estático

884 El objetivo de esta sección es poder configurar, filtrar, visualizar, agrupar y estudiar métricas de
 885 la red en detalle. Para realizar estas tareas esta sección se divide en varias subsecciones:
 886 Visualización de coautoría, Artículos asociados, Estructura de la red, Comparación con
 887 modelos y Comunidades. En la [Fig. 3.8](#) podemos ver la pantalla de Visualización de coautoría.
 888

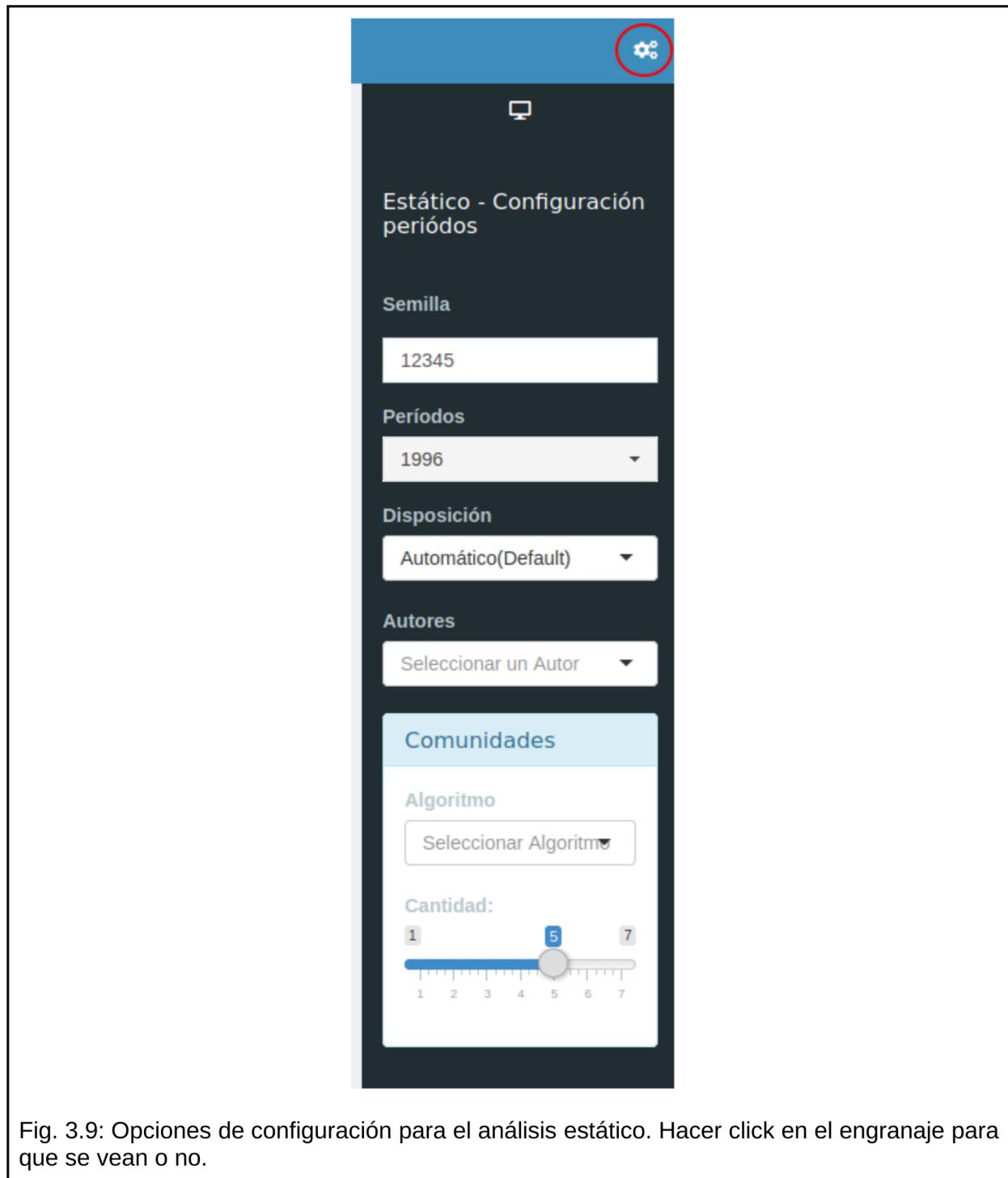


Fig. 3.8: Ejemplo de resultado obtenido en la sección de Análisis estático > Visualización de coautoría > Red.

889 Configuración

890 Para configurar el análisis a realizar, se debe hacer click sobre el engranaje como se ve en la
 891 [Fig. 3.9](#). En esta sección se configura la semilla (seed) para asegurar la reproducibilidad en
 892 partes aleatorias de las ejecuciones, como podrían ser algunas de las “disposiciones” de la red
 893 (*layouts*) o las ejecuciones de algunos algoritmos, por ejemplo, en la parte de Comunidades,
 894 las “caminatas aleatorias cortas” (*short random walks*).

895 Luego, la opción periodos permite seleccionar los períodos a analizar. Al hacerlo, se
 896 desprenden opciones de la red resultante en particular, como ser la disposición de la
 897 visualización de Red, si se desea seleccionar algún autor en particular, y la opción de
 898 búsqueda de comunidades en la red actual.
 899



901 Visualización

902 Utilizando la aplicación es posible obtener diferentes tipos de visualización del grafo: la vista de
 903 red, Biofabric y Mapa de calor (*heatmap*). Esta sección fue armada con controles de la [librería](#)
 904 [shiny](#).

905

906 La visualización en forma de red ([Fig. 3.10](#)) muestra a los Autores como nodos y su relación de
 907 coautoría en forma de vértices. El tamaño de los nodos representa la suma total de fuerza de
 908 coautoría del autor. El color de las aristas está basado en la suma de fuerza de colaboración de
 909 la relación y pueden seleccionarse tres categorías relativas al grafo actual (que representan
 910 baja, media o alta relación de coautoría) para filtrar la visualización de aristas. Esta
 911 visualización fue realizada con la librería *visnetwork*.

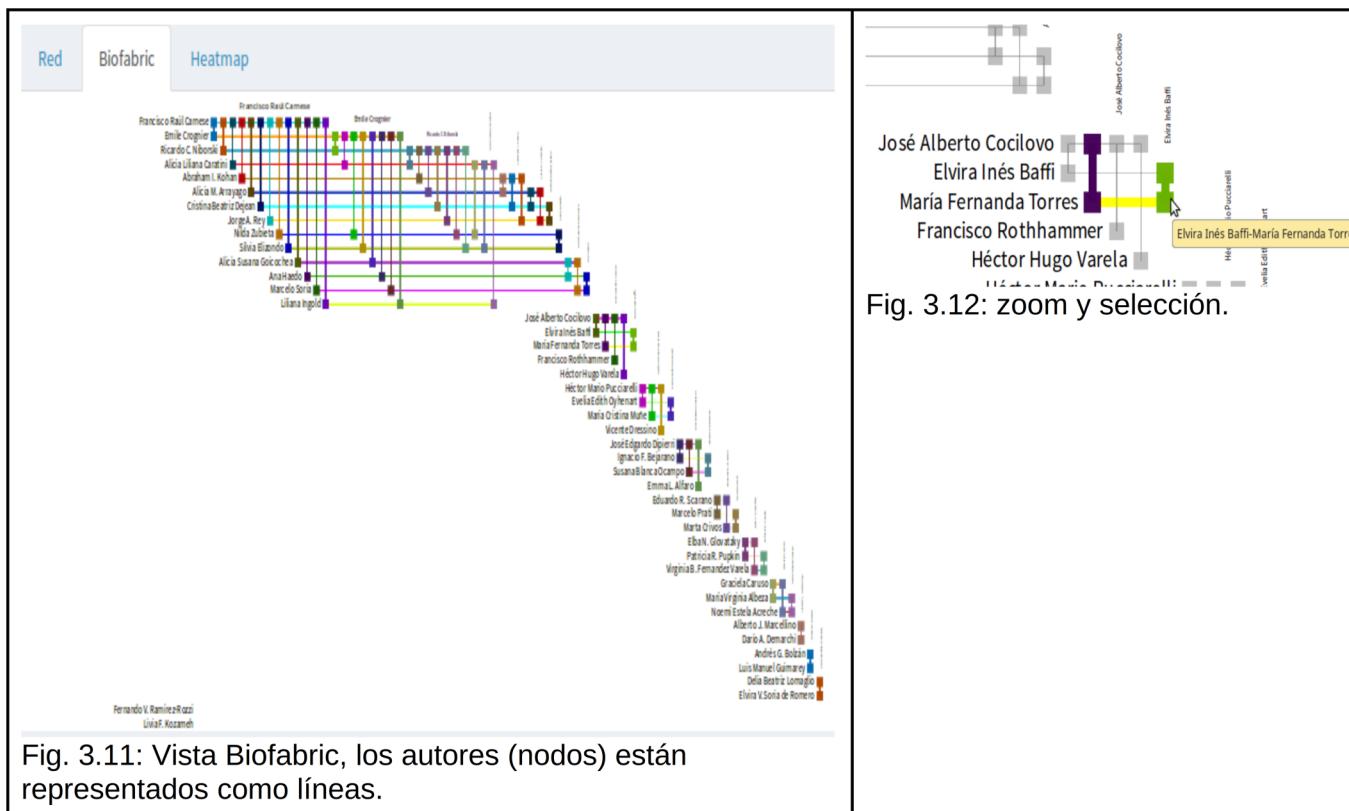
912



Fig. 3.10: La vista de red en la cual se puede filtrar e interactuar para afectar a la sección de artículos asociados. El filtro fue aplicado en este caso para mostrar las relaciones que tengan una fuerza de colaboración mayor o igual a 3.

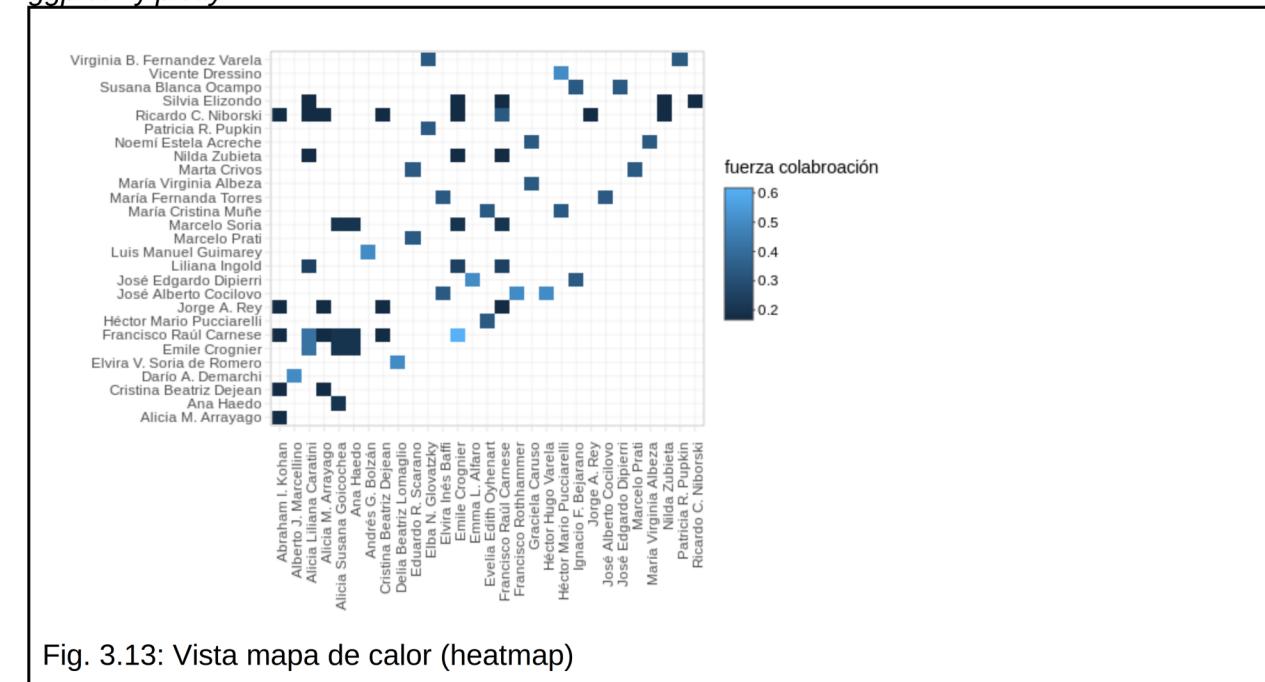
913

914 La Visualización Biofabric ([Longabaugh 2012](#)), es otra opción interactiva para ver las relaciones
 915 entre autores ([Fig. 3.11](#) y [Fig. 3.12](#)). En esta se toman los nodos como líneas evitando, de esta
 916 forma, el problema de tener muchos vértices superponiéndose en forma desordenada. Esta
 917 sección fue armada con la adaptación de R disponible en la librería *RBioFabric*.



918

919 Finalmente, la visualización de tipo mapa de calor (*heatmap*) (Fig. 3.13) permite observar la
 920 fuerza de colaboración entre pares de autores. Esta adaptación fue realizada con las librerías
 921 *ggplot2* y *plotly*.



922

923 Artículos asociados

924 Esta sección está asociada a la visualización en formato de red. Al seleccionar un nodo Autor,
 925 se muestran los datos asociados a dicho Autor incluyendo las visualizaciones de subgrafo y
 926 heatmap. En el resumen se muestra quién es el Autor seleccionado, en qué períodos participó,
 927 la cantidad de artículos relacionados con dicho autor, la cantidad de autores asociados y la
 928 fuerza colaboración total para dicho autor (Fig. 3.14). Es importante considerar que cualquier
 929 métrica o valor tendrá aplicado el filtro correspondiente a los períodos seleccionados en la
 930 configuración de la sección de análisis estático. La realización de esta visualización está
 931 basada en el artículo de [Newman \(2004\)](#).
 932

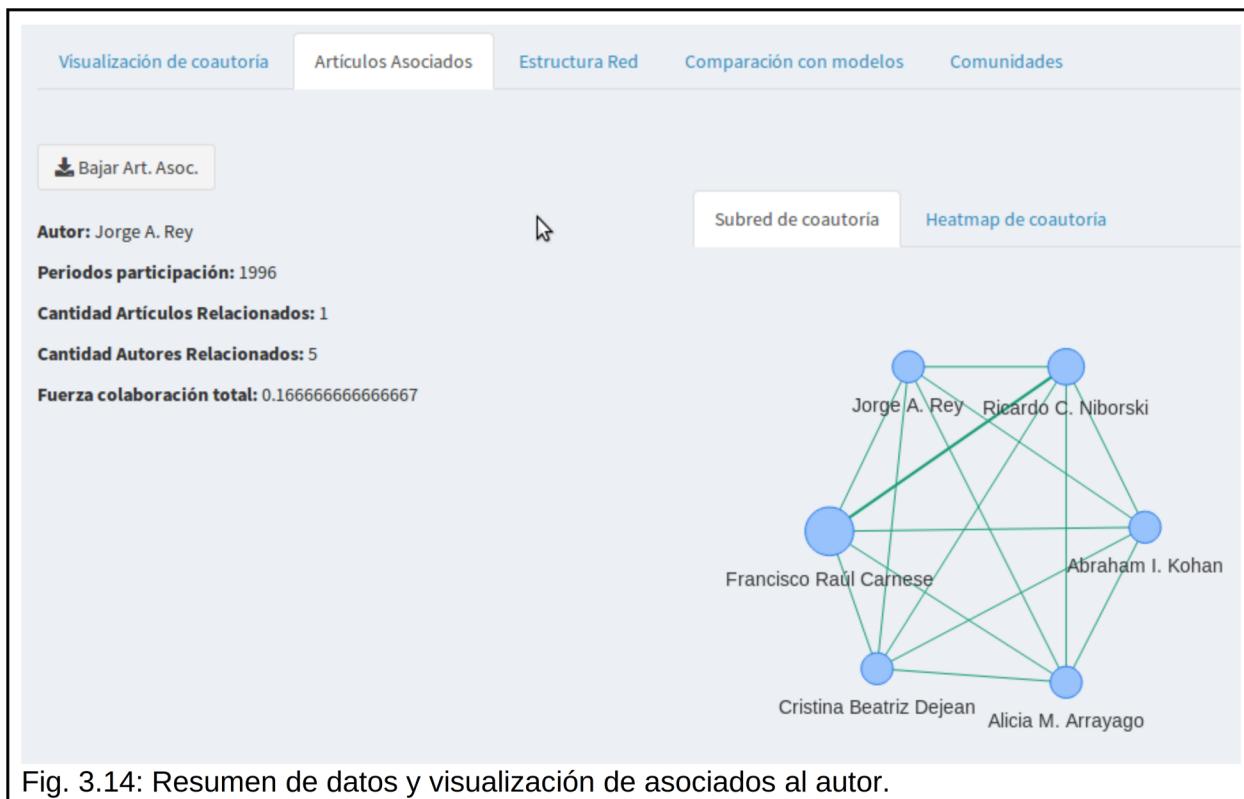


Fig. 3.14: Resumen de datos y visualización de asociados al autor.

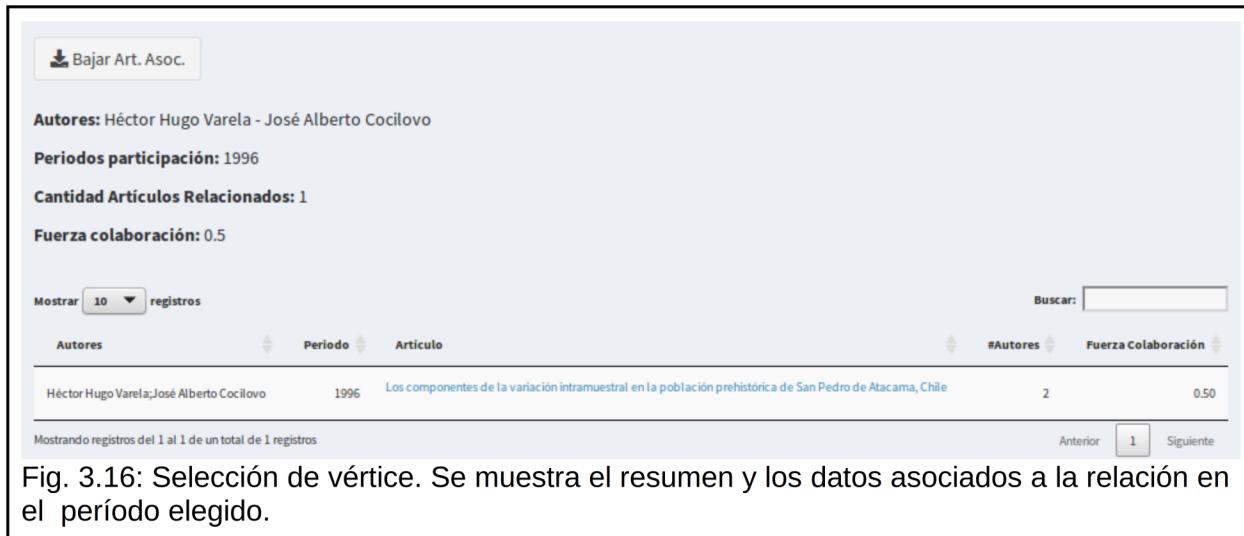
933
934
935
936
937

Fig. 3.15: Selección de autor. Se muestran los datos asociados al autor en el período seleccionado.

938
939

Por otro lado, al seleccionar un vértice se pueden ver los autores involucrados en la relación,

940 los períodos de participación de esa relación, la cantidad de artículos relacionados y finalmente
 941 la fuerza colaboración asociada a esa relación ([Fig. 3.16](#)).
 942



The screenshot shows a software interface for managing research collaborations. At the top, there is a button labeled "Bajar Art. Asoc." (Download Associated Articles). Below it, the following information is displayed:

- Autores:** Héctor Hugo Varela - José Alberto Cociollo
- Periodos participación:** 1996
- Cantidad Artículos Relacionados:** 1
- Fuerza colaboración:** 0.5

Below this summary, there is a search bar labeled "Buscar:" and a table with the following columns: Autores, Periodo, Artículo, #Autores, and Fuerza Colaboración.

Autores	Periodo	Artículo	#Autores	Fuerza Colaboración
Héctor Hugo Varela;José Alberto Cociollo	1996	Los componentes de la variación intramuestral en la población prehistórica de San Pedro de Atacama, Chile	2	0.50

At the bottom of the interface, it says "Mostrando registros del 1 al 1 de un total de 1 registros" and has navigation buttons for "Anterior" and "Siguiente".

Fig. 3.16: Selección de vértice. Se muestra el resumen y los datos asociados a la relación en el período elegido.

943 Estructura de la red

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

947

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

951 *Métricas de la red*

952 Esta sección dispone de diferentes métricas que se utilizan habitualmente en análisis de redes,
 953 que fueron seleccionadas en base a las propuestas por [Newman \(2004\)](#). Para este caso bajo
 954 estudio, se quitó el Promedio de colaboradores, dado que en la revista no se cuenta con
 955 información explícita para una correcta distinción entre el autor principal y los colaboradores
 956 sino que solamente figura el orden en la página de internet. Además se agregaron otras
 957 métricas como: el porcentaje que representa el cliqué más grande, la densidad de la red, la
 958 cantidad de autores en el componente más grande, la cantidad de componentes, la cantidad de
 959 contribuciones y la cantidad de relaciones ([Tabla 3.2](#)).
 960

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: $n_{autores} * (n_{autores}-1) / 2$
# 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 $n_{autores_clique_maximo} / n_{autores} * 100$, 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.

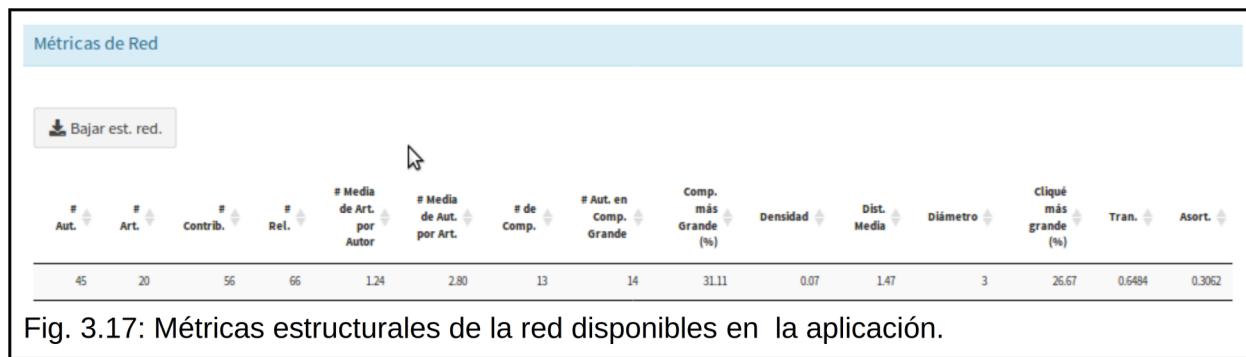


Fig. 3.17: Métricas estructurales de la red disponibles en la aplicación.

967 *Métricas de los nodos*

968 Esta sección incluye distintas “solapas” donde se puede acceder a los datos (es decir, las
969 métricas propiamente dichas), la distribución de grado, el histograma de variables, y la
970 densidad de variables. Para una revisión de las medidas dentro de las redes se puede
971 consultar a [Newman \(2016\)](#) y a [Freeman \(1978\)](#). Las métricas que se incluyeron se describen
972 en la [Tabla 3.3](#).

973

974 Tabla 3.3. 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.
Page Rank	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.

975

976 En la [Fig. 3.18](#) se muestra la disposición en la aplicación de la sección Métricas de Nodos.
977 Para la sección de datos está disponible el cuadro de búsqueda que agiliza la búsqueda de
978 algún autor en particular.

Métricas de Nodos							
Datos	Distribución de grado	Histograma variables	Densidades variables				
 Bajar est. nodos							
Mostrar	10	registros		Buscar: <input type="text"/>			
Autor	Fuerza Colaboración	Grado	Intermediación	Centralidad Autovector	Cercanía	Page Rank	# Triángulos
Abraham L. Kohan	0.1667	5	0.1429	0.3090	0.0007062	0.0188	10
Alberto J. Marcellino	0.5000	1	0.0000	0.0000	0.0005165	0.0236	0
Alicia Liliana Caratini	0.4167	6	1.1667	0.6537	0.0007042	0.0274	12

Fig. 3.18: Métricas de los nodos, sección Datos.

981 La distribución de grado es un gráfico ampliamente utilizado en el estudio de redes. Analizando
 982 la distribución se puede ver qué tipo de nodos predomina más dentro de una red en particular
 983 ([Fig. 3.19](#)).

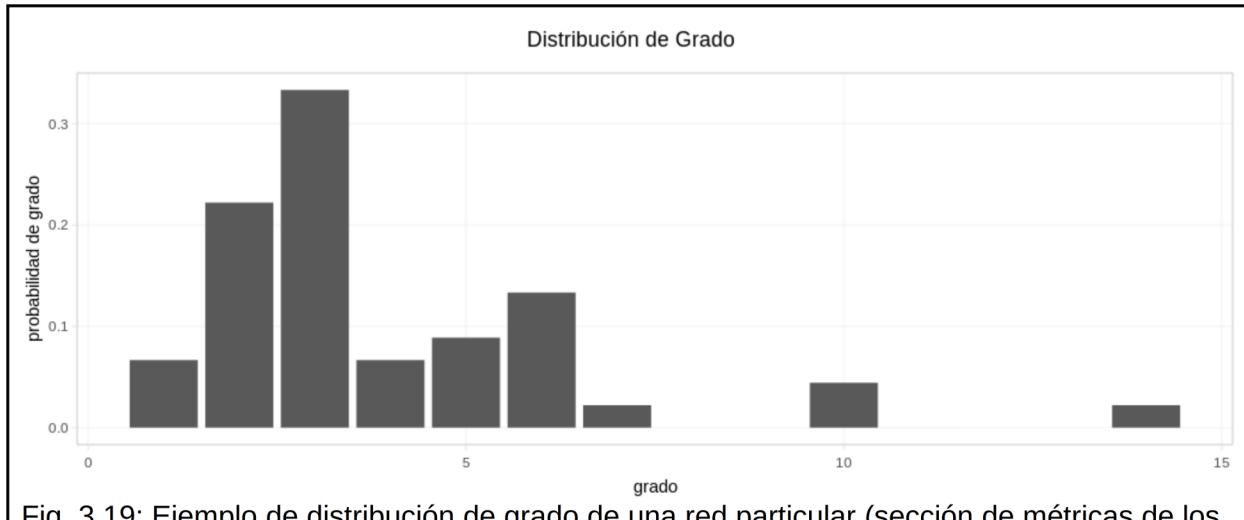


Fig. 3.19: Ejemplo de distribución de grado de una red particular (sección de métricas de los nodos).

986 En la sección Histograma variable ([Fig. 3.20](#)) podemos obtener los histogramas de las métricas
 987 que se mencionaron anteriormente para poder visualizar de forma gráfica los datos.
 988

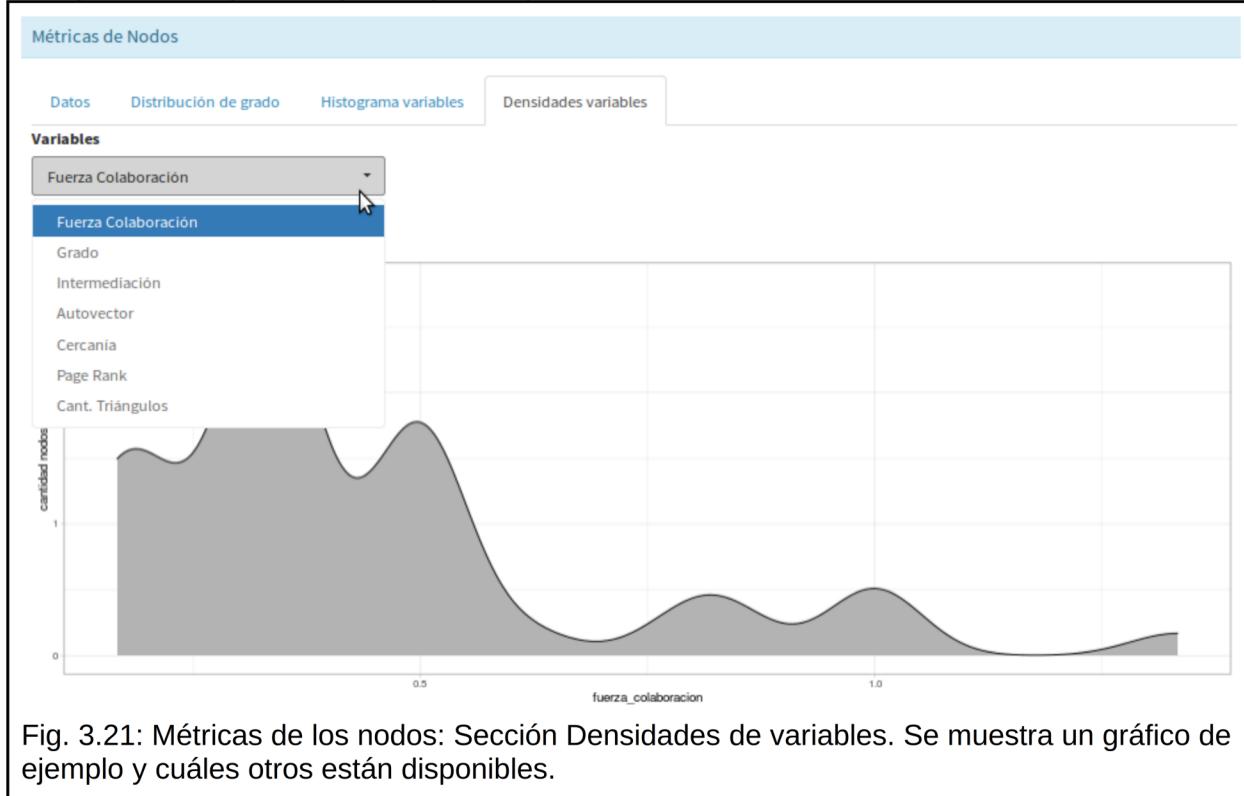


989

990

991

Estas métricas son muchas veces representadas en forma de gráfico de densidades ([Fig. 3.21](#)) en lugar de histogramas, por lo que la aplicación brinda ambas funcionalidades.



992 *Detalles de los componentes*

993 Un componente es una porción de red en donde todos los nodos están conectados directa o
994 indirectamente por lo menos por una arista ([Morel et al. 2009](#)).

995

996 El análisis de componentes da un vistazo general de la estructura de la red ([Fig. 3.22](#)) y provee
997 información valiosa en relación a su nivel de fragmentación ([Morel et al. 2009](#)).

998

999 Para extraer los componentes se utilizó la [librería igraph](#).

1000

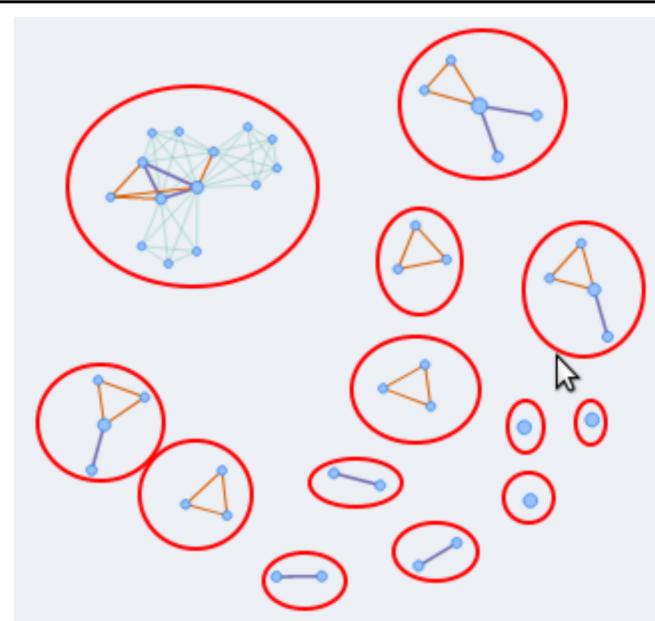


Fig. 3.22: Ejemplo de gráfico de componentes obtenido para la red de la RAAB correspondiente al año 1996. Los componentes se resaltan en rojo.

1001

1002 En esta sección pueden obtenerse las métricas de grafo para cada componente (Estructura
1003 subgrafo), los autores pertenecientes a cada componente, y detalles del componente
1004 seleccionado y sus nodos asociados (Detalle subgrafo seleccionado y Detalle nodos del
1005 subgrafo seleccionado). En este último caso puede verse qué artículos están relacionados a
1006 dicho componente, así como distintas métricas de los nodos correspondientes .

1007 Las métricas que pueden obtenerse para cada componente ([Fig. 3.23](#)) son las mismas que las
1008 que fueron utilizadas para la red total ([Tabla 3.2](#)).
1009

Estructura Subgrafos

Bajar métricas

Mostrar 10 registros Buscar:

componente_id	cant_autores	cant_relaciones	densidad_red	distancia_media	num_dist_lejanos	num_largest_cliques	porc_largest_cliques	num_cliques	num_transitivity	num assort_degree
1	14	41	0.49024940354601	1.5464205464025	3	12	185.7142857142857	6	0.640340463414034	-0.309002443890024
7	3	3	0.5	1.5	2	3	60	3	0.375	-0.8222222222222226
8	4	4	0.0000000000000007	1.333333333333333	2	3	73	3	0.6	-0.714285714285714
9	4	4	0.0000000000000007	1.333333333333333	2	3	73	3	0.6	-0.714285714285714
5	3	3	1	1	1	3	100	3	1	
6	3	3	1	1	1	3	100	3	1	
11	3	3	1	1	1	3	100	3	1	
2	2	1	1	1	1	2	100	2		
3	2	1	1	1	1	2	100	2		
4	2	1	1	1	1	2	100	2		

Mostrando registros del 1 al 10 de un total de 13 registros Anterior 1 2 Siguiente

Autores por subgrafo

Detalle subgrafo seleccionado

Detalle nodos del subgrafo seleccionado

Fig. 3.23: Sección Detalles de los componentes: Estructura Subgrafos. Se muestran los datos de métricas de red para distintos componentes a modo de ejemplo.

1010

1011 En la subsección Autores por subgrafo se detalla la cantidad y los autores que participan de
 1012 cada componente. Es posible realizar búsquedas por autores, ordenar por cantidades, y
 1013 también exportar los datos ([Fig. 3.24](#)).
 1014

Autores por subgrafo

Bajar autores

Mostrar 10 registros Buscar:

componente_id	cantidad_autores	autores
1	14	Abraham I. Kahan; Alicia Liliana Caratini; Alicia M. Arzayago; Alicia Susana Golcoches; Ana Haedo; Cristina Beatriz Dejean; Emile Crognier; Francisco Raúl Camese; Jorge A. Rey; Liliana Ingold; Marcelo Soria; Nilda Zuberti; Ricardo C. Nibosić; Silvia Elizondo
2	2	Alberto J. Maredline; Dario A. Demarchi
3	2	Andrés G. Bolzan; Luis Manuel Guimarey
4	2	Delia Beatriz Lomaglio; Elvira V. Sofía de Romeo
5	3	Eduardo R. Scarano; Marcelo Prati; Marta Civos
6	3	Elba N. Giovatelly; Patricia R. Pupkin; Virginia B. Fernández Varela
7	5	Elvira Inés Baffi; Francisco Rothhammer; Héctor Hugo Varela; José Alberto Cicciolio; María Fernanda Torres
8	4	Emma L. Alfaro; Ignacio F. Bejarano; José Edgardo Dipietri; Susana Blanca Ocampo
9	4	Evelia Edith Oyenart; Héctor Mario Pucciarelli; María Cristina Murie; Vicente Dressino
10	1	Fernando V. Ramírez-Rozzi

Mostrando registros del 1 al 10 de un total de 13 registros Anterior 1 2 Siguiente

Fig. 3.24: Detalle de la subsección Autores por subgrafo.

1015

1016 En Detalle del subgrafo se encuentra el detalle de los artículos relacionados al componente
 1017 seleccionado en la subsección Autores por subgrafo ([Fig. 3.25](#)). Es posible realizar búsquedas,
 1018 que aplican a todas las columnas, y exportar los datos asociados.
 1019

1020

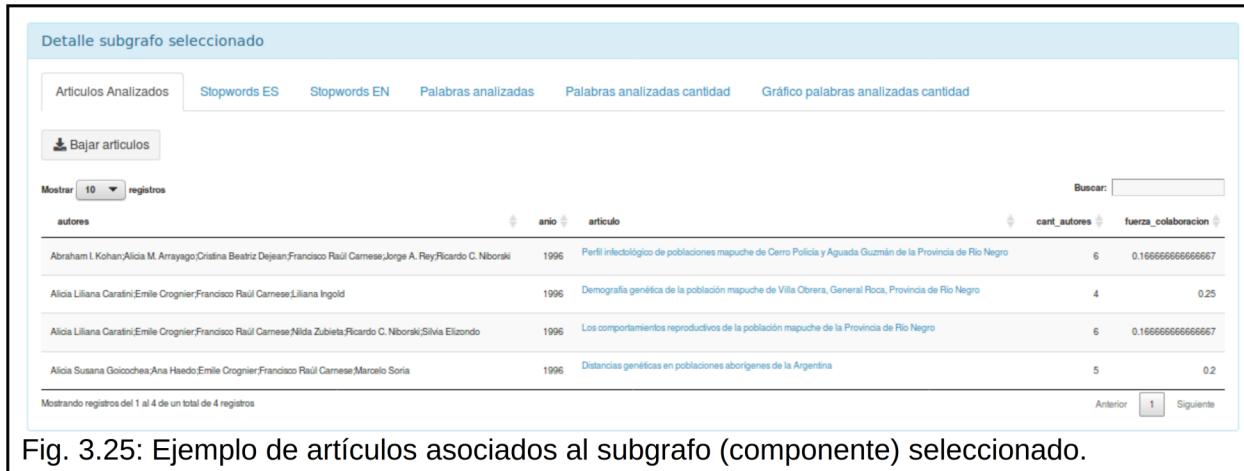


Fig. 3.25: Ejemplo de artículos asociados al subgrafo (componente) seleccionado.

1021

1022 También se cuenta con secciones relacionadas con análisis de texto. Puede verse, por
 1023 ejemplo, la frecuencia de palabras en los títulos de los artículos asociados al componente
 1024 seleccionado ([Fig. 3.26](#)).
 1025



Fig. 3.26: Ejemplo de cantidad de palabras en los artículos asociados al componente seleccionado.

1026

1027 Finalmente, se ofrece un detalle de los nodos del subgrafo o componente seleccionado ([Fig.](#)
 1028 [3.27](#)) con las mismas funcionalidades descriptas anteriormente: métricas de nodos, distribución
 1029 de grado, histograma de variables y densidad de variables.
 1030

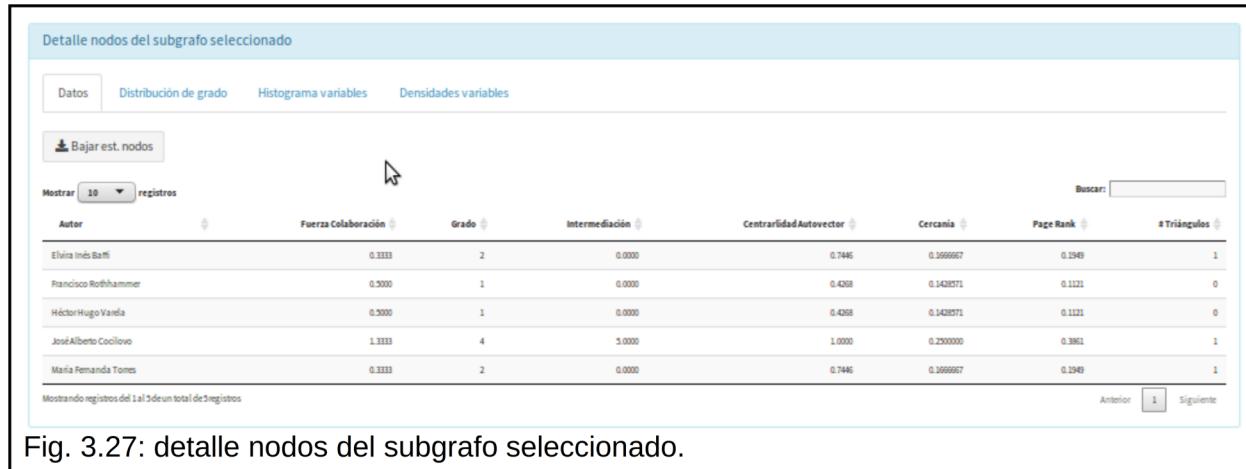


Fig. 3.27: detalle nodos del subgrafo seleccionado.

1031 Comparación de modelos

1032 En los análisis de redes es de utilidad realizar comparaciones con modelos. En palabras de
1033 [Kolaczyk-Csárdi \(2014\)](#):

1034

En la práctica, los modelos de redes de grafos son usados para diferentes propósitos. Estos incluyen: 1) la prueba de “significancia” de ciertas características predefinidas en una determinada red , 2) el estudio de mecanismos propuestos para generar ciertas propiedades que se observan comúnmente en redes del mundo real (como distribuciones de grado amplias o efectos de mundo pequeño) o 3) la evaluación de potenciales factores predictivos de relaciones.

1035

1036 Esta subsección está compuesta por tres partes: Datos del modelo, Modelos y Verificaciones
1037 de propiedades. Para la creación de los grafos a partir de modelos se utilizó la [librería igraph](#).

1038

1039 Para los estudios de Mundo Pequeño (*Small World*) y Libre escala (*Scale-Free*) se tienen en
1040 cuenta ciertas propiedades de la red, que se usan para hacer las comparaciones de la red con
1041 modelos. Las propiedades analizadas fueron las siguientes ([Tabla 3.4](#)):

1042

1043 Tabla 3.4. 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 test 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 test 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.
---	--

1044

1045

1046 La forma en la cual se puede visualizar esta tabla en la aplicación se muestra en la [Fig. 3.28](#)

1047

Fig. 3.28: Sección de comparación con modelos. Se observan los datos útiles de la red que fueron elegidos para la comparación.

1048

1049 En la solapa de modelos disponemos de tres modelos que se pueden simular para comparar
1050 con la red actual seleccionada previamente. Para cada uno de estos modelos existen distintas
1051 secciones: configuración de parámetros ([Fig. 3.29](#)), ejecución del modelo, histogramas para
1052 comparar, y los datos resultantes de la simulación.

1053

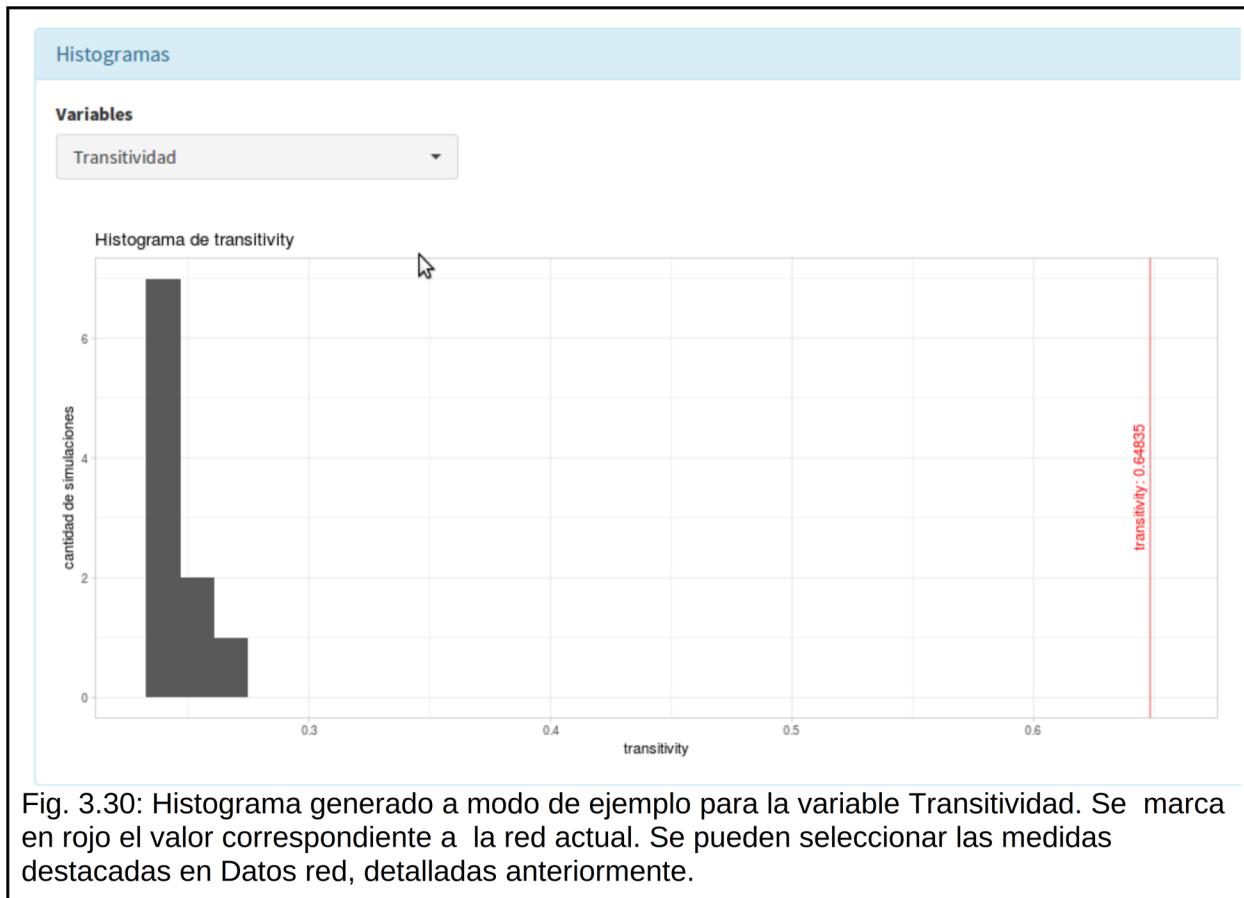
1054 Los modelos incluidos en la aplicación son el de Red Aleatoria según el modelo Erdos-Renyi
1055 configurado a través de nodos y aristas, Mundo pequeño de Watts-Strogatz y Libre escala de
1056 Barabasi-Albert. Para mayor detalle de estos modelos se recomienda leer el capítulo 5 de
1057 [Kolaczyk-Csárdi \(2014\)](#).

1058

Fig. 3.29: Configuración de los diferentes modelos incluidos en la aplicación. Los parámetros que figuran por default para cada uno están asociados a métricas de la red actual seleccionada.

1059

1060 Luego de la ejecución del modelo podemos evaluarlo de forma visual a través de histogramas
1061 de los modelos en comparación con la red ([Fig. 3.30](#)). Para crear los histogramas se utilizó la
1062 librería ggplot.



1065 Por otro lado, se pueden explorar los datos asociados a la estructura de la red de cada una de
 1066 las iteraciones del modelo teniendo en cuenta las métricas presentadas anteriormente al
 1067 describir las métricas de interés para la comparación con modelos ([Fig. 3.31](#)).
 1068

Datos simulación							
Bajar simulación	Mostrar 10 registros	Buscar:					
transitividad	diametro	dist.camino medio	ley pot.alta	alta > 1	ley pot. signif. test KS	test KS significativo	
0.0309042230247997	5	3.07262471255162	12.6171626742252	true	0.93770940230126	false	
0.024743052912067	5	3.06171205983639	7.32471177611344	true	0.93770940230126	false	
0.0293122086113033	6	3.05398036622325	15.76412816668902	true	0.93770940230126	false	
0.0275299515676778	5	3.06129414223027	17.7504480226853	true	0.9999981129000393	false	
0.0234703873561257	5	3.06222204125497	7.68104207362733	true	0.99429316932892	false	
0.0253069746021642	5	3.04629951336245	7.74573478760112	true	0.99429316932892	false	
0.0246050669377235	6	3.06201388402101	16.893047777102	true	0.99429316932892	false	
0.02630391001525105	5	3.05739369290063	8.34119909273701	true	0.97233139503003	false	
0.02293506082937797	6	3.06429884934178	11.2436041153782	true	0.843712120902998	false	
0.02649842271229338	5	3.05693840572257	5.90325949158025	true	0.843712120902998	false	

Fig. 3.31: Datos asociados a la simulación según un modelo en particular.

1070 Finalmente en la parte de Verificaciones podemos contestar dos preguntas que se hacen

1071 frecuentemente en los análisis de redes: si la red es de Mundo Pequeño y si la red es de Libre
1072 Escala.

1074 **Validación de Mundo Pequeño**

1075 Para la validación de Mundo Pequeño utilizamos dos métodos de validación: el método 1
 1076 descripto en [Bialonski et al. \(2010\)](#) y el método de validación 2 realizado con la función
 1077 *smallworldness* de la librería [qgraph](#) que se basa en lo planteado por [Humphries-Gurney](#)
 1078 ([2008](#)). Visualmente, este análisis consta de tres paneles: dos de parámetros utilizados (uno
 1079 por cada método de validación) y uno de la comparación de ambos métodos ([Fig. 3.32](#)).
 1080

<p>Verificación 1</p> <p>Validación 1:</p> <p>para validar small world se tomó el criterio de comparar la red actual contra redes aleatorias (lo seleccionado en el tab de Modelos / Aleatorio).</p> <p>la comparación se lleva a cabo en el promedio del camino mas corto VS la media del promedio del camino más corto de las N redes simuladas.</p> <p>para mas detalles ver: Bialonski (2010) donde se detalla el procedimiento.</p> <p>Mundo pequeño: si Delta ~=1 y Gamma > 1</p> <p>Delta = L/Lr. y Gamma = C/Cr.</p> <p>L: promedio camino más corto</p> <p>Lr: promedio de promedio camino más corto de las simulaciones</p> <p>C: transitivity</p> <p>Cr: promedio de transitivity de las simulaciones</p> <p>Delta: 0.434632992787933</p> <p>Gamma: 10.4603591478984</p> <p>Es mundo pequeño: NO</p>	<p>Verificación 2</p> <p>Validación 2:</p> <p>Una red puede ser considerada 'mundo pequeño' si la 'smallworldness' es mayor a 1 (qgraph::smallworldness)</p> <p>Un punto de vista más estricto dice de la red llevarla a 'smallworldness' >= 3 (Humphries & Gurney, 2008)</p> <p>smallworldness: 1.23583815580434</p> <p>mayor a 1: Si</p> <p>mayor a 3: NO</p>
<p>Comparación de Validaciones</p> <p>Recordar que para Validación 2, se vuelven a ejecutar simulaciones.</p> <p>Sobre las simulaciones de Validación 1, se puede ver su detalle en su pestaña</p> <p>Sobre las simulaciones de Validación 2, se puede ver su detalle en su pestaña</p> <p>para cada variable: valor validación 1 valor validación 2</p> <p>transitivity [red]: 0.648351648351648 0.648351648351648</p> <p>transitivity [promedio simulaciones]: 0.0619817770293199 0.134285714285714</p> <p>promedio distancia del camino mas corto [red]: 1.472 39.5040404040404</p> <p>promedio distancia del camino más corto [promedio simulaciones]: 3.38676544216749 10.1116565656566</p>	

Fig. 3.32: Paneles de criterios de evaluación de mundo pequeño.

1082 La diferencia de valores que aparece en la comparación de métodos ([Fig. 3.32](#)) está asociada
 1083 al método de generación de las redes aleatorias. En la validación 1 se usa el método de Erdos-
 1084 Renyi implementado a traves del método `sample_gnm` de la librería [igraph](#) que se basa en la
 1085 publicación de [Erdős-Renyi \(1959\)](#); mientras que en la función de la librería [qgraph](#) se utiliza la

1086 función para generar los aleatorios de la librería [igraph](#) `degree.sequence.game` pasando el
1087 método de generación “*simple.no.multiple*”. Para más información se puede ver la
1088 documentación de la librería [igraph](#).

1089 *Validación de Libre Escala*

1090 Según lo escrito por [Barabási \(2016\)](#):

1091

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

1092

1093 Para verificar que la distribución de grado de la red sigue la ley de potencia utilizamos la
1094 función `fit_power_law` de la librería [igraph](#), que se basa en lo planteado por [Clauset, et al.](#)
1095 ([2009](#)) y [Newman \(2005\)](#). Esta función recibe como parámetros los valores a ajustar a la ley de
1096 potencia, los utilizados dentro de la herramienta es una lista con el grado de todos los nodos de
1097 la red, y el tipo de implementación `plfit`.

1098 En nuestro caso utilizamos como valores a ajustar los grados de los nodos de la red y la
1099 implementación `plfit` que utiliza máxima verosimilitud para determinar el alfa dado un parámetro
1100 `xmin`. Cuando no se brinda el parámetro `xmin`, como es nuestro caso, el algoritmo trata de
1101 buscar el valor óptimo para maximizar el p-valor de un test de Kolmogorov-Smirnov entre la
1102 distribución ajustada y la muestra original. La función utiliza el método de [Clauset, et al. \(2009\)](#)
1103 para calcular los parámetros de la distribución ajustada.

1104

1105 Al aplicar la función, se obtiene el valor alfa correspondiente y los resultados de la prueba de
1106 Kolmogorov-Smirnov. Acorde a [Barabási \(2016\)](#) para que la red analizada sea de Libre Escala
1107 el valor alfa debe ser mayor a 2. Luego puede verificarse la hipótesis de ajuste de la
1108 distribución de grado a la distribución de ley de potencia mediante la prueba de Kolmogorov-
1109 Smirnov.

1110 Por lo tanto, si el p-valor de la prueba de Kolmogorov-Smirnov es mayor a 0.05 y asimismo el
1111 alfa es mayor a 2, la red es considerada de Libre escala. En la [Fig. 3.33](#) se observa la
1112 implementación de esta prueba en la aplicación.

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: 2.36183466980806

El exponente es mayor a 2: TRUE

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

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

El test de ajuste de Kolmogorov-Smirnov significativo?

p-value test KS: 0.746254010970353

menor a 0.05: FALSE

Es de Libre Escala.

Fig. 3.33: Implementación de la evaluación de Libre Escala.

1114 Comunidades

1115 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.

1116

1117 La necesidad de búsqueda de comunidades está asociada a buscar patrones según ciertas
1118 propiedades de la red. Esta búsqueda de patrones se puede realizar de diferentes maneras,
1119 para lo cual existen distintos algoritmos.

1120

1121 Por lo tanto, para poder analizar comunidades de una red específica, primero se debe
1122 seleccionar el algoritmo con el cual se van a buscar dichas comunidades. Esta opción está
1123 disponible en el panel de configuración de la sección de Comunidades, así como la posibilidad
1124 de seleccionar cuántas comunidades se desean ver en la visualización de la red ([Fig. 3.34](#)).
1125 Para más detalles de los algoritmos disponibles ver el Anexo: Detalle de algoritmos para
1126 búsqueda de comunidades.
1127

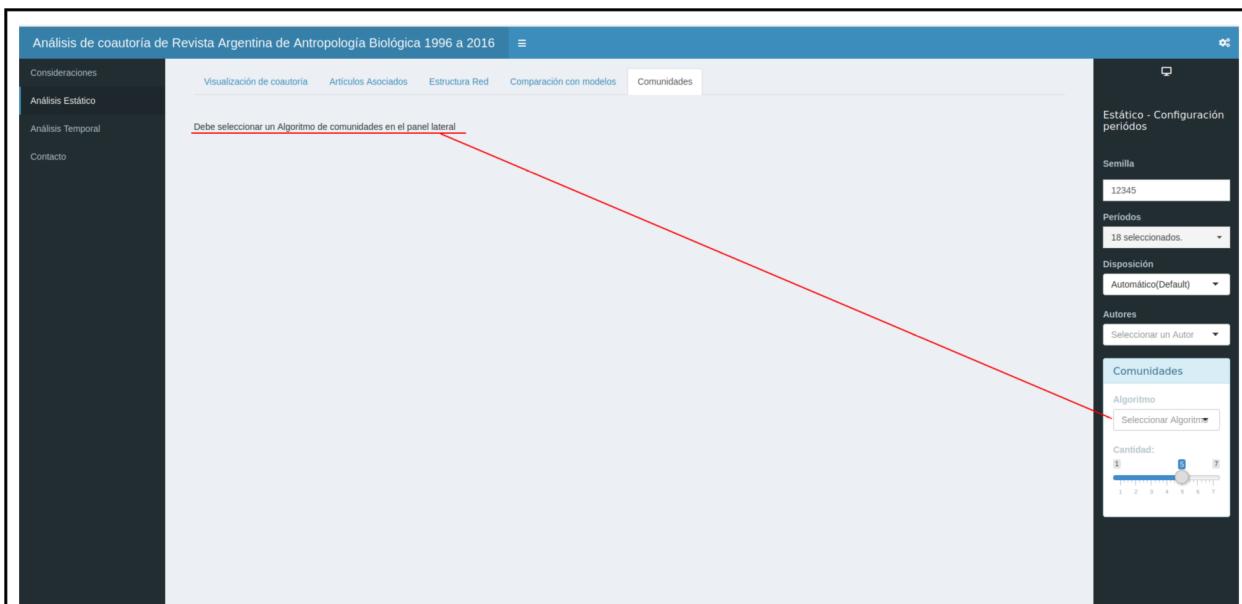


Fig. 3.34: Panel de configuración de búsqueda de comunidades.

1128

1129 Luego de seleccionar el algoritmo, en la sección de Visualización de coautoría se puede ver
1130 cómo están distribuidas la totalidad de las comunidades obtenidas ([Fig. 3.35](#)) .
1131

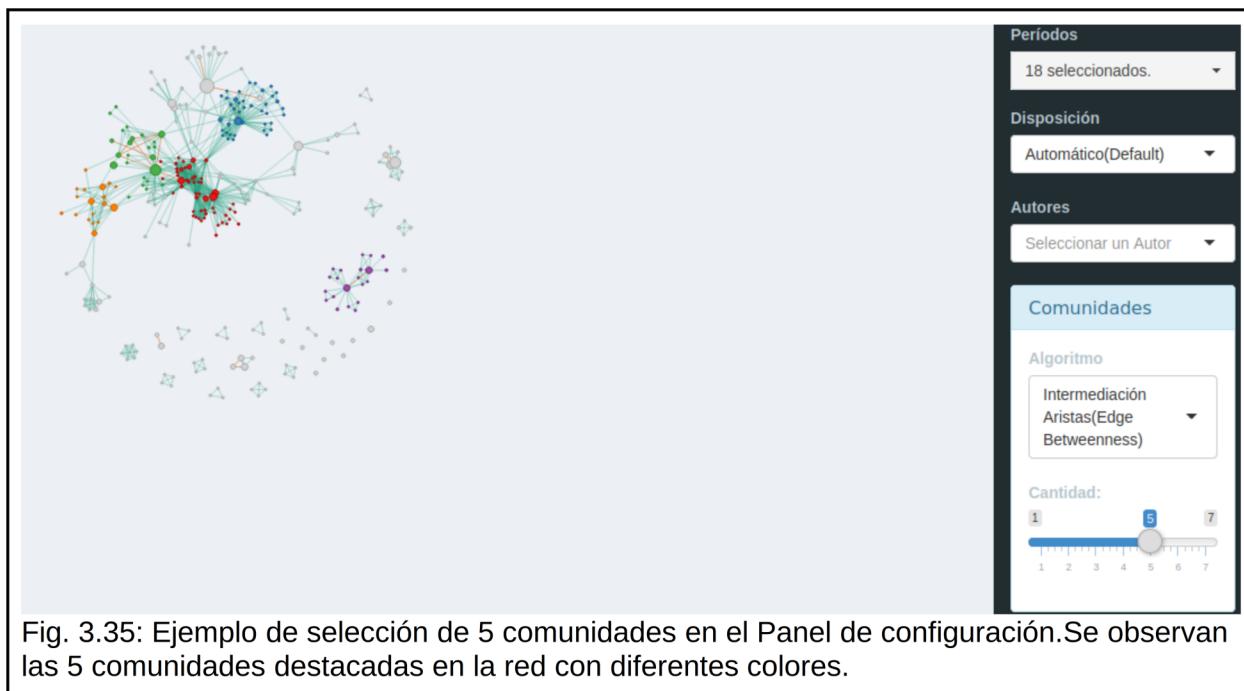


Fig. 3.35: Ejemplo de selección de 5 comunidades en el Panel de configuración. Se observan las 5 comunidades destacadas en la red con diferentes colores.

1132

1133 Para complementar la visualización, se puede proceder con un análisis de la composición de
 1134 los subgrafos resultantes de la búsqueda de comunidades, con ese objetivo se dispuso de la
 1135 sección Comunidades (Fig. 3.36). Como detalle se describe qué algoritmo fue utilizado para la
 1136 búsqueda, la cantidad total de comunidades y, para evaluar la relevancia de los resultados de
 1137 los algoritmos, está disponible la medida de modularidad (Newman-Girvan 2004). Una vez
 1138 seleccionado el algoritmo se puede ver en el panel de comunidades el valor de esta medida
 1139 teniendo en cuenta y sin tener en cuenta el peso de las relaciones (Fig. 3.36). Cabe recordar
 1140 que los pesos de las relaciones en este caso están dados por la fuerza de colaboración de
 1141 cada relación.

1142

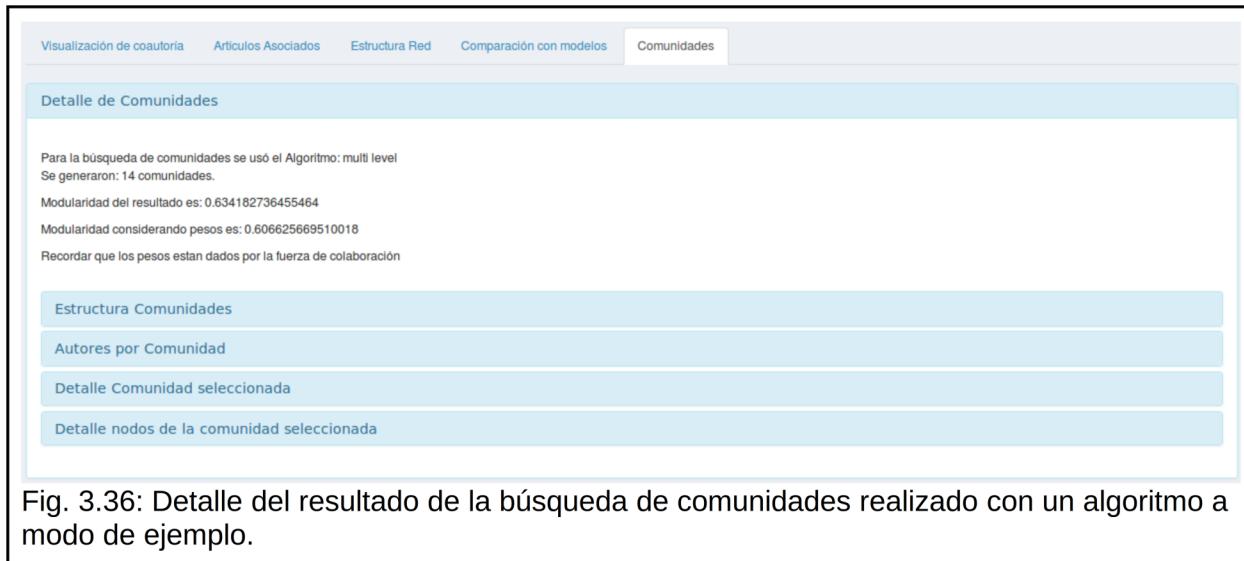


Fig. 3.36: Detalle del resultado de la búsqueda de comunidades realizado con un algoritmo a modo de ejemplo.

1143

1144 Además es posible realizar un análisis de la estructura de cada comunidad ([Fig. 3.37](#)), en el
 1145 que se puede ver qué autores forman parte de ella y los artículos asociados. Esto es similar a
 1146 lo disponible para los componentes de la red, solo que esta vez el subgrafo es cada comunidad
 1147 en vez de cada componente.

1148

1149

Autores por Comunidad		
 Bajar autores		
Mostrar	10	registros
componente_id	cantidad_autores	autores
1	9	Abraham I. Kohan;Alicia Liliana Caradini;Alicia M. Arrayago;Cristina Beatriz Dejean;Francisco Raúl Camese;Jorge A. Rey;Nilda Zubietta;Ricardo C. Niborski;Silvia Elizondo
2	2	Alberto J. Marcellino;Dario A. Demarchi
3	4	Alicia Susana Gómezchea;Ana Heedo;Emile Cognier;Marcelo Soria
4	2	Andrés G. Bolívar;Luis Manuel Guimarey
5	2	Delia Beatriz Lomiglio;Elvira V. Soria de Romero
6	3	Eduardo R. Scarano;Marcelo Prati;Marta Civico
7	3	Elba N. Giovattoni;Patricia R. Pupkin;Virginia B. Fernandez Varela
8	5	Elvira Inés Battif;Francisco Rothhammer;Héctor Hugo Varela;José Alberto Coddio;María Fernanda Torres
9	4	Emma L. Alfaro;Ignacio F. Bejarano;José Edgardo Dipieri;Susana Blanca Ocampo
10	4	Evelia Edith Oyenart;Héctor Mario Pucciarelli;María Cristina Muñiz;Violente Dressino

Mostrando registros del 1 al 10 de un total de 15 registros

Anterior 1 2 Siguiente

Fig. 3.37: Resultado de la búsqueda de comunidades: autores por comunidad.

1150

1151 Asimismo, es posible obtener las mismas métricas que las descriptas para el análisis de
 1152 componentes, solo que en este caso, los subgrafos son las comunidades en vez de los
 1153 componentes ([Fig. 3.38](#)).

1154

Estructura Comunidades											
 Bajar métricas											
Mostrar	10	registros									Buscar:
componente_id	cant_autores	cant_relaciones	densidad_red	distancia_media	num_dist_lejanos	num_largest_cliques	porc_largest_cliques	num_cliques	num_transitivity	num_assort_degree	
1	9	24	0.6666666666666667	1.333333333333333	2	6	66.66666666666667	6	0.789473684210526	-0.302325581395349	
2	2	1	1	1	1	2	100	2			
3	4	6	1	1	1	4	100	4	1		
4	2	1	1	1	1	2	100	2			
5	2	1	1	1	1	2	100	2			
6	3	3	1	1	1	3	100	3	1		
7	3	3	1	1	1	3	100	3	1		
8	5	5	0.5	1.5	2	3	60	3	0.375	-0.8055555555555556	
9	4	4	0.6666666666666667	1.333333333333333	2	3	75	3	0.6	-0.714285714285714	
10	4	4	0.6666666666666667	1.333333333333333	2	3	75	3	0.6	-0.714285714285714	

Mostrando registros del 1 al 10 de un total de 15 registros

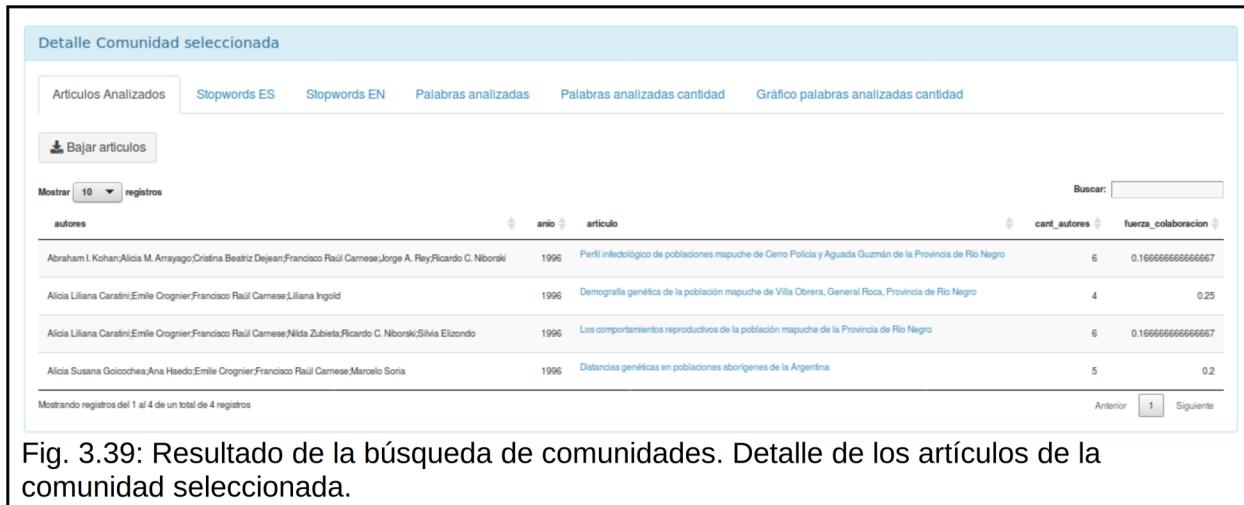
Anterior 1 2 Siguiente

Fig. 3.38: Métricas de los subgrafos de comunidades

1155

1156

1157 Luego de seleccionar una comunidad, en la tabla de autores por comunidad se dispone de los
 1158 detalles de la misma, entre los que se encuentran los artículos asociados ([Fig. 3.39](#)).
 1159



The screenshot shows a table titled "Detalle Comunidad seleccionada" with the following columns: autores, anio, artículo, cant_autores, and fuerza_colaboracion. The table contains four rows of data:

autores	anio	artículo	cant_autores	fuerza_colaboracion
Abraham I. Kohan;Alicia M. Arrayago;Cristina Beatriz Dejean;Francisco Raúl Carnese;Jorge A. Rey;Ricardo C. Niborski	1996	Perfil infectológico de poblaciones mapuche de Cerro Policia y Aguada Guzmán de la Provincia de Río Negro	6	0.1666666666666667
Alicia Liliana Caratini;Emile Cognier;Francisco Raúl Carnese;Liliana Ingold	1996	Demografía genética de la población mapuche de Villa Obrera, General Roca, Provincia de Río Negro	4	0.25
Alicia Liliana Caratini;Emile Cognier;Francisco Raúl Carnese;Nilda Zubielta;Ricardo C. Niborski;Silvia Elizondo	1996	Los comportamientos reproductivos de la población mapuche de la Provincia de Río Negro	6	0.1666666666666667
Alicia Susana Goicochea;Ana Haedo;Emile Cognier;Francisco Raúl Carnese;Marcelo Soria	1996	Distancias genéticas en poblaciones aborígenes de la Argentina.	5	0.2

Mostrando registros del 1 al 4 de un total de 4 registros

Fig. 3.39: Resultado de la búsqueda de comunidades. Detalle de los artículos de la comunidad seleccionada.

1160
 1161 De dichos artículos se pueden ver algunos resultados del análisis de texto realizado sobre los
 1162 títulos de los artículos, tales como: las palabras más frecuentes de esa comunidad, la cantidad
 1163 de palabras frecuentes y los filtros aplicados de *stop words* en español e inglés (dado que
 1164 existen algunos títulos en dicho idioma) ([Fig. 3.40](#)).
 1165

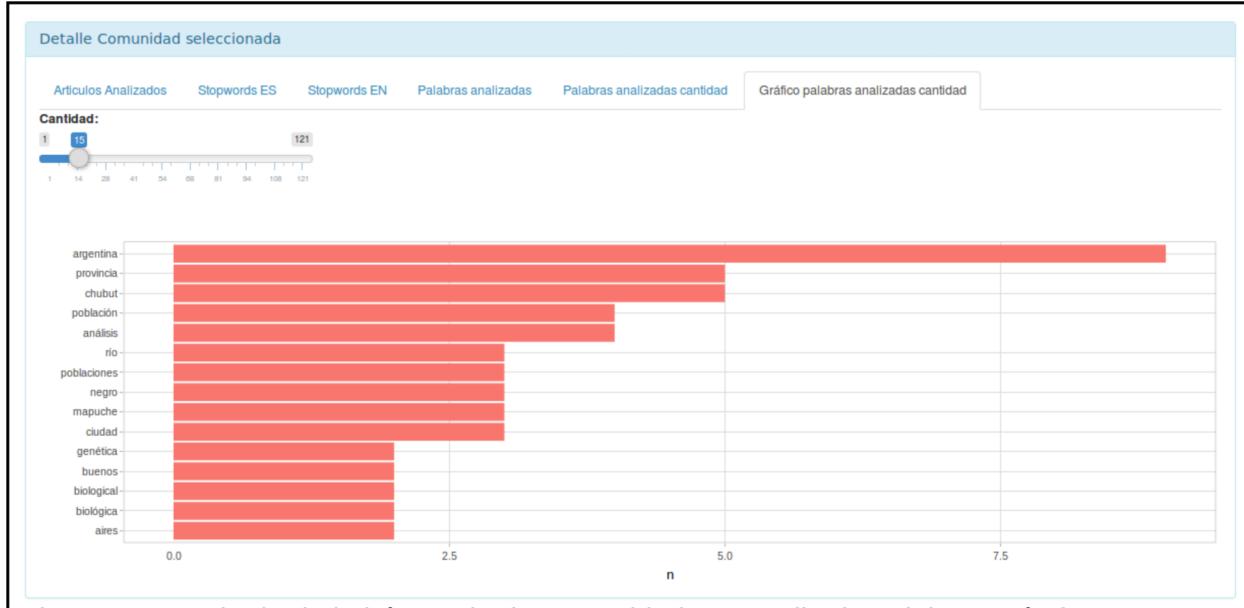


Fig. 3.40: Resultado de la búsqueda de comunidades. Detalle de palabras más frecuentes en los títulos de los artículos.

1166
 1167

1168 De manera similar a lo realizado para la sección de componentes, también se dispone del
 1169 detalle de los nodos de la comunidad seleccionada ([Fig. 3.41](#)).
 1170

Detalle nodos de la comunidad seleccionada							
Datos	Distribución de grado	Histograma variables	Densidades variables				
							Buscar: <input type="text"/>
Mostrar 10 registros							
Author	Fuerza Colaboración	Grado	Intermediación	Centrabilidad Autovector	Cercanía	Page Rank	# Triángulos
Abraham I. Kohan	0.1667	5	0.1429	0.5773	0.0909091	0.0976	10
Alicia Liliana Caratti	0.4167	4	0.0000	0.6119	0.0769231	0.0981	6
Alicia M. Arrayago	0.1667	5	0.1429	0.5773	0.0909091	0.0976	10
Cristina Beatriz Dejean	0.1667	5	0.1429	0.5773	0.0909091	0.0976	10
Francisco Raúl Carnese	0.7833	8	4.0000	1.0000	0.1000000	0.1824	16
Jorge A. Rey	0.1667	5	0.1429	0.5773	0.0909091	0.0976	10
Nilda Zubia	0.1667	4	0.4762	0.4756	0.0833333	0.0819	6
Ricardo C. Niborski	0.3333	8	8.0000	0.9266	0.1111111	0.1655	16
Silvia Elizondo	0.1667	4	0.4762	0.4756	0.0833333	0.0819	6

Fig. 3.41: Detalle de los nodos que forman parte de la comunidad seleccionada.

1171

1172 Sección 3: Análisis temporal

1173 En esta sección se realizó un análisis temporal de la red desde su primera publicación en 1996
 1174 hasta el periodo 2016 puesto como cota por 20 años desde su primera edición.

1175

1176 Como vistazo inicial del **Análisis temporal** se planteó una visualización dinámica. Para su
 1177 generación se utilizó el grafo acumulado desde el periodo 1996 como $T=0$ hasta 2016 como
 1178 $T=18$. De esta manera, a medida que se avanza en los tiempos aparecen nuevos nodos y se
 1179 modifican las relaciones.

1180

1181 Luego de tener una idea general de cómo se generaron las relaciones en el tiempo se pueden
 1182 obtener diferentes medidas que sirven para analizar diversos aspectos de la red como por
 1183 ejemplo qué nodos cumplen distintos roles. Dependiendo del objetivo, puede ponerse el foco
 1184 en una o varias de estas medidas para contestar las preguntas de interés. La herramienta
 1185 presenta dos enfoques de análisis de las medidas: de forma individual por periodo o de forma
 1186 acumulada total. Esto nos permite analizar de forma general las relaciones de autoría dentro de
 1187 la RAAB, de manera similar al trabajo realizado por [Koseoglu \(2016\)](#).

1188

1189 Finalmente la exploración de diferentes agregaciones temporales puede realizarse en la
 1190 sección Análisis estático de periodos, mediante la selección de los períodos específicos de
 1191 interés.

1192

1193 Animación del grafo de manera acumulada

1194 Esta herramienta muestra cómo se generaron las relaciones entre los nodos y cómo fueron
1195 conectándose los diversos componentes a lo largo del tiempo. Esto ayuda a focalizarnos en el
1196 periodo de interés para posteriormente realizar un análisis más detallado en la sección **Análisis**
1197 **estático**.

1198 En la animación se pueden ver los nodos en escala de azules según su fuerza de colaboración
1199 y los colores de las aristas según la fuerza de relación entre los autores (medida en fuerza de
1200 colaboración) (Fig. 3.42). A la derecha de la animación hay una tabla para ver las equivalencias
1201 entre períodos y años.

1202

1203 En el detalle de los nodos de la animación podemos ver quién es el autor, cuánta es la fuerza
1204 de colaboración acumulada, los períodos de participación y cantidad de artículos por periodo.
1205 En el detalle de las aristas podemos ver los autores participantes de dicha relación, la fuerza de
1206 colaboración resultante de las coautorías de los autores, los períodos en los cuales participan, y
1207 cuántos artículos contribuyeron en esos períodos.

1208

1209 Para poder realizar esta animación se usaron los paquetes igraph, ndtv, intergraph,
1210 networkDynamic y network. Para más detalles ver el [anexo de librerías](#).

1211

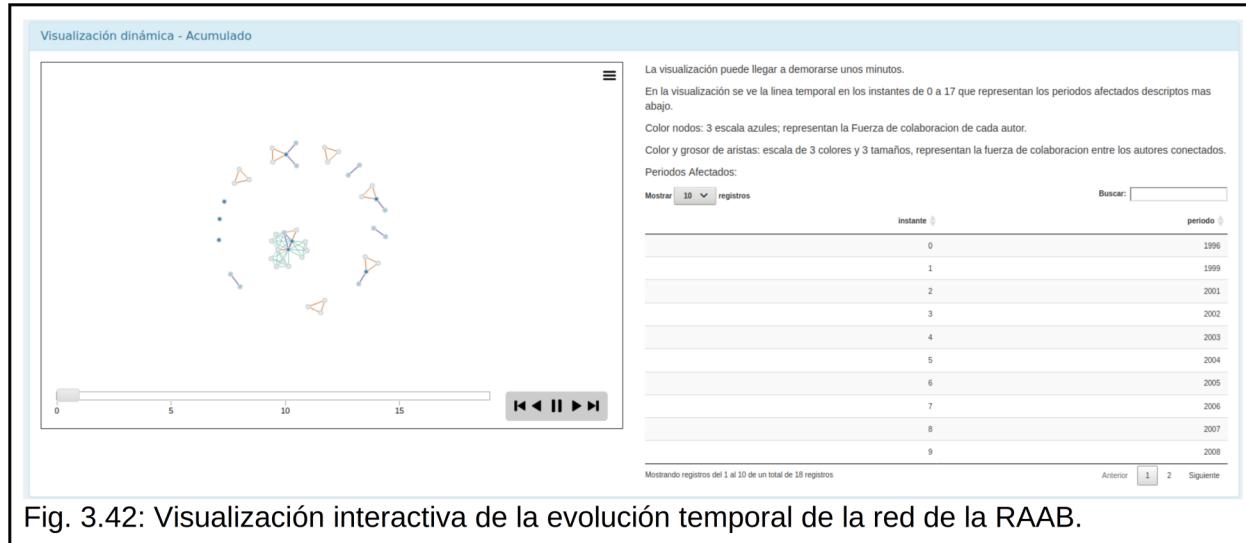


Fig. 3.42: Visualización interactiva de la evolución temporal de la red de la RAAB.

1212

1213 Medidas de cada periodo

1214 Esta sección fue diseñada para ver las medidas en su evolución por año. Las medidas
1215 disponibles son las mismas detalladas para la sección Métricas de la red en Estructura de la
1216 red. ([Fig. 3.43](#))

1217 Los gráficos fueron realizados con la [librería ggplot](#) y se hicieron interactivos utilizando la
1218 [librería plotly](#).

1219



Fig. 3.43: Métricas de la red para cada instante anual. Se muestran como ejemplo la cantidad de autores y la cantidad de artículos.

1220 Medidas Acumuladas a cada periodo

1221 Dado que en cada nueva edición de la revista se fueron incorporando más trabajos, resultó de
1222 interés ver la interacción entre los viejos nodos y los nuevos nodos, cuándo aparecen los
1223 nuevos y qué tanto aporte realizan en el total de la revista. Las medidas que se incorporaron en
1224 la visualización acumulada en el tiempo son las mismas que las detalladas previamente para
1225 Estructura de la red > Métricas de la red ([Fig. 3.44](#)).

1226

1227 Los gráficos fueron realizados con la [librería ggplot](#) y se hicieron interactivos utilizando la
1228 [librería plotly](#).

1229

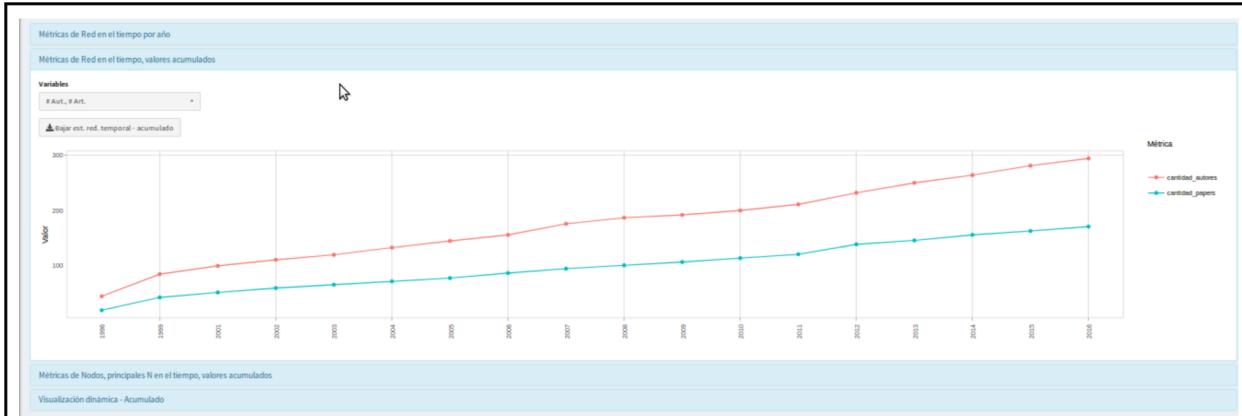


Fig. 3.44: Métricas acumuladas para cada instante anual. Se muestran como ejemplo la cantidad de autores y la cantidad de artículos.

1230

1231 Ranking de autores en el tiempo

1232 En esta sección es posible comparar a los autores a lo largo del tiempo según alguna métrica
1233 elegida y una cantidad de autores por año especificada (N) ([Fig. 3.45](#)).

1234 Se obtiene entonces un gráfico con las posiciones en el ranking de los N autores elegidos. Las
1235 líneas sirven para ver entre cuáles autores fueron variando las posiciones a lo largo del tiempo.
1236 Si se desea ver el valor de la métrica elegida, se puede visualizar en el detalle interactivo. Los
1237 empates entre autores se visualizan como múltiples puntos del mismo color en el mismo
1238 periodo, unidos por una línea vertical.

1239 Las métricas disponibles son las mismas que las empleadas en la sección Estructura de nodos,
1240 específicas anteriormente. Los gráficos fueron realizados con la [librería ggplot](#) y se hicieron
1241 interactivos utilizando la [librería plotly](#).

1242

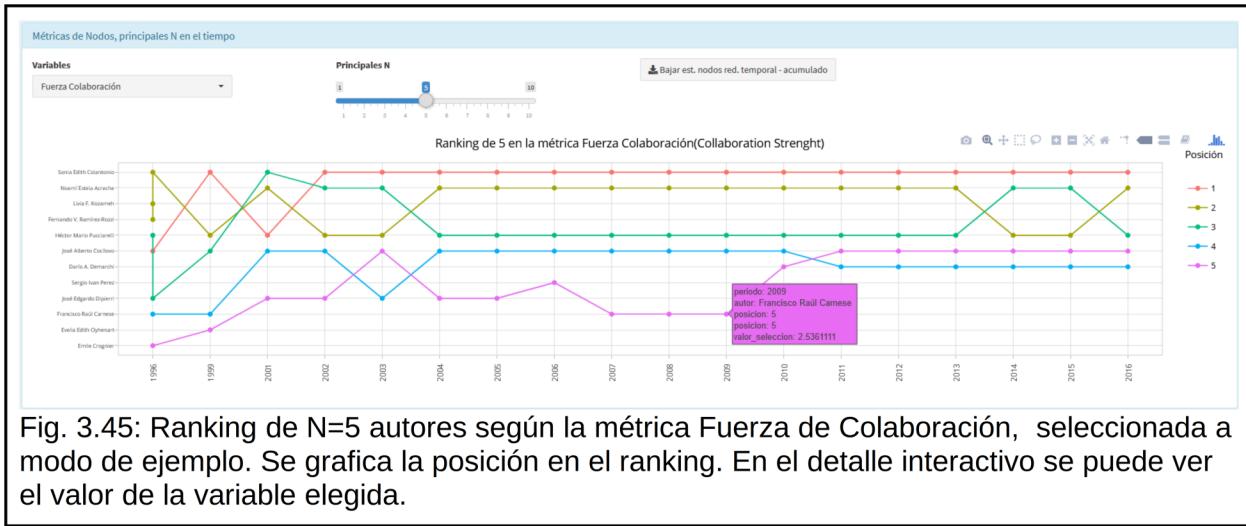


Fig. 3.45: Ranking de N=5 autores según la métrica Fuerza de Colaboración, seleccionada a modo de ejemplo. Se grafica la posición en el ranking. En el detalle interactivo se puede ver el valor de la variable elegida.

1243

1244 Ranking de autores en el tiempo para medidas acumuladas

1245 Esta sección es similar a la anterior, solo que se visualiza el color de la métrica acumulada en
1246 el tiempo ([Fig. 3.46](#)). Al igual que en el caso anterior, se elige una métrica de interés y la
1247 cantidad de autores a visualizar (N). En caso de empate aparecen todos los autores que estén
1248 entre los primeros N.

1249 Las métricas disponibles son las mismas que las empleadas en la sección Estructura de nodos,
1250 especificadas anteriormente. Los gráficos fueron realizados con la [librería ggplot](#) y se hicieron
1251 interactivos utilizando la [librería plotly](#).

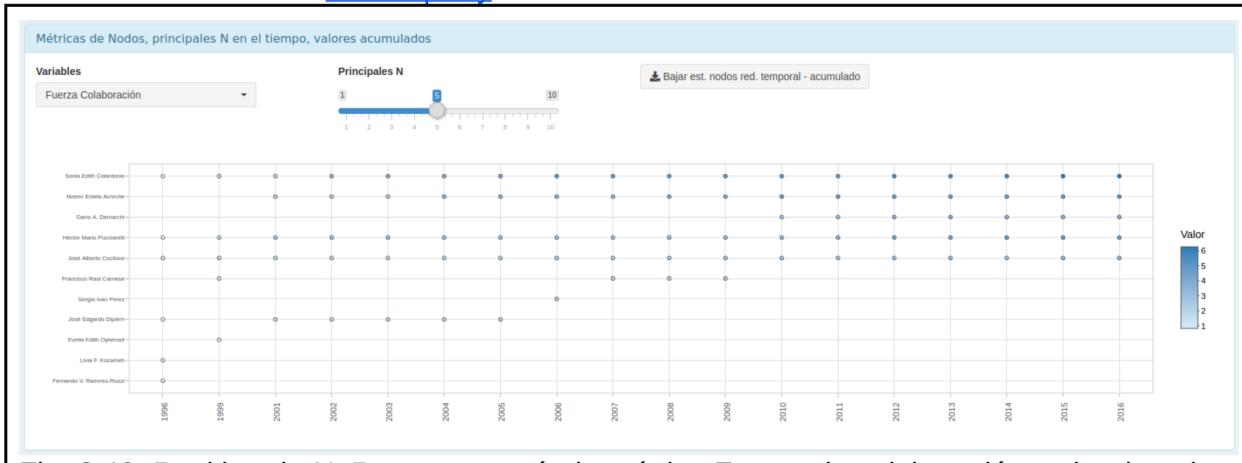


Fig. 3.46: Ranking de N=5 autores según la métrica Fuerza de colaboración, seleccionada a modo de ejemplo. En este caso se destaca en gradiente de color el valor acumulado de la métrica.

1252

1253 Resumen del capítulo

1254 En este capítulo se presentó la aplicación desarrollada y se describieron, para cada una de las
1255 secciones que la componen, su fundamentación, la metodología aplicada y las herramientas
1256 utilizadas en su desarrollo, su implementación con ejemplos y visualizaciones, así como
1257 también sus limitaciones.

1258

1259 En el próximo capítulo se usará esta herramienta para analizar el caso específico de la RAAB.

1260

1261 Capítulo 4: Análisis de la Red

1262 Introducción

1263 El periodo temporal analizado fue entre los años 1996 y 2016. Se optó por esta cota temporal
 1264 como límite factible para el análisis en el marco de la presente tesis, y puesto que conforma un
 1265 periodo de 20 años desde la primera publicación de la revista.

1266 Se utilizaron sólo los Trabajos originales para el análisis de la red de coautoría de la revista,
 1267 siguiendo a [Ordoñez et al. \(2009\)](#) quienes mencionan que uno de los criterios de calidad que
 1268 hacen a una revista especializada es el componente investigativo y original de sus artículos.

1269

1270 En el presente capítulo analizamos la red en su totalidad (es decir, de 1996 a 2016) y luego por
 1271 períodos acotados, a fin de evaluar posibles cambios en los patrones generales a lo largo del
 1272 tiempo.

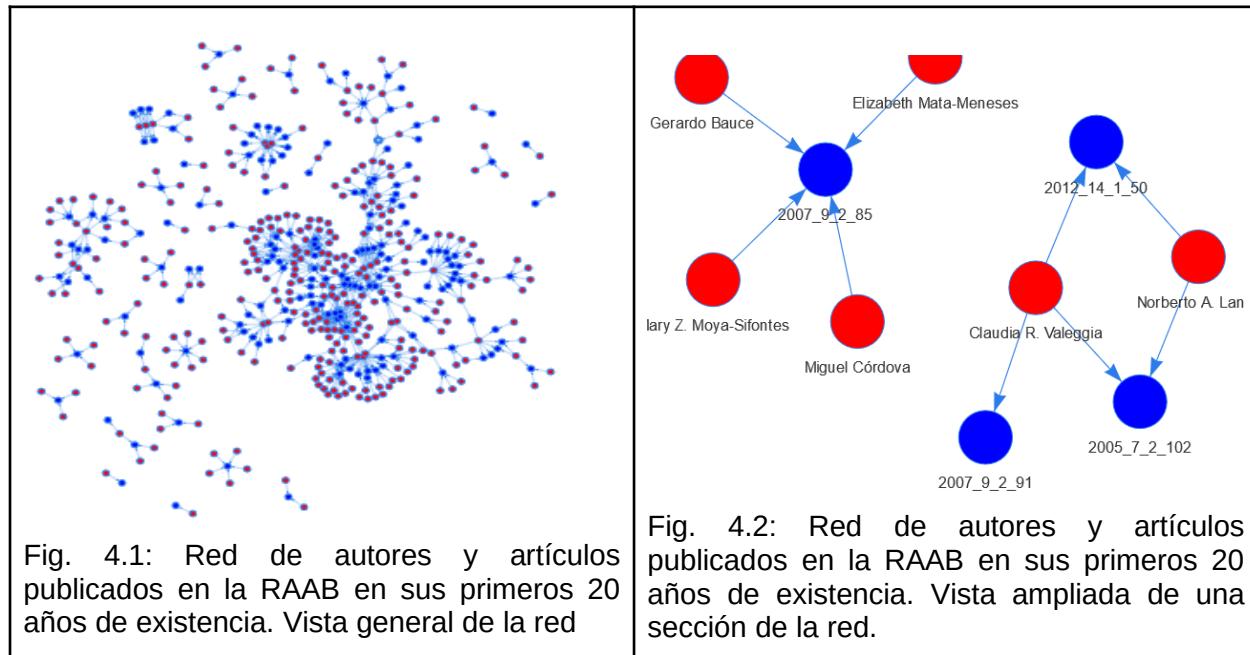
1273

1274 En el análisis contestamos las hipótesis de investigación planteadas en esta tesis así como
 1275 también otras preguntas que fueron surgiendo al familiarizarnos con los datos.

1276

1277 Antes de comenzar con el análisis, cabe recordar algunas características de la red bajo estudio.
 1278 En primer lugar, un autor puede ser autor de múltiples obras y una obra puede ser elaborada
 1279 por múltiples autores. Esto genera una red de autoría compuesta por dos tipos diferentes de
 1280 nodos, lo que se denomina un grafo bipartito. En el caso de la RAAB, analizada en este caso,
 1281 los autores son los científicos y las obras analizadas son los trabajos originales ([Fig. 4.1](#)).
 1282 Además, observamos que la red es dirigida, dado que los autores generan las obras ([Fig. 4.2](#)).

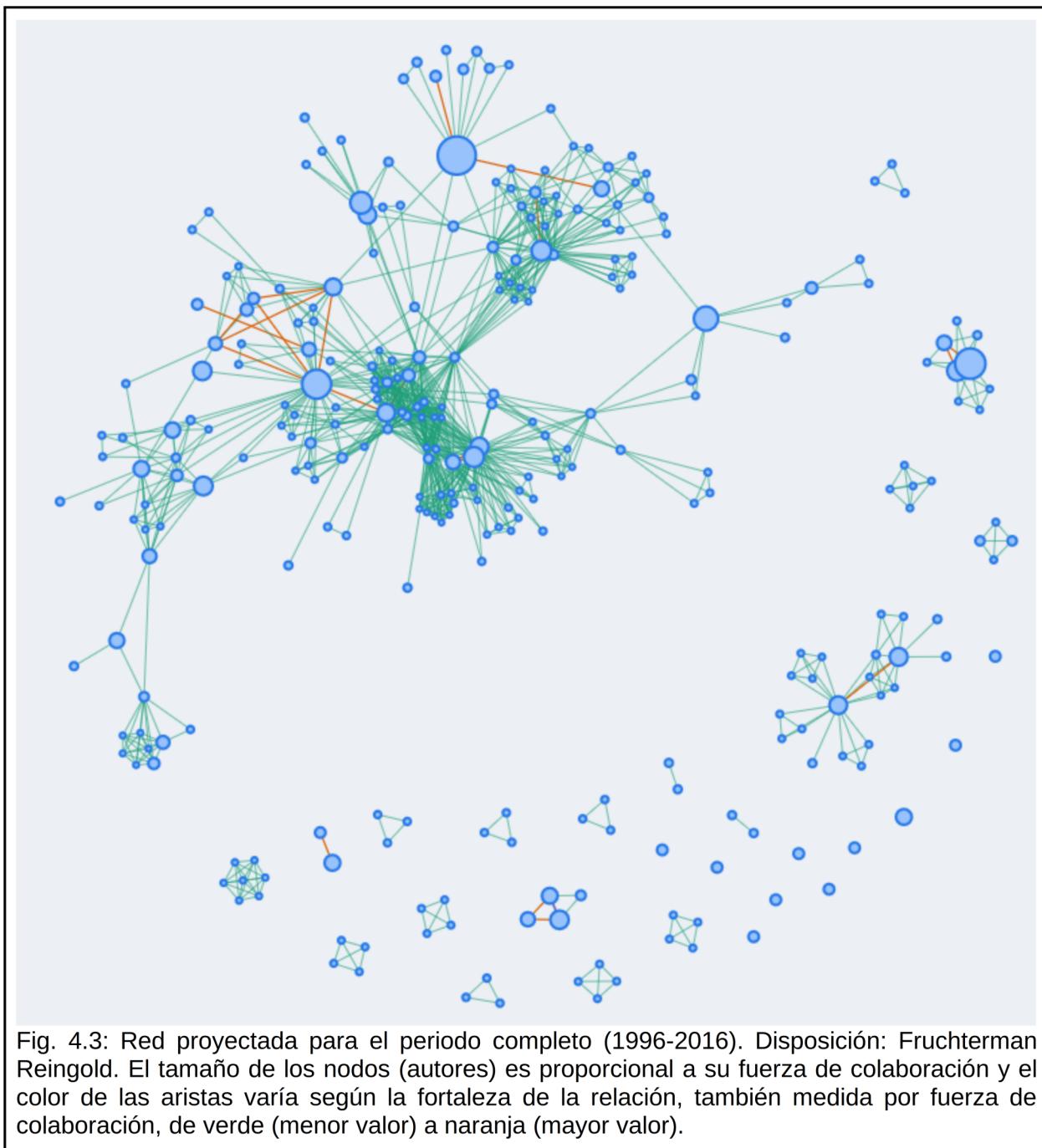
1283



1284

1285 Finalmente, para transformar esta red bipartita en una de coautoría utilizamos una proyección a
 1286 la dimensión de los autores, quedándonos así con los nodos autores y generando aristas de
 1287 “coautoría” o “trabajo conjunto”, a las cuales se les asignó un peso según la cantidad de

1288 trabajos involucrados. En este caso, observamos que la red pasó a ser no dirigida ([Fig. 4.3](#)),
 1289 dado que la obra generada es la relación y, en caso de existir múltiples relaciones, se simplificó
 1290 en una arista con mayor peso. Es decir que a mayor cantidad de trabajos en conjunto de un
 1291 dado par de autores, más peso tuvo la relación o arista que los unió. Para este análisis, no se
 1292 tuvo en cuenta el orden de autores en cada trabajo realizado; es decir, todos se encontraron en
 1293 la misma jerarquía.
 1294



1296 Algunas de las preguntas que surgieron al analizar este grafo ([Fig. 4.3](#)) fueron: ¿Los autores
 1297 individuales aportan igualmente a la red que aquellos que participan en equipos de dos

1298 personas? ¿Cuál es el aporte de los equipos de más de dos personas?

1299

1300 Para abordar estas cuestiones se tuvieron en cuenta diferentes métricas. Cabe destacar una en
1301 particular para este tipo de redes: la **fuerza de colaboración**. Esta métrica se define como un
1302 artículo repartido en la cantidad de autores (1 / N autores) ([Newman 2004](#)).

1303

1304 En la vista de la red de la [Fig. 4.3](#), se puede ver el tamaño de los autores según su fuerza de
1305 colaboración y, de la misma forma, el color de las aristas varía según la fortaleza de la relación
1306 medida por fuerza de colaboración.

1307

1308 Sin embargo, con esta métrica se destacan principalmente los autores que han publicado de
1309 manera individual por sobre los equipos, a veces interdisciplinarios, que pueden realizar otro
1310 tipo de aportes a la disciplina. Por lo tanto, en el caso de los nodos se utilizaron otras métricas
1311 de redes como el **Grado** de cada autor, que mide con cuántos otros autores se conecta
1312 directamente el autor, y la **Intermediación**, que mide la importancia de un autor según su
1313 participación en caminos para llegar de un autor a otro, generando equipos.

1314

1315 Como los autores pueden trabajar en forma conjunta o individual, o en equipos que no se
1316 conectan entre ellos, se espera encontrar una red inconexa, esto significa que no todos los
1317 nodos están conectados entre sí.

1318

1319 Cabe recordar finalmente el aspecto temporal del presente análisis puesto que puede decirse
1320 que la red pasó por varias etapas para llegar a su estadio final, en este caso el año 2016. Las
1321 distintas redes de coautoría de la RAAB pueden ser analizadas entonces por etapa/año o bien
1322 de forma agregada, es decir, teniendo en cuenta lo acumulado en etapas anteriores hasta la
1323 etapa seleccionada.

1324

1325 Teniendo presentes todos estos aspectos detallados previamente, a continuación se presentan
1326 los análisis realizados para poner a prueba las cuatro hipótesis de investigación planteadas en
1327 la presente tesis.

1328

1329

1330 **Hipótesis 1: La cantidad de autores aumenta en el tiempo debido
1331 al crecimiento de la disciplina en el país, lo que se refleja en un
1332 aumento en la cantidad de nodos en la red**

1333
1334 La primera hipótesis que planteamos trata de verificar si existe un crecimiento de la disciplina
1335 Antropología Biológica a nivel local. Para constatar esta hipótesis elegimos analizar las
1336 publicaciones de la Revista de Antropología Biológica (RAAB), que es el principal elemento de
1337 difusión de la Asociación Argentina de Antropología Biológica (AAAB). Revisar los artículos
1338 publicados en la RAAB nos da una idea si se fomenta el crecimiento local, una red interna de
1339 Argentina, en contraste de solo la red externa internacional. Estimamos que en ambas redes
1340 hay desarrollo, este trabajo se focaliza en la red interna argentina tomando la RAAB como
1341 representante de la materia a nivel local.

1342
1343 Tomamos como métrica la cantidad de autores en el tiempo, sin embargo esta métrica se
1344 puede ver de diferentes formas, acumulada en cada instante temporal o individualizando cada
1345 instante. Estas aproximaciones pueden otorgar diferentes puntos de vista sobre el contexto de
1346 estudio.

1347
1348 Si analizamos la cantidad en cada instante temporal tendremos mayor cantidad de autores
1349 según potenciales eventos en la existencia de la revista, como puede ser su inicio o congresos.
1350 Se estima que estos eventos afectan la producción de trabajos originales, haciendo crecer o
1351 reduciendo la cantidad de trabajos originales, y por ende, la cantidad de autores participantes
1352 en la red de coautoría de trabajos originales de la RAAB. Este punto de vista nos ayuda a
1353 comprender mejor cómo se comportan las dinámicas asociadas a los eventos de mayor o
1354 menor producción. Tal vez para plantear nuevas preguntas como cuánto tiempo tardan en
1355 aparecer nuevos trabajos siendo disparados por interacción en congresos.

1356
1357 Si tomamos las cantidades acumuladas de autores, estimamos que crecen en el tiempo en
1358 caso de crecer la revista, o mantenerse sin incrementos en caso de no crecer. Crecer implica
1359 que nuevos autores aparecen en la red, que estos pueden conformar nuevos equipos, o tal vez
1360 se integran a equipos existentes. Al revisar la cantidad acumulada van a ser de nuestro interés
1361 los saltos que hay entre edición y edición viendo la variación que hay entre cada momento.
1362 Mayor variación entre periodos significa que nuevos autores participan de la red, menor
1363 variación significa que la red se mantiene por los autores existentes.

1364
1365 Para tener un mejor contexto, se evaluó la cantidad de autores y publicaciones por año desde
1366 la primera edición de la RAAB ([Tabla 4.1](#)).

1367
1368 Recordemos que estas medidas corresponden generalmente a dos ediciones de la revista por
1369 año, lo cual en cierto sentido es una restricción a la cantidad de artículos publicables por año y
1370 quizás también a la de autores. Además, en esta primera aproximación no solo se tuvieron en
1371 cuenta trabajos originales, sino también otro tipo de publicaciones ([Tabla 10](#)).

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	2	1	0	0	0	1	1	0
	Cartas al Editor	0	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
	Comentarios Bibliográficos	0	0	0	0	0	0	0	0	0	2	0	0	1	0	1	2	1	0	0
	Comunicaciones Breves	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1
	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	7
edición	1	20	24	10	11	2	8	123	11	2	1	13	9	10	12	8	6	7	8	10
	2	0	0	1	0	7	0	7	0	10	5	0	0	0	0	0	5	10	11	7
	especial	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0

Tabla 4.1: Cantidad de publicaciones por año, discriminado en secciones y ediciones.* Al momento de la confección de la tabla, el año 2018 no estaba completo.

1373

1374 Al observar la cantidad total de publicaciones en la RAAB por año ([Fig. 4.4](#)) y discriminada en
 1375 ediciones (primera, segunda y especial) ([Fig. 4.5](#)) pudo verse una gran disparidad para el año
 1376 2005.

1377

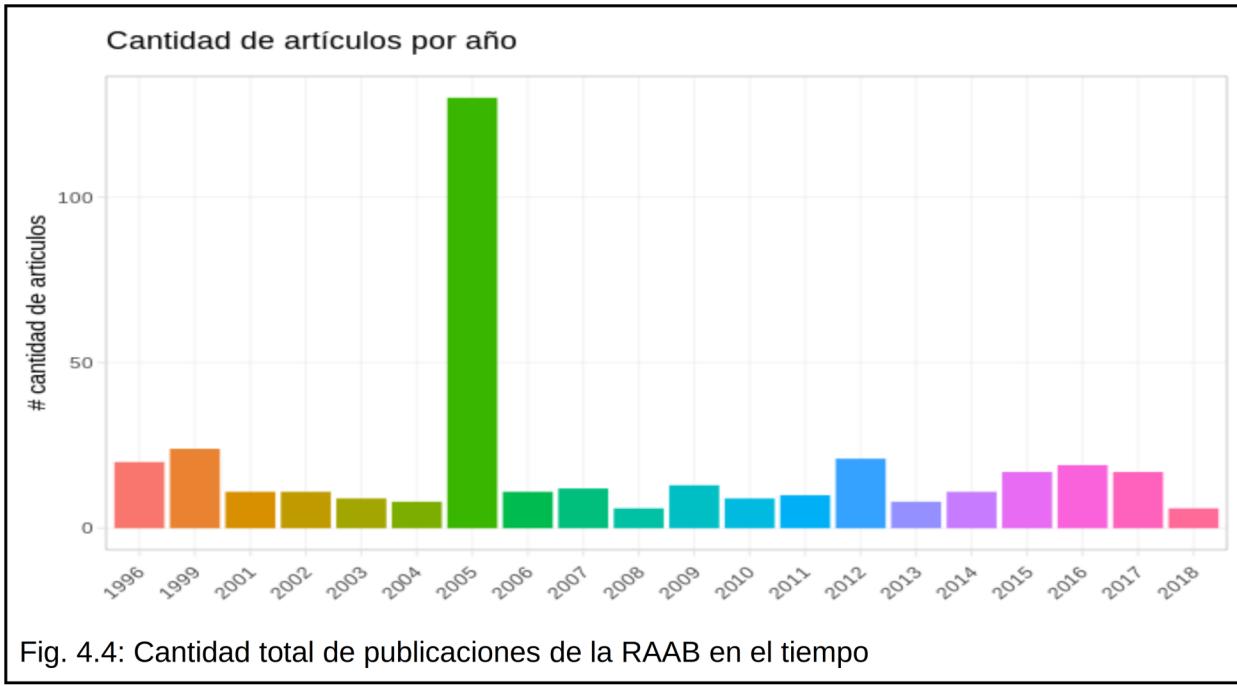
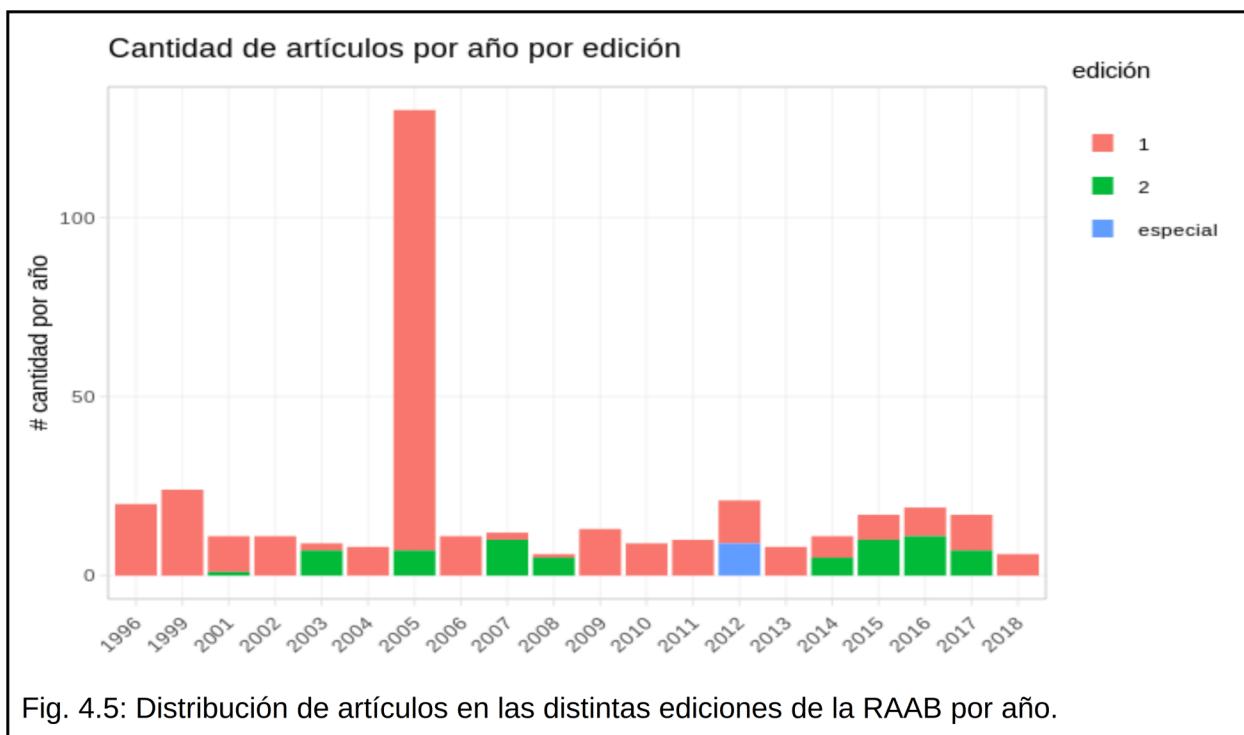


Fig. 4.4: Cantidad total de publicaciones de la RAAB en el tiempo

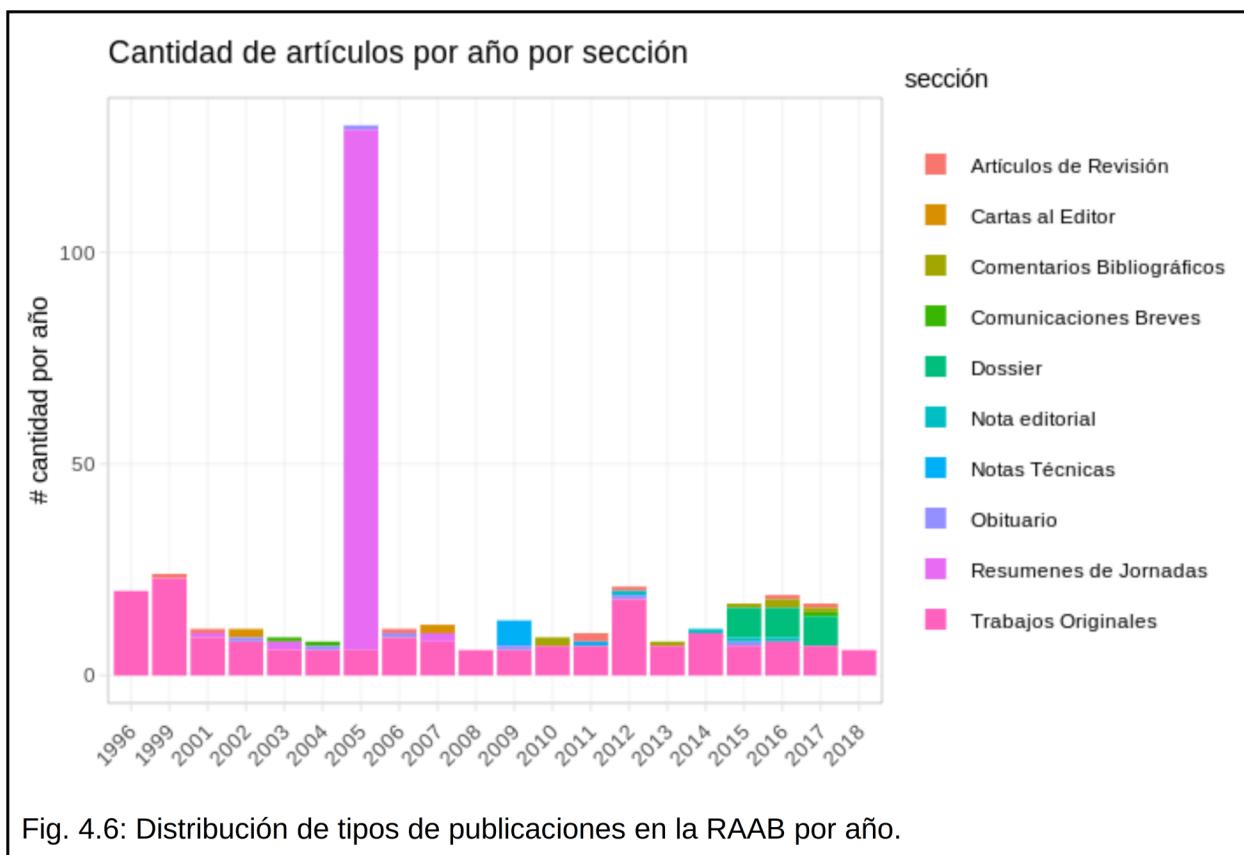
1378

1379



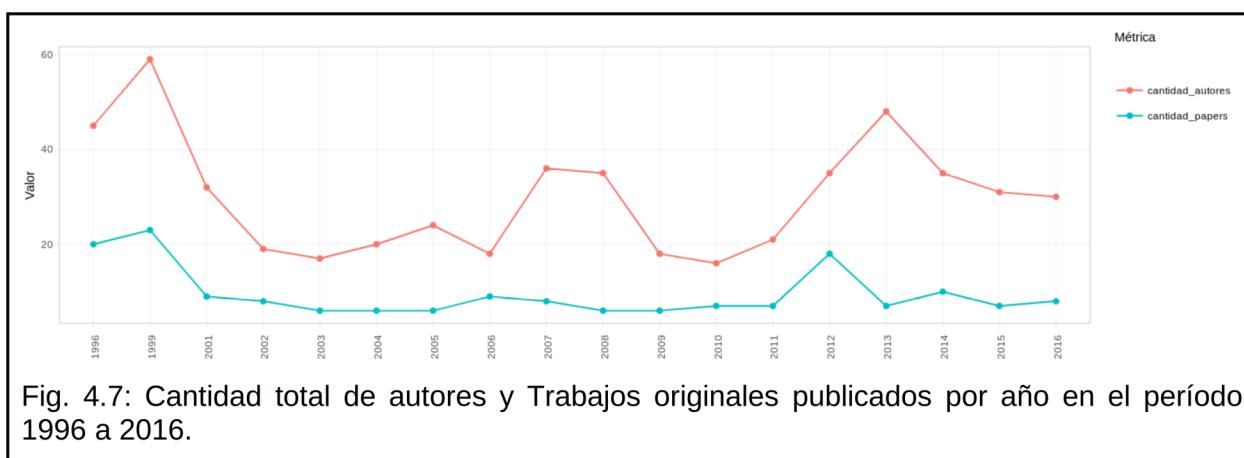
1380

1381 Cuando se evaluó la composición por secciones (es decir, tipo de publicación) se observó que
 1382 esta diferencia en la cantidad total de publicaciones en el año 2005 se debe a la inclusión de
 1383 los resúmenes de los trabajos presentados en las Jornadas Nacionales de Antropología
 1384 Biológica ([Fig. 4.6](#)).
 1385



1386

1387 Esta contextualización inicial puso en evidencia la complejidad de incorporar distintos tipos de
 1388 publicaciones al análisis. Por lo tanto, y retomando lo planteado originalmente, en esta tesis
 1389 analizamos la cantidad de autores y artículos de tipo Trabajo original, en el período de tiempo
 1390 acotado entre 1996 y 2016 ([Fig. 4.7](#)).
 1391



1392

1393 Observamos que en la etapa inicial la RAAB tuvo un fuerte impulso, tanto en cantidad de
 1394 autores como artículos. Sin embargo, el patrón general no es el mismo para ambas medidas, lo
 1395 cual amerita su análisis por separado.

1396 Por un lado, con respecto al número total de Trabajos originales, se observó que a partir del
 1397 2001 se alcanzó un valor que corresponde a la media para la RAAB, y que se mantuvo sin

1398 grandes variaciones en los años que siguieron ([Fig. 4.7](#)). Este patrón es de esperarse en
 1399 revistas científicas puesto que, por lo general, el número de publicaciones que se admiten en
 1400 cada edición está pautado previamente y se mantiene constante o sin mayores variaciones a lo
 1401 largo de los años. En este sentido, el crecimiento de una revista no implica necesariamente un
 1402 aumento en la cantidad de artículos publicados, sino que el mantenimiento de la cantidad de
 1403 publicaciones a lo largo de los años también puede considerarse como síntoma de “buena
 1404 salud”. Contrariamente, la disminución en la cantidad de artículos sí podría vincularse a
 1405 períodos de menor actividad de la revista y/o de la disciplina asociada.

1406 Distinto es lo que ocurrió con la cantidad de autores. En este caso, descartando el impulso
 1407 inicial de los dos primeros años, se observó un aumento en líneas generales pero con la
 1408 existencia de picos en determinados años, no necesariamente asociados a un incremento en el
 1409 número de publicaciones ([Fig. 4.7](#)). Por ejemplo, en 2013 hubo un fuerte aumento en la
 1410 cantidad de autores, quedando como segundo año con mayor cantidad de autores para
 1411 trabajos originales, luego del impulso inicial. Esto podría estar asociado al incremento en la
 1412 cantidad de artículos originales publicados en 2012, que corresponde al tercer lugar en la línea
 1413 de vida de la RAAB para este tipo de publicaciones. Sin embargo, otro aumento en la cantidad
 1414 de autores se observó en los años 2007 y 2008 pero sin asociarse a un incremento marcado en
 1415 el número de publicaciones ([Fig. 4.7](#)). Este aspecto del análisis se profundizará en el siguiente
 1416 apartado.

1417 Ahora que comprendemos mejor el contexto anual de los artículos, y su punto de vista
 1418 individual de cada instante queda revisar cómo se comportan los valores acumulados de la
 1419 cantidad de autores a lo largo del tiempo ([Fig. 4.8](#)).
 1420

1421

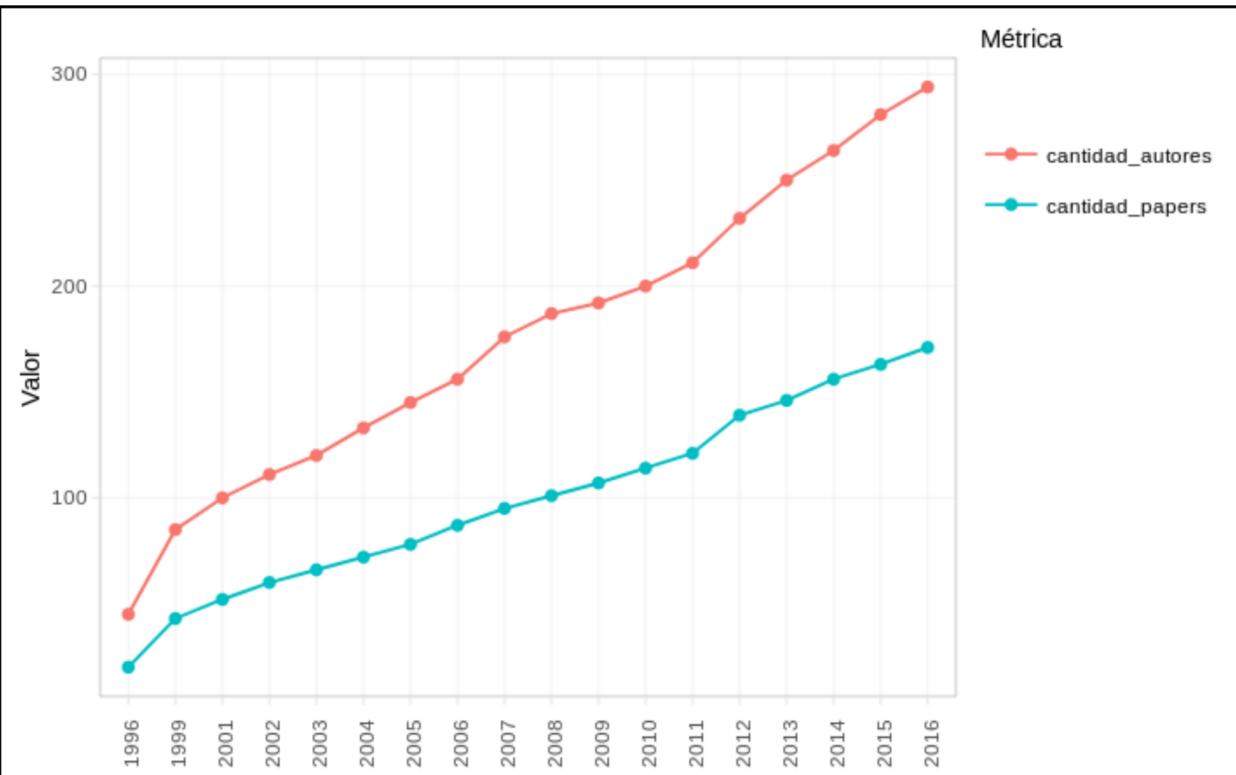


Fig. 4.8: Cantidad acumulada de autores y trabajos originales en el período de 1996 a 2016.

1422

1423 Podemos ver las diferencias mencionadas anteriormente acerca del ritmo de crecimiento en la

1424 cantidad de autores y de artículos. En particular, para la cantidad de autores vemos que existen
 1425 ciertos ciclos de cambios en la pendiente de crecimiento acumulado, asociados a los picos
 1426 señalados para los años 1999 (impulso inicial), 2007 y 2012-2013. En todos los casos, estos
 1427 incrementos en la pendiente parecen sucederse por períodos de aumento relativamente
 1428 constante ([Fig. 4.8](#)). Observamos, además, que los aumentos en la cantidad de autores son
 1429 mayores y más variables que en el número de artículos, cuyo incremento acumulado se
 1430 mantiene relativamente constante por períodos mayores de tiempo. En función de esto,
 1431 intuimos que estas diferencias podrían deberse a una mayor cantidad de artículos trabajados
 1432 en conjunto, aspecto que se profundizará en el apartado siguiente.
 1433 Tomando en consideración todo lo anterior, podemos decir que no tenemos evidencias para
 1434 rechazar la hipótesis 1 en cuanto al crecimiento de la cantidad de autores en la RAAB a lo largo
 1435 del tiempo.

1436 Autores en artículos por año

1437 Elaboramos un poco más respecto de los múltiples autores y vemos que en diferentes
 1438 momentos varían los patrones de cantidad de autores y cantidad de artículos, no son
 1439 relaciones lineales, deben existir momentos donde se hacen más artículos con múltiples
 1440 autores y momentos donde la faceta individual prevalece. Una pregunta interesante a realizar
 1441 es si con el pasar del tiempo se reducen los artículos individuales y se fortalecen los artículos
 1442 en equipo. Vemos la [Fig. 4.9](#) :

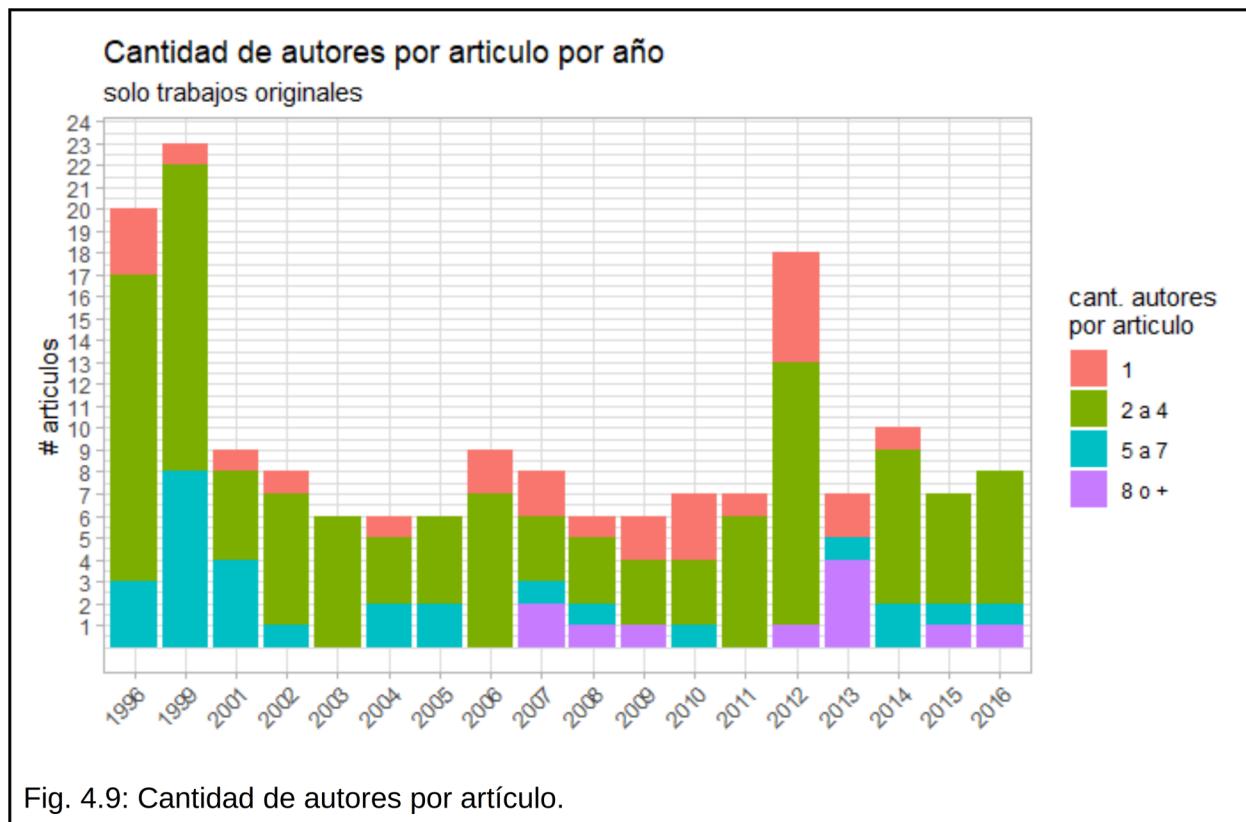


Fig. 4.9: Cantidad de autores por artículo.

1444 Vemos que en general predominan los trabajos originales que son realizados en equipo, entre
 1445 2 y 7 personas. A partir del 2007, vemos que se agregan equipos de 8 o más personas. En
 1446 2015 y 2016, vemos la ausencia de trabajos originales individuales durante más de 1 periodo.
 1447 Valores correspondientes a la [Fig. 4.9](#) en la [Tabla 4.2](#).

1449

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

Tabla 4.2: cantidad de artículos con cantidad de autores por año.

1450

1451 Hipótesis 2: Los autores fundadores se refuerzan en el tiempo, lo
 1452 cual puede ser medido por grado, intermediación y fuerza de
 1453 colaboración

1454 Luego de ver el crecimiento de la revista en cuanto a cantidad de autores y artículos, interesó
 1455 evaluar si ese aumento estuvo dado por nuevos autores o bien, si correspondió a un
 1456 reforzamiento de los autores fundadores. La idea detrás de esta hipótesis es revisar si el
 1457 avance de la revista depende de los autores fundadores, o si está impulsado por la
 1458 incorporación de nuevos autores en grupos de trabajo en principio independientes de los
 1459 fundadores.

1460

1461 Para ello, planteamos otra de nuestras hipótesis respecto de los autores participantes: los
 1462 autores fundadores se refuerzan en el tiempo, lo cual puede ser medido por grado,
 1463 intermediación y fuerza de colaboración. Recordemos que los fundadores son los autores que
 1464 publicaron en la primera edición de la revista.

1465

1466 Cada una de estas métricas nos brindan formas de evaluar los nodos de la red desde
 1467 diferentes puntos de vista. Por otro lado, evaluar la pregunta de si se refuerzan o no en el
 1468 tiempo da diferentes visiones respecto de los roles que cumplen los autores dentro de la red.
 1469 Estos roles pueden ser Concentrador (*Hub*), quién trabaja más con los demás y que genera
 1470 nuevos equipos y es medido por el Grado; Autoridad (*Authorities*), que son generadores de
 1471 contenido, es decir, quienes más aportes hacen en cantidad de trabajos originales, lo cual está
 1472 relacionado con la Fuerza de Colaboración; y finalmente quienes generan puentes entre otros
 1473 autores, quienes tal vez puedan generar nuevas tendencias con mayor facilidad, uniendo
 1474 equipos con los que trabajaron alguna vez o, también, que su desaparición de la red puede
 1475 desconectar equipos, asociados a la distancia de Intermediación. Para profundizar de estos
 1476 roles se puede verificar [Kleinberg \(1999\)](#).

1477

1478 A fin de evaluar estas tres métricas a lo largo del tiempo, se clasificó a los autores para
 1479 ordenarlos en un *ranking* de cinco puestos según cada una de ellas ([Fig. 4.10](#), [Fig. 4.11](#) y [Fig. 4.12](#)). Para interpretar correctamente las siguientes figuras es necesario comprender que las
 1480 líneas de colores están asociadas a la posición en el ranking, esto ayuda a visualizar si los
 1481 autores se mantienen o varían en el tiempo (eje X). Cuando vemos líneas rectas horizontales
 1482 significa que ese autor se mantiene por un tiempo en la misma posición del ranking. Mientras
 1483 que la ausencia de las mismas significa que hay mucha variación y que los autores no se
 1484 mantienen estables en una misma posición a lo largo del tiempo. Por otro lado, las líneas

1486 verticales indican que se comparte el puesto del ranking entre varios autores.

1487

1488 De manera complementaria se presentan tablas, para verlo desde un punto de vista numérico
 1489 (Tabla 4.3, [Tabla 4.4](#) y [Tabla 4.5](#)), se puede ver una tabla a lo largo del tiempo de los primeros
 1490 5 puestos. Se toman en cuenta 3 indicadores por cada puesto: **mantiene** con valor 0 para no y
 1491 1 para sí, respecto de si el autor se mantiene en el puesto respecto del tiempo anterior.
 1492 **es_fundador**, para representar si el autor estuvo en la primera edición de la revista o no, con
 1493 valor 0 para representar que NO es de la primera edición, y valor 1 para representar que sí es
 1494 de la primera edición. Finalmente la variable **valor**, que representa el valor obtenido de la
 1495 métrica analizada para el puesto en ese año.

1496 En la última columna se muestra **ratio_fundador** que representa del total de los años
 1497 analizados, cuánto corresponde a si fue fundador o no. Para esta métrica, un valor de 1
 1498 equivale a 100 %, significa que en todos los años evaluados fue un autor fundador el que
 1499 estuvo en esa posición. Si fuese menor significa que hubo momentos en los cuales el autor no
 1500 fue fundador.

1501

1502 A continuación se muestran las figuras correspondientes a cada métrica seguidas por el detalle
 1503 de cada una.

1504

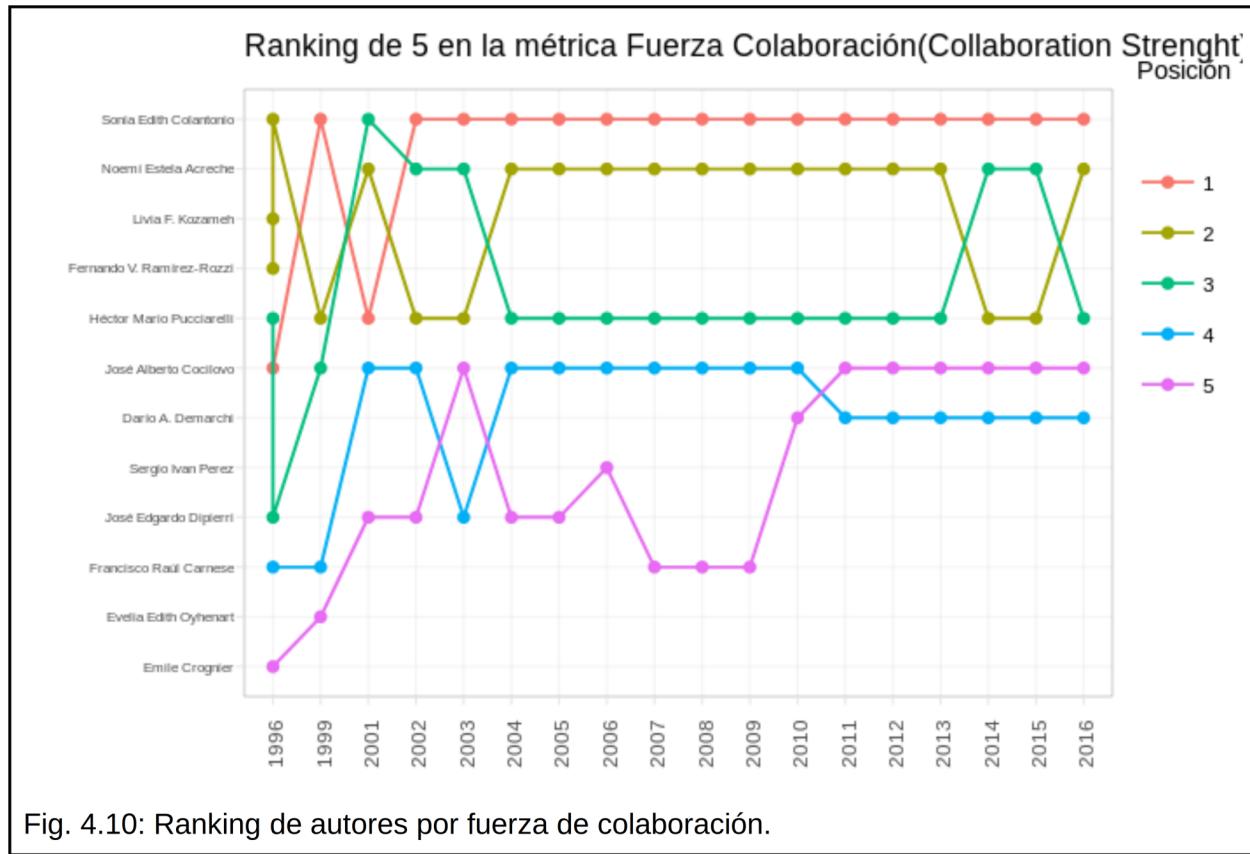


Fig. 4.10: Ranking de autores por fuerza de colaboración.

1505

1506 Recordemos que la métrica fuerza de colaboración es artículo / cantidad Autores. Esto significa
 1507 que el ranking de la métrica representa quienes mejor están posicionados respecto a
 1508 producción de contenido de trabajos originales. Si en un año se ven dos puntos ([Fig. 4.10](#)),
 1509 significa que esos autores 'empataron' en cuanto a fuerza de colaboración. Vemos que los
 1510 autores fundadores se refuerzan en el tiempo y compiten por sus posiciones en los primeros
 1511 tres puestos. Los puestos 4 y 5 son más variables, pero a partir de 2011 se mantienen.

1512 Por lo tanto, vemos que respecto a la métrica fuerza de colaboración, la generación de
 1513 contenido se mantiene bastante estable a partir del 2002. Como análisis complementario
 1514 quedaría pendiente ver la variación interanual para evaluar cuánto crecen estos valores año a
 1515 año. Los valores correspondientes a la [Fig. 4.10](#), están en la [Tabla 4.3](#).

1516

1517

1518

métrica	#	variable	1996	1999	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	ratio fundador
Fuerza de colaboración	1	mantiene es_fundador	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.00
		valor	1.3333	2.0000	2.6619	3.3333	3.8333	3.8333	4.3333	4.6667	4.6667	4.6667	4.6667	4.6667	5.3333	5.3333	5.5833	6.2500	6.2500		
	2	mantiene es_fundador	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	0	1	0	1.00
		valor	1.0000	1.9952	2.1667	2.9119	2.9119	3.4167	3.4167	3.4167	3.4167	3.7500	3.7500	4.4500	4.4500	4.4500	4.4500	4.5286	4.5286	4.7833	
	3	mantiene es_fundador	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	0	1	0	1.00
		valor	0.8333	1.9167	2.0000	2.1667	2.1667	2.9119	2.9119	2.9119	2.9119	3.1119	3.1119	3.1119	3.3619	4.1119	4.2786	4.4500	4.4500	4.5286	
	4	mantiene es_fundador	0	1	0	1	0	0	1	1	1	1	1	1	1	0	1	1	1	1	1.00
		valor	0.7833	1.3833	1.9167	1.9167	1.9762	2.1667	2.1667	2.5000	2.5000	2.5000	2.8333	2.8333	3.0833	3.5833	3.5833	3.5833	3.5833	3.5833	
	5	mantiene es_fundador	0	0	0	1	0	0	1	0	0	1	1	0	0	1	1	1	1	1	0.94
		valor	0.6167	1.3429	1.7262	1.7262	1.9167	1.9762	2.1429	2.3095	2.4111	2.4111	2.5361	2.5833	2.8333	2.8333	2.8333	2.8333	3.1667	3.1667	

Tabla 4.3: Valores del ranking de Fuerza de Colaboración. Análisis de si se mantiene y si es fundador.

1519

1520 Para Fuerza de Colaboración, se aprecia que salvo el quinto puesto del ranking, se mantienen
 1521 los fundadores en el tiempo; y en el quinto puesto el 94% del tiempo son autores fundadores.
 1522

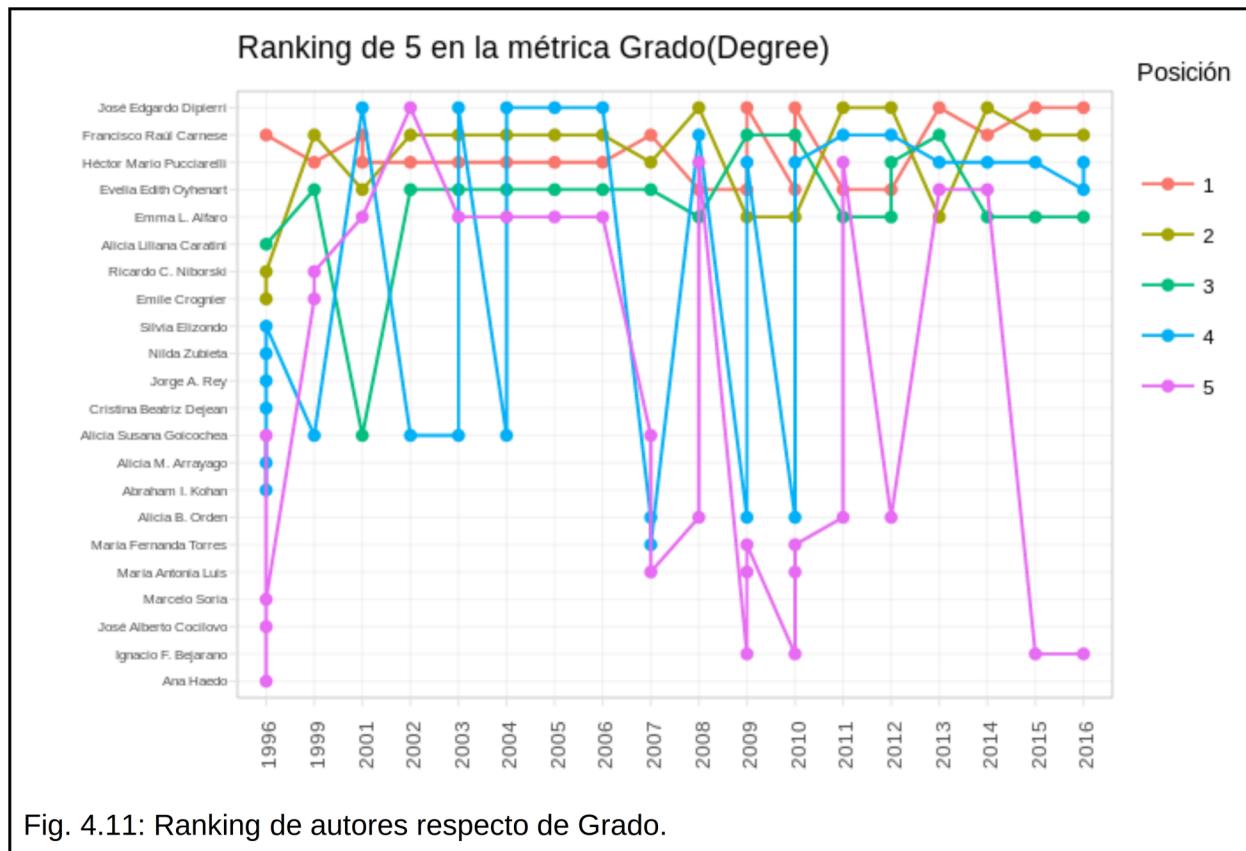


Fig. 4.11: Ranking de autores respecto de Grado.

1523

1524 En la [Fig. 4.11](#) se observa la métrica Grado que indica cuánto se conecta directamente este
 1525 nodo con otros nodos. En el contexto de red de coautoría hace referencia a con cuántos otros

1526 autores publicó ese año. Este ranking es competitivo, anualmente vemos más empates que en
 1527 el de fuerza de colaboración ([Fig. 4.10](#)). Con respecto a los autores fundadores, vemos que los
 1528 primeros cuatro puestos compiten bastante, entre 2002 y 2006 estuvieron estables, mientras
 1529 que los puestos 4 y 5 tienden a variar más.

1530

1531 Es interesante ver cómo algunos nodos evolucionan en el tiempo. Por ejemplo, el primer nodo
 1532 aparece en 2001 en el puesto 4, luego en 2002 baja al puesto 5, en 2003 vuelve al 4 y se
 1533 mantiene hasta 2006. En 2007, no aparece. Luego en 2008 aparece con más fuerza, saltando
 1534 al puesto 2, en 2009 y 2010, escaló hasta el primer puesto y en los años siguientes, se
 1535 mantiene entre el puesto 1 y el 2. Esto significa que, en el contexto de la revista analizada,
 1536 dicho autor es un gran generador de colaboraciones e integrador de nuevos nodos a la red. Lo
 1537 cual tiene gran valor al impulsar el interés por la revista dentro de la disciplina en el país.

1538

1539 Como análisis complementario se podría incorporar la variación interanual para ver cuánto
 1540 crecen estos valores año a año. Recordemos que el Grado es la cantidad de nuevos nexos que
 1541 genera el autor (son nuevos porque no estamos considerando fortalecimiento de una relación
 1542 por hacer más trabajos en conjunto, para eso está la Fuerza de Colaboración).

1543

metrica	#	variable	1996	1999	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	ratio fundador
			mantiene	0	0	1	1	1	1	1	0	0	1	1	1	1	0	0	0		
grado	1	es_fundador	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.00
	2	es_fundador	9	21	15	24	24	24	26	26	27	33	33	33	34	34	42	43	44	44	1.00
	3	es_fundador	6	13	13	15	15	15	18	18	23	32	32	32	33	33	39	42	42	42	1.00
	4	es_fundador	5	10	12	13	13	13	16	16	18	31	30	30	32	32	38	40	40	40	0.83
	5	es_fundador	4	9	11	12	12	12	15	15	17	30	29	29	30	30	37	37	39	39	0.94

Tabla 4.4: Valores del ranking de Grado. Análisis de si se mantiene y si es fundador.

1544

1545 Para Grado, se aprecia que salvo el cuarto y quinto puestos del ranking, se mantienen los
 1546 fundadores en el tiempo; y en el cuarto puesto el 83% del tiempo son fundadores y el 94% del
 1547 tiempo son fundadores para el quinto puesto ([Tabla 4.4](#)).

1548

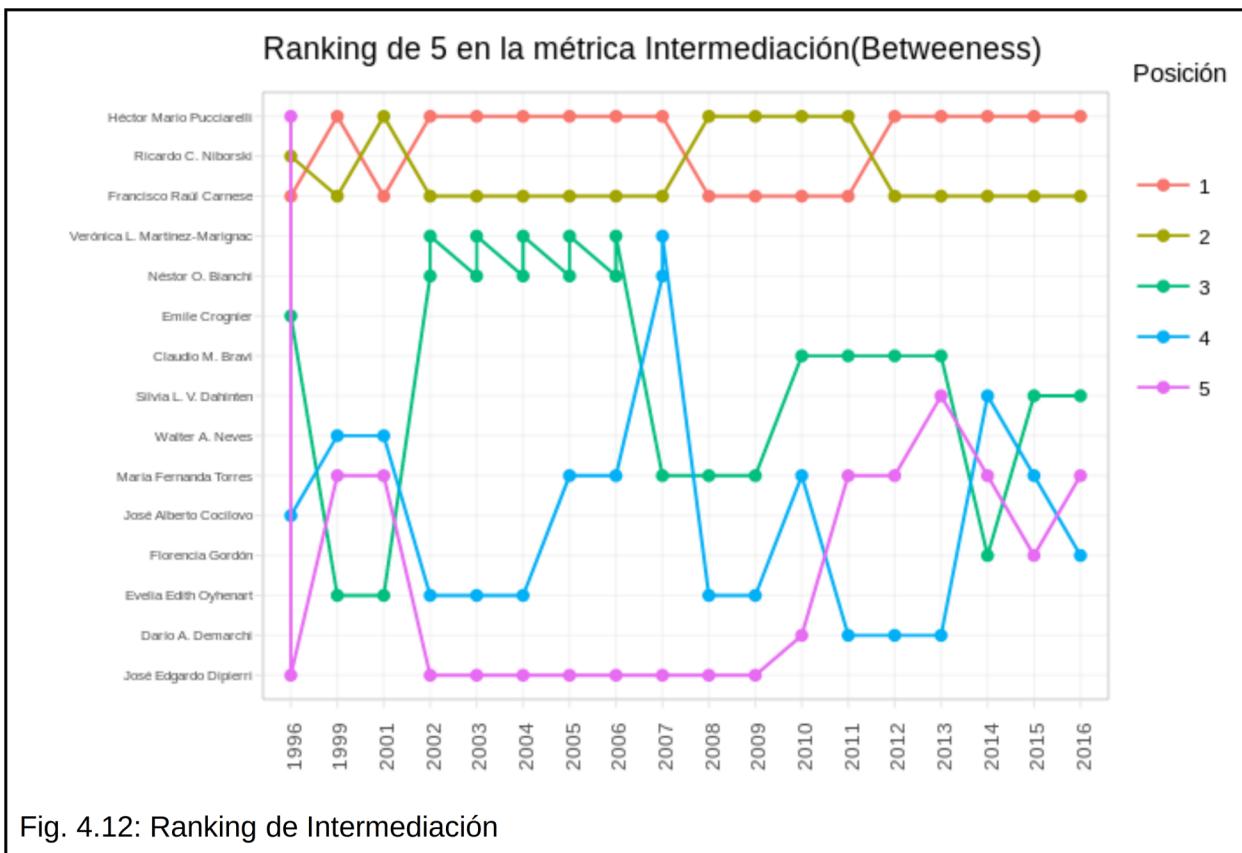


Fig. 4.12: Ranking de Intermediación

1549

1550 La métrica de Intermediación mide la importancia de un autor según su participación en
 1551 caminos para llegar de un autor a otro. En la [Fig.4.12](#) vemos que hay más empates para esta
 1552 métrica que en comparación a Fuerza de Colaboración, pero menos que en el caso de Grado.
 1553 Los primeros dos puestos se mantienen bastante estables sin embargo los 3, 4, y 5, varían
 1554 bastante. Respecto de los autores fundadores, se mantienen los primeros 2 puestos mientras
 1555 que el resto cambia. Llama la atención como el puesto 5 se mantiene estable durante el
 1556 periodo de 2002 a 2009.

1557

métrica	#	variable	1996	1999	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	ratio fundador
intermediación	1	mantiene es_fundador	0	0	0	0	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1.00
		valor	26.500	658.500	736.018	1601.177	1656.177	1691.292	1853.668	1885.902	2272.349	1990.606	2087.649	2438.818	2522.497	2777.240	4124.605	6355.776	6614.722	7059.152	
		mantiene es_fundador	0	0	0	0	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1.00
	2	mantiene es_fundador	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1.00
		valor	12.533	591.979	721.052	1064.907	1085.463	1069.504	1297.290	1309.977	1777.675	1398.516	1452.230	2161.094	2225.381	2715.160	2663.442	3595.673	3728.957	3891.051	
		mantiene es_fundador	0	0	1	0	1	1	1	1	1	0	1	1	1	1	0	1	1	1	0.33
	3	mantiene es_fundador	0	0	1	0	1	1	1	1	1	1	0	1	1	1	0	1	0	1	0.33
		valor	10.333	210.500	259.614	381.683	411.163	418.163	542.122	550.622	723.950	916.141	1027.498	1368.792	1502.125	1985.458	2554.467	2223.364	2435.435	2503.777	
		mantiene es_fundador	0	0	1	0	1	1	1	1	1	0	0	1	1	1	1	0	1	0	0.72
	4	mantiene es_fundador	0	0	1	0	1	1	1	1	1	0	0	1	1	1	1	1	0	1	0.72
		valor	5.000	158.667	180.000	291.614	293.614	300.500	482.103	540.090	635.622	727.105	758.956	1141.215	1236.000	1749.000	2184.000	2223.058	2357.307	2444.122	
		mantiene es_fundador	0	0	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0	0.89	
	5	mantiene es_fundador	0	0	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0.89
		valor	2.000	145.794	154.595	259.465	292.673	297.212	424.005	430.032	490.295	490.767	553.594	1096.000	1179.789	1252.436	2029.277	2111.820	2332.303	2401.666	

Tabla 4.5: Valores del ranking de intermediación. Análisis de si se mantiene y si es fundador.

1558

1559 Para Intermediación, el primero y segundo puesto son estables respecto de los fundadores. Sin
 1560 embargo, el tercero puesto, tiene un 33% del tiempo autores fundadores. El cuarto puesto tiene
 1561 autores fundadores un 72% del tiempo, y el quinto puesto tiene autores fundadores un 89% del

1562 tiempo ([Tabla 4.5](#)).

1563

1564 Vemos que en las tres figuras hay muchos nombres que varían según la métrica considerada,
1565 lo cual significa que diferentes autores han cumplido diferentes funciones en la RAAB dentro
1566 del período analizado. En los ranking que involucran Grado e Intermediación se comparten más
1567 autores, sin embargo en Fuerza de Colaboración varían.

1568

1569 Estas métricas son contextuales en cada tiempo de evaluación y respecto a la red en dicho
1570 momento, por lo tanto mantenerse entre los primeros 5 significa que se refuerzan en el tiempo,
1571 como puede ser validado en los números de cada métrica.

1572

1573 Respecto de la hipótesis acerca de que los fundadores se refuerzan en el tiempo, los
1574 resultados indican que no habría evidencias para rechazarla dado que mayormente se
1575 mantienen los autores fundadores entre los primeros 5 puestos de todas las métricas a lo largo
1576 del tiempo.

1577 **Hipótesis 3: Los componentes más grandes absorben a los más
1578 pequeños a lo largo del tiempo, dado que autores periféricos
1579 pasan a trabajar en conjunto con autores principales**

1580

1581 Resultó de interés contestar esta hipótesis dado que representa la integración de individuos o
1582 equipos a la comunidad local mayormente conectada, que es el subgrafo representado por el
1583 componente mayor. Recordemos que un componente es un grupo de nodos que están
1584 conectados entre sí y está desconectado de otros grupos. Los autores individuales son
1585 componentes en sí mismos. Cuando existe más de un componente se califica a la red total
1586 como una red inconexa, lo cual es natural en redes de coautoría dado que dentro de una
1587 comunidad científica, existen autores individuales y grupos que todavía no realizaron
1588 colaboraciones a través de ninguno de sus autores, o que publican en otros lugares además de
1589 la revista analizada.

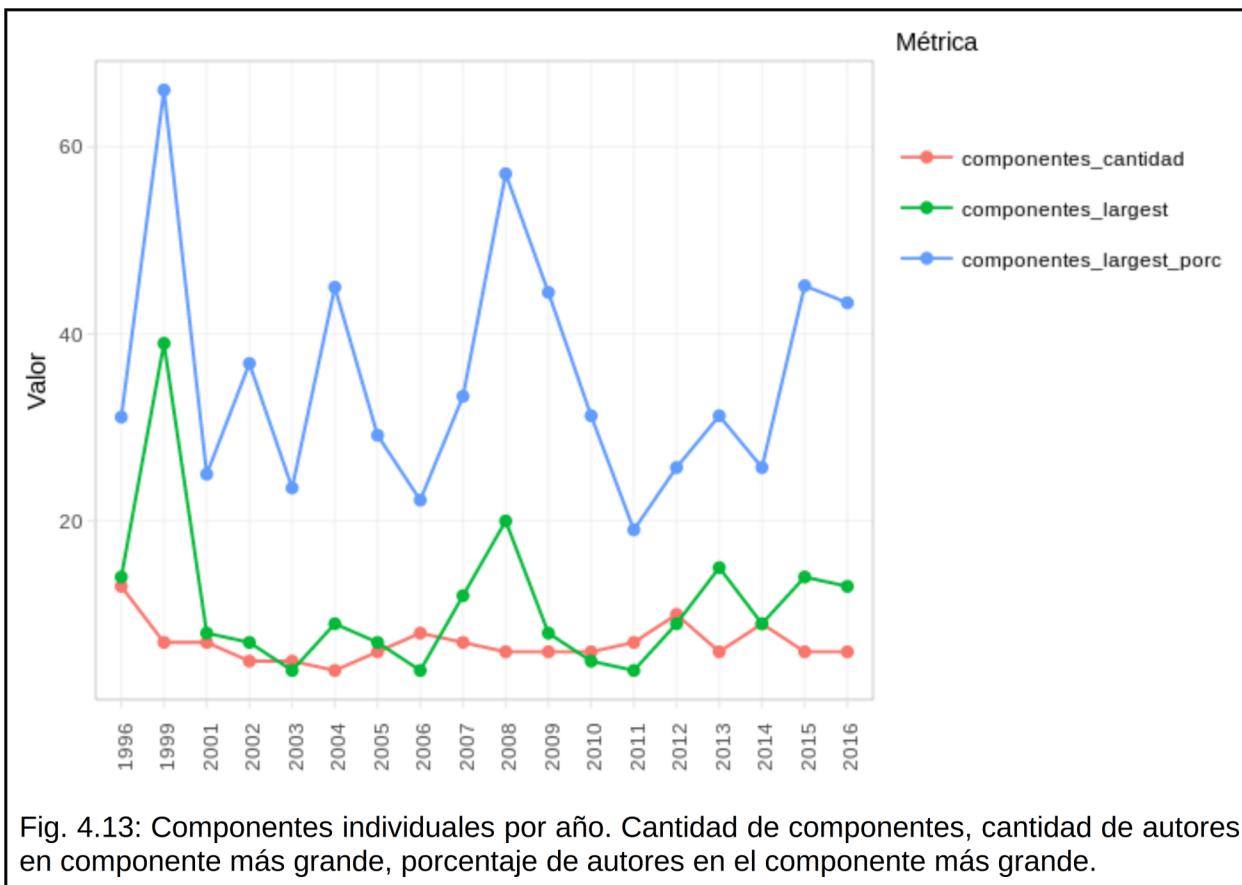
1590

1591 En cuanto al orden, esta hipótesis fue planteada en tercer lugar porque necesitamos un
1592 contexto previo para realizar este análisis; luego de ver la cantidad de autores por artículos por
1593 año, es interesante ver cómo se van generando los componentes, cómo se relacionan entre
1594 ellos y su integración en el tiempo.

1595

1596 Como vimos anteriormente, podemos realizar el análisis desde el punto de vista anual o
1597 agregado obteniendo diferente información respecto de su evolución temporal.

1598

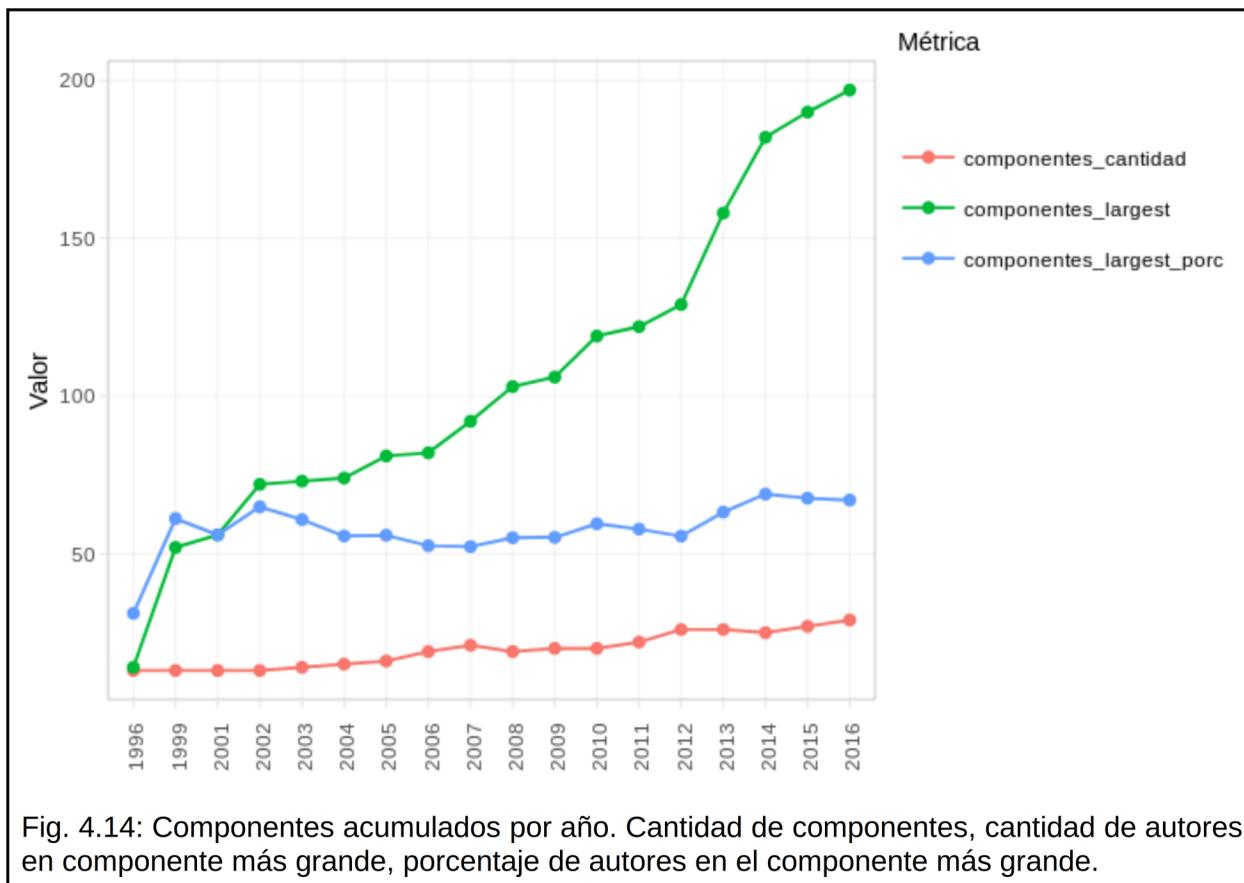


1599

1600 Desde un punto de vista anual ([Fig. 4.13](#)), vemos que en 1999 tuvo su pico de participación de
1601 autores involucrados en el componente principal para una misma edición. Observando ese año
1602 en el gráfico de autores por artículo ([Fig. 4.9](#)), vemos que es el año con mayor cantidad de
1603 trabajos originales presentados, de estos su mayoría artículos entre 2 y 7 autores. Volviendo a
1604 los componentes, vemos que hay un segundo pico en 2008. Si vemos el detalle de
1605 composición mencionado en la hipótesis 1, esto es resultado de un crecimiento en la cantidad
1606 de autores y cantidad de artículos, que podría estar relacionado con que en ese año se publicó
1607 un número especial. Se observa que fue el 3er número de mayor publicación en la historia de la
1608 RAAB, no se tenían esas cantidades desde los números iniciales del 1996 y 1999. En cuanto a
1609 cantidad de artículos es similar a la de los primeros números, solo que existe un artículo de 8 o
1610 más personas, más artículos individuales, y no hay artículos entre 5 y 7 personas.
1611

1611

1612 Este punto de vista nos da una noción de cuando la red tiene incorporaciones de autores y
1613 cómo se comportan respecto de realizar trabajos en conjunto de manera anual. Por otro lado, el
1614 acumulado para ver cómo queda la red total agregada en su evolución temporal, nos brinda
1615 información acerca de cuándo se realizan absorciones importantes respecto del componente
1616 mayor.
1617



1618

1619 A diferencia del gráfico anterior ([Fig. 4.13](#)) podemos ver los saltos de cantidad de autores en el
1620 componente mayor. Estos saltos significan que un componente grande fue absorbido por el
1621 componente mayor. Lo podemos ver en las transiciones de 1996 a 1999, de 2012 a 2013 y de
1622 2013 a 2014 ([Fig. 4.14](#)).

1623

1624 Teniendo en cuenta las observaciones de 2012, se ve que alguno de los componentes
1625 originados, fue absorbido en el 2013 al realizar trabajos de colaboración. A continuación vamos
1626 a verificar estas supuestas absorciones a través de revisar las transiciones estructurales de los
1627 grafos de forma visual.

1628 Verificación de absorción de componentes en la red

1629 Veamos la red en estos instantes para verificar ([Fig. 4.15](#)). En la aplicación se pueden ver de
1630 forma dinámica las transiciones.

1631

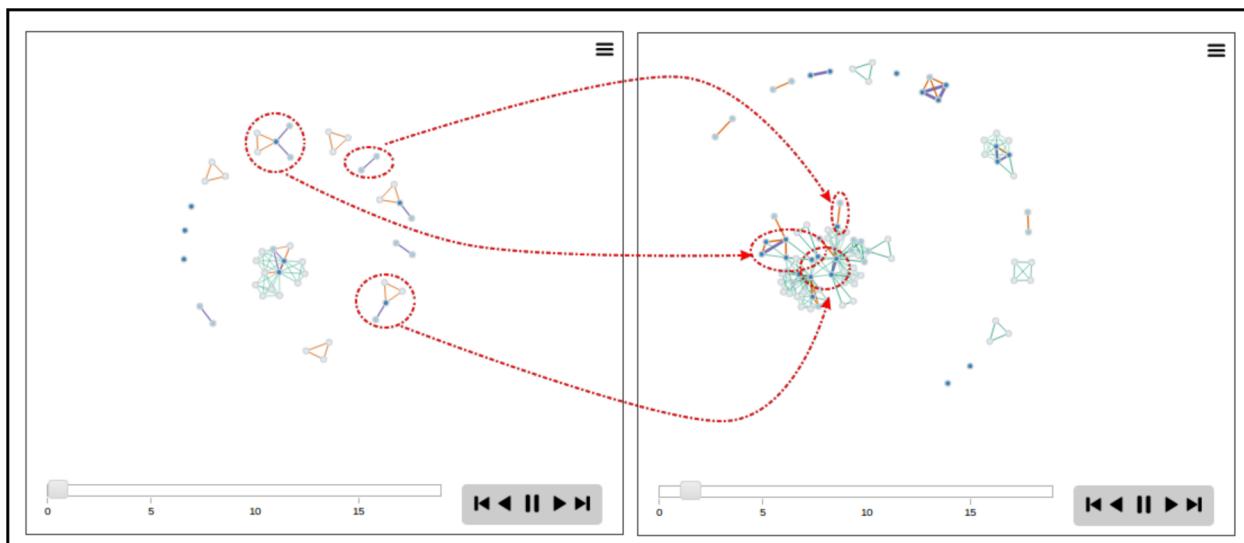


Fig. 4.15: 1996 a 1999, los círculos son los componentes de forma aproximada. Las flechas señalan la transición.

1632

1633 Vemos que en la primera transición el componente mayor absorbe 2 componentes muy
 1634 importantes del tiempo anterior de 5 y 4 autores. Y un tercer componente de 2 nodos. Además
 1635 aparecen nuevos nodos en el nuevo año haciendo que pase de 14 a 52 autores en 1999
 1636 (crecimiento de 38 autores).

1637

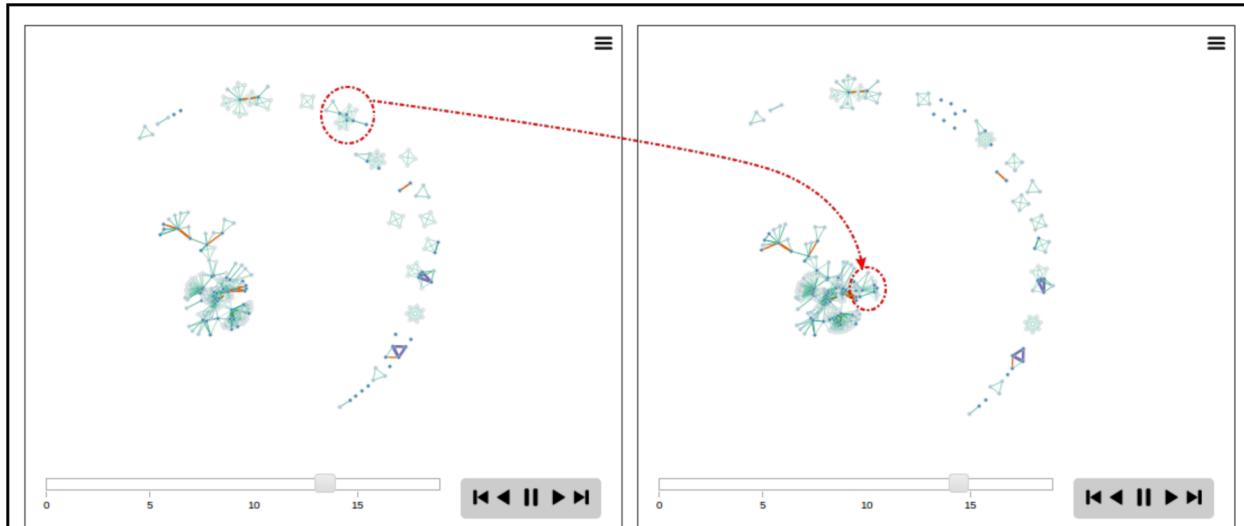


Fig. 4.16: 2012 a 2013, los círculos son los componentes de forma aproximada. Las flechas señalan la transición.

1638

1639 Para el período siguiente ([Fig. 4.16](#)) vemos que uno de los componentes más grandes (12
 1640 autores) es absorbido por el componente mayor. De forma complementaria se agregan
 1641 directamente nuevos nodos al componente mayor pasando de 129 a 158 autores en 2013
 1642 haciendo un crecimiento de 29 autores.

1643

1644

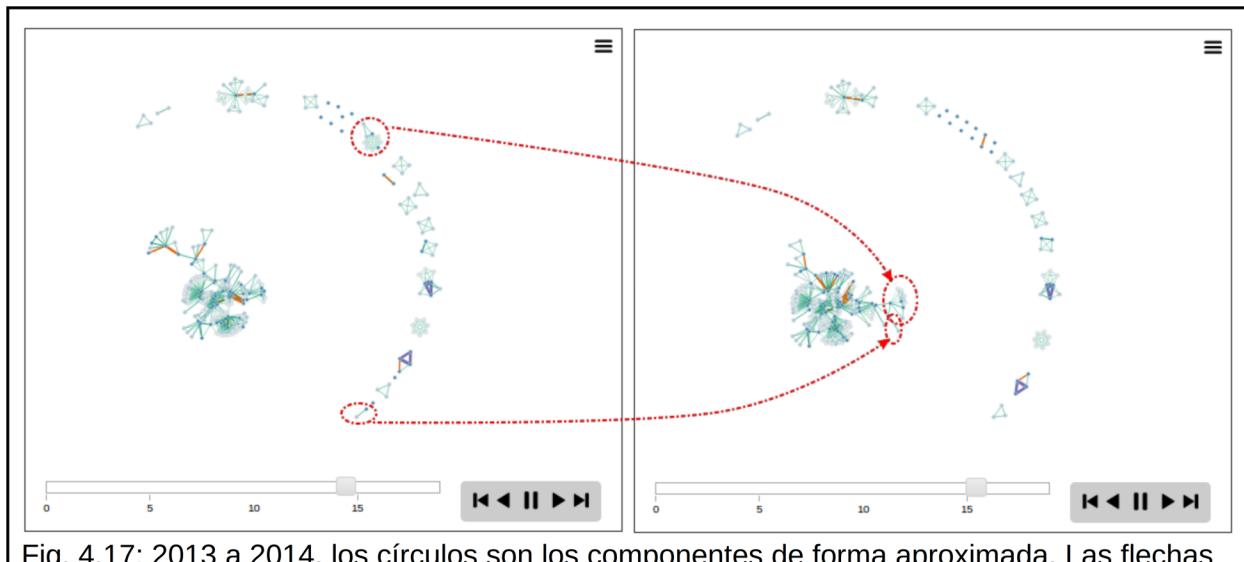


Fig. 4.17: 2013 a 2014, los círculos son los componentes de forma aproximada. Las flechas señalan la transición.

1645

1646 Nuevamente, en el período de 2013 a 2014 ([Fig. 4.17](#)), vemos que uno de los componentes
1647 más grandes (9 autores) es absorbido por el componente mayor. De forma complementaria
1648 también absorbe un componente menor de 2 autores además se agregan directamente nuevos
1649 nodos al componente mayor pasando de 158 a 182 en 2014 haciendo un crecimiento de 24
1650 autores.

1651

1652 Para terminar de revisar la hipótesis vamos a revisar numéricamente en la [Tabla 4.6](#).

1653

1654 Tabla 4.6: Cantidad de nodos Autores en componente mayor.

1655 Cantidad total de nodos, Cantidad de nodos en componente mayor en tiempo inicial de
1656 comparación (t0), Cantidad de nodos en componente mayor en tiempo final de comparación
1657 (t1), porcentaje que representa en (t1) , Cantidad de nuevos nodos en el componente, Cantidad
1658 de nodos absorbidos por el componente entre t0 y t1. Entendemos por absorbidos, nodos que
1659 ya existían en la red, y pasaron a formar parte del componente mayor.

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

1660

1661 De esta forma los resultados son concordantes con la hipótesis de que los componentes
1662 periféricos pasan a ser parte del componente mayor dado que se generan nuevos trabajos
1663 entre los autores de los componentes. Como complemento podemos ver que de los nuevos
1664 nodos que se agregan a este componente mayor, una parte corresponde a nuevos nodos
1665 publicando en la revista, además de los nodos ya existentes.

1666 Relación con actividades de investigación

1667 De manera complementaria se intentó corroborar lo planteado por [Fatt et al. \(2010\)](#):
 1668 *El componente gigante podría significar la principal actividad de investigación; mientras que los*
 1669 *otros componentes pueden ser agrupamientos especializados o sub-comunidades.*

1670
 1671 Este análisis se realizó sobre la red total. Para ello, revisamos los títulos de los artículos
 1672 creados por los 5 componentes con mayor cantidad de autores. Los resultados se pueden ver
 1673 en la [Tabla 4.7](#).

1674
 1675 Tabla 4.7: Artículos, autores, y palabras de mayor frecuencia en los títulos. Para los primeros 5
 1676 componentes.

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, étnicos, ethnic, congenitas, congenital, anomalies, anomalías, america

1677
 1678 En las siguientes figuras vemos la frecuencias de las palabras obtenidas por componente para
 1679 los primeros tres puesto que, los grupos 4 y 5 solo tienen publicado un artículo, lo que lleva a
 1680 que las palabras que figuran son solo las que aparecen en dicho artículo ([Tabla 4.7](#)). Por lo
 1681 tanto, estos últimos dos componentes quedaron fuera del análisis de frecuencia de palabras.
 1682 Por otro lado, consideraremos también que el componente principal tiene 114 artículos, mientras
 1683 que el segundo tiene solo 10 artículos, y el tercero 11 lo cual repercute en las frecuencias de
 1684 palabras de las figuras.
 1685

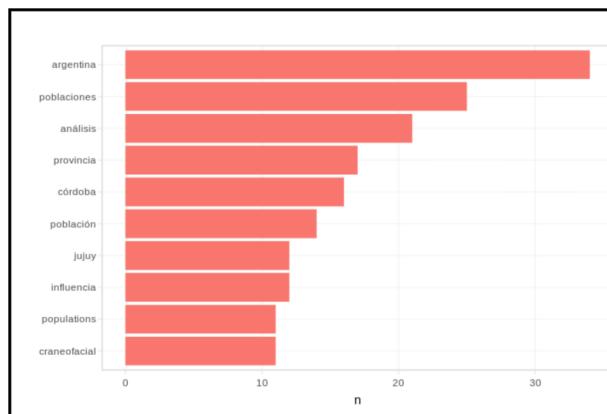


Fig. 4.18: Componente 1

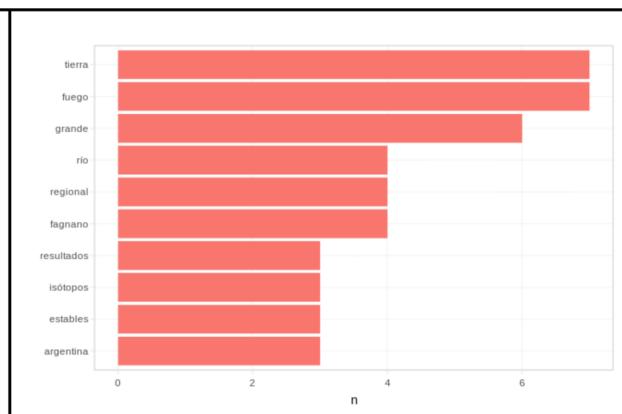


Fig. 4.19: Componente 2

1686
 1687 Lo que llama la atención del 1er componente ([Fig. 4.18](#)) es que se destacan algunas
 1688 localidades, como Córdoba y Jujuy, a pesar de ser un componente muy heterogéneo a nivel

1689 geográfico. Sin embargo, la presencia del término “craneofacial”, estaría indicando que este
 1690 componente tiene una relación estrecha con esa línea de investigación en varias regiones del
 1691 país. Vemos también que aparecen palabras en inglés, como “populations”.

1692
 1693 Con respecto al segundo componente ([Fig. 4.19](#)), puede verse que es bien local de Tierra del
 1694 Fuego (“tierra del fuego”, “río grande”), aunque también parece asociado a una línea de
 1695 investigación por la presencia de las palabras “isótopos” y “estables”.
 1696

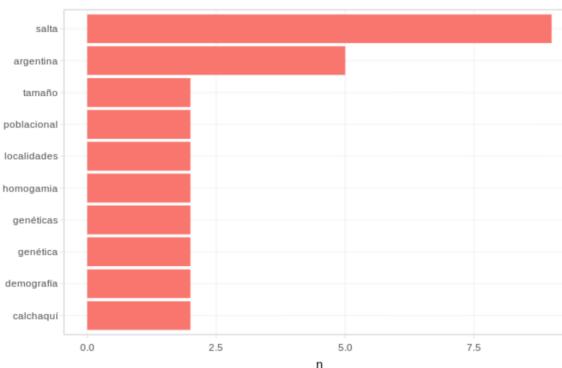


Fig. 4.20: Componente 3

1697
 1698 En el componente 3 ([Fig. 4.20](#)) también vemos que se destaca otra localidad, Salta, haciendo
 1699 que lo geográfico sea el factor común, mientras que parecería ser heterogéneo en términos de
 1700 las líneas de investigación que incluye.
 1701

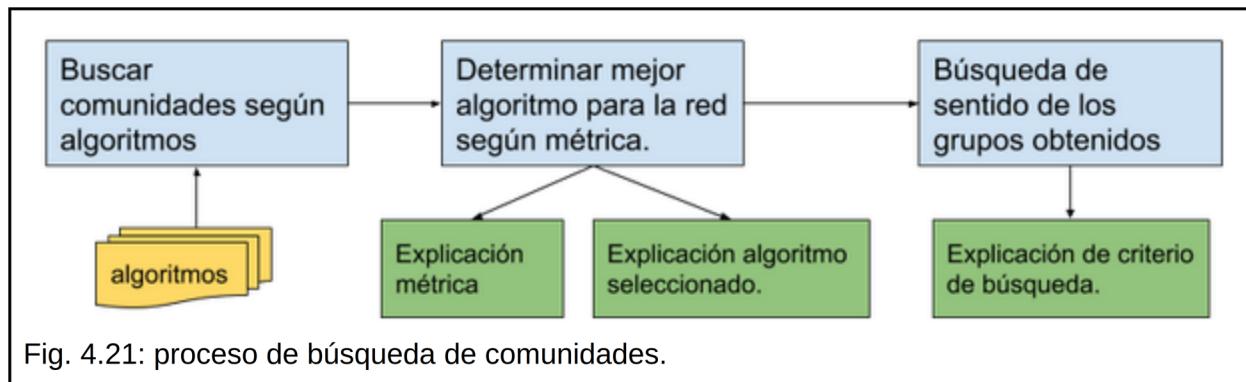
1702 Vemos que los resultados obtenidos pueden interpretarse acorde al postulado de que el
 1703 componente principal abarca la principal actividad de investigación, mientras que los
 1704 secundarios abarcan temas más especializados ([Fatt et al. 2010](#)). Aún teniendo en cuenta la
 1705 cota de 10 palabras en el análisis realizado, puede verse que, en términos geográficos, el
 1706 primer componente es más heterogéneo que los dos más pequeños. Además, en relación a las
 1707 actividades de investigación, es probable que la línea que surge del componente más grande
 1708 (asociada a la palabra “craneofacial”) sea la más frecuente dentro de la disciplina a nivel
 1709 nacional, lo cual repercute en una mayor cantidad de artículos publicados en la RAAB. Queda
 1710 profundizar el enfoque para comprender con mayor detalle estos componentes así como su
 1711 composición de temas, sobre todo porque el componente mayor contiene múltiples artículos
 1712 que seguramente no fueron representados en las primeras 10 palabras. De la misma forma
 1713 sería interesante, a futuro, revisar en el tiempo los diferentes componentes y sus temas para
 1714 evaluar si mientras se va armando el componente gigante va absorbiendo las diferentes
 1715 comunidades, enriqueciendo así los temas tratados a nivel revista. También interesa, a futuro,
 1716 explorar estas temáticas en un análisis de texto más complejo, teniendo en cuenta, por
 1717 ejemplo, los conceptos compuestos por varias palabras (como ‘genética poblacional’ o ‘tierra
 1718 del fuego’).

1719 Búsqueda de comunidades

1720 Además del análisis de componentes que representa cómo se van uniendo a través de
 1721 trabajos en conjunto, se realizó una búsqueda de comunidades resultantes según diferentes
 1722 algoritmos. Estos algoritmos buscan según diferentes criterios que suelen ser útiles en la
 1723 búsqueda de patrones en redes. Para determinar que algoritmo aplica mejor para la red
 1724 estudiada se midieron a través del criterio de modularidad. De esta forma resta interpretar el

1725 resultado para el caso de la RAAB, ver qué comunidades surgen y si tienen cierto sustento
 1726 temático como ocurrió con los componentes.

1727
 1728 En primer lugar, verificamos si encontramos agrupamientos bajo algún criterio determinado por
 1729 los algoritmos disponibles y también lo complementamos con análisis de comunidades dentro
 1730 de la red ([Fig. 4.21](#)).
 1731



- 1733 Explicación de la métrica de evaluación de algoritmos: Modularidad
- 1734 Para la selección del algoritmo utilizamos la métrica de modularidad. Esta métrica fue
1735 propuesta por [Newman-Girvan \(2004\)](#), la implementación utilizada es la que se encuentra en el
1736 el paquete igraph, que utiliza la versión de [Clauset, Newman y Moore \(2004\)](#).
- 1737
- 1738 Siguiendo a [Barabási \(2016\)](#) puede comprenderse mejor la métrica de la siguiente forma:
- 1739
- 1740 Para calcular la modularidad de una red con N nodos y conexiones L , particionados en N_c
1741 comunidades; y cada una de las comunidades tiene N_{ci} nodos y L_{ci} conexiones, donde c es el
1742 numero de comunidad 1 a N .
- 1743
- 1744 Si el número de L_{ci} es mayor que el número de L esperado para la secuencia de grado de la
1745 red, entonces los N_{ci} podrían ser parte de una comunidad. Entonces para cada comunidad se
1746 calcula la diferencia entre la matriz de conexión de la red actual(A_{ij}), y la matriz de conexión de
1747 los mismos nodos pero conectados aleatoriamente(p_{ij}), preservando el grado para los nodos.
- 1748

$$M_C = \frac{1}{2L} \sum_{(i,j) \in C} (A_{ij} - p_{ij})$$

A_{ij} matriz de conexiones actuales,
 P_{ij} matriz conexiones aleatorias
 C comunidad dentro de comunidades
 L conexiones totales de la red
 $\frac{1}{2}$ porque es un grafo no dirigido.

$M_C > 0$: cuanto mayor Modularidad ,mejor el particionamiento.
 $M_C == 0$: todos los nodos en 1 sola partición
 $M_C < 0$: cada nodo como una partición separada.

- 1749
- 1750 Por lo tanto, cuanto mayor sea la Modularidad significa que la configuración de conexiones
1751 actual es mejor que una conexión aleatoria y mejor que otras configuraciones.
- 1752
- 1753 En la [Tabla 4.8](#) podemos ver la comparación de los algoritmos de búsqueda de comunidades
1754 en la red agregada total (periodo 1996 a 2016).
- 1755
- 1756 Tabla 4.8: Comparación de los algoritmos utilizados para la búsqueda de comunidades.

Algoritmo Utilizado	Cantidad de comunidades	Valor modularidad	Valor modularidad con peso (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 líder (Leading Eigenvector)	35	0.6551	0.6385
Etiquetas Propagates (Propagating Labels)	44	0.6459	0.642

1758

1759

1760 En la [Tabla 4.8](#) podemos ver que el algoritmo mejor posicionado es *Optimización de modularidad multinivel*, sea con o sin considerar los pesos de las aristas. En el segundo puesto 1761 podemos ver que si no consideramos los pesos, *Intermediación de aristas* queda seleccionado, 1762 mientras que si consideramos los pesos, *Búsqueda de comunidades infomap* queda mejor 1763 ranqueado.

1764

1765

1766 Explicación del algoritmo seleccionado: Optimización de modularidad multi-nivel

1767 A continuación se detalla solamente cómo funciona el algoritmo ganador dentro de los 1768 evaluados (para más información de los algoritmos probados ver el [anexo de los algoritmos de 1769 comunidades](#)). Para comprender mejor cómo es que se realizan las divisiones con este 1770 algoritmo analizamos el trabajo de [Blondel \(2008\)](#) y la explicación del libro de [Barabási \(2016\)](#).

1771

1772 El algoritmo tiene N Pases y cada Pase debe cumplir 2 pasos. El paso 1 es de asignar a todos 1773 los nodos a una comunidad propia, y luego analizar cuánto gana en modularidad por 1774 incorporarse a una comunidad vecina. Si supera, se muda de comunidad. Este paso se repite 1775 hasta no poder mejorar la modularidad de forma local ([Fig. 4.22](#)).

1776

1777 Luego el paso 2 es crear un nuevo grafo que va agregar los nodos de cada comunidad, 1778 generando 1 nodo representando a cada comunidad. Las conexiones representan la cantidad 1779 de conexiones con las otras comunidades, y los bucles consigo misma representan todas las 1780 conexiones internas de la comunidad ([Fig. 4.22](#)).

1781

1782 Luego de esto se termina el Pase, arrancando un siguiente Pase en el que se repiten los pasos 1783 1 y 2. El algoritmo termina cuando no se puede aumentar más la modularidad ([Fig. 4.22](#)).

1784

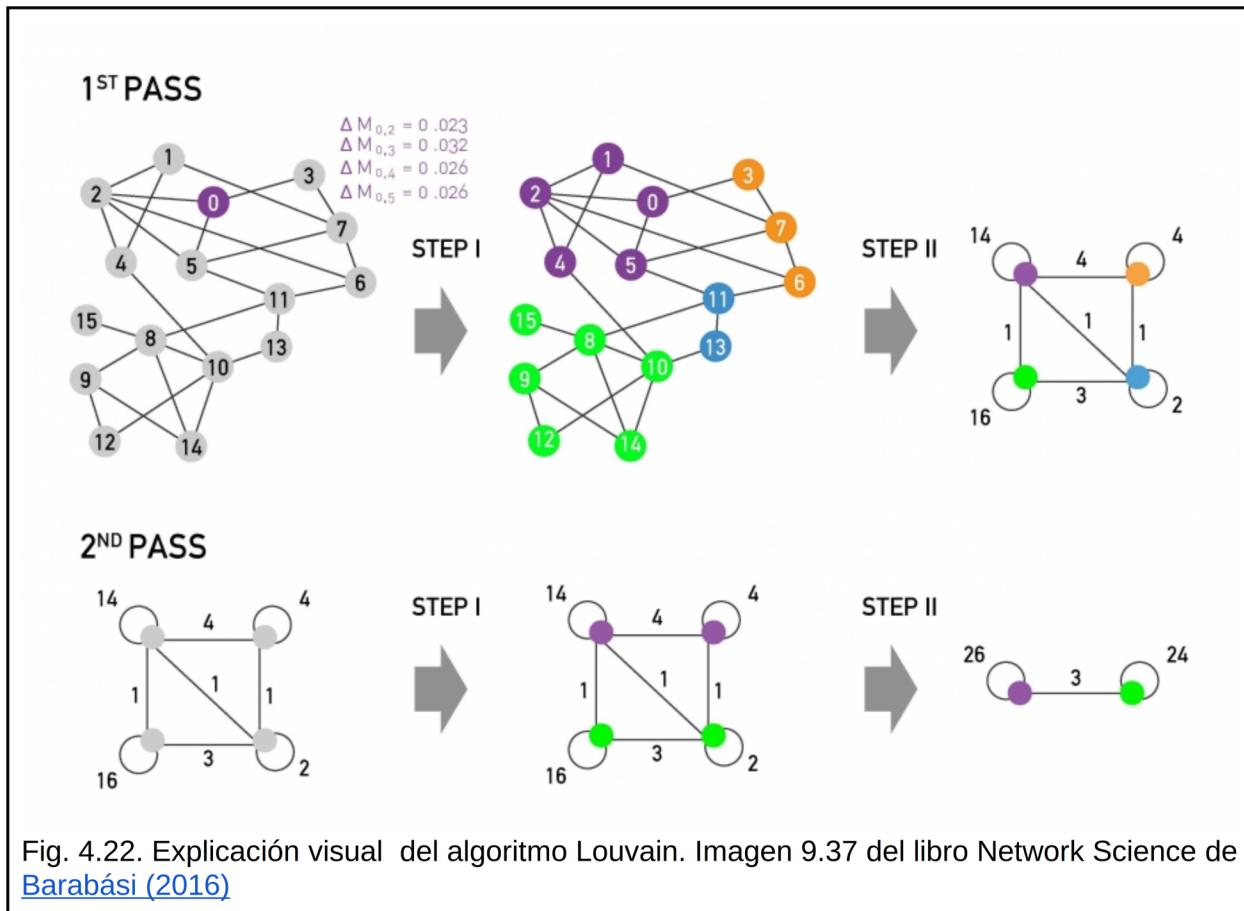
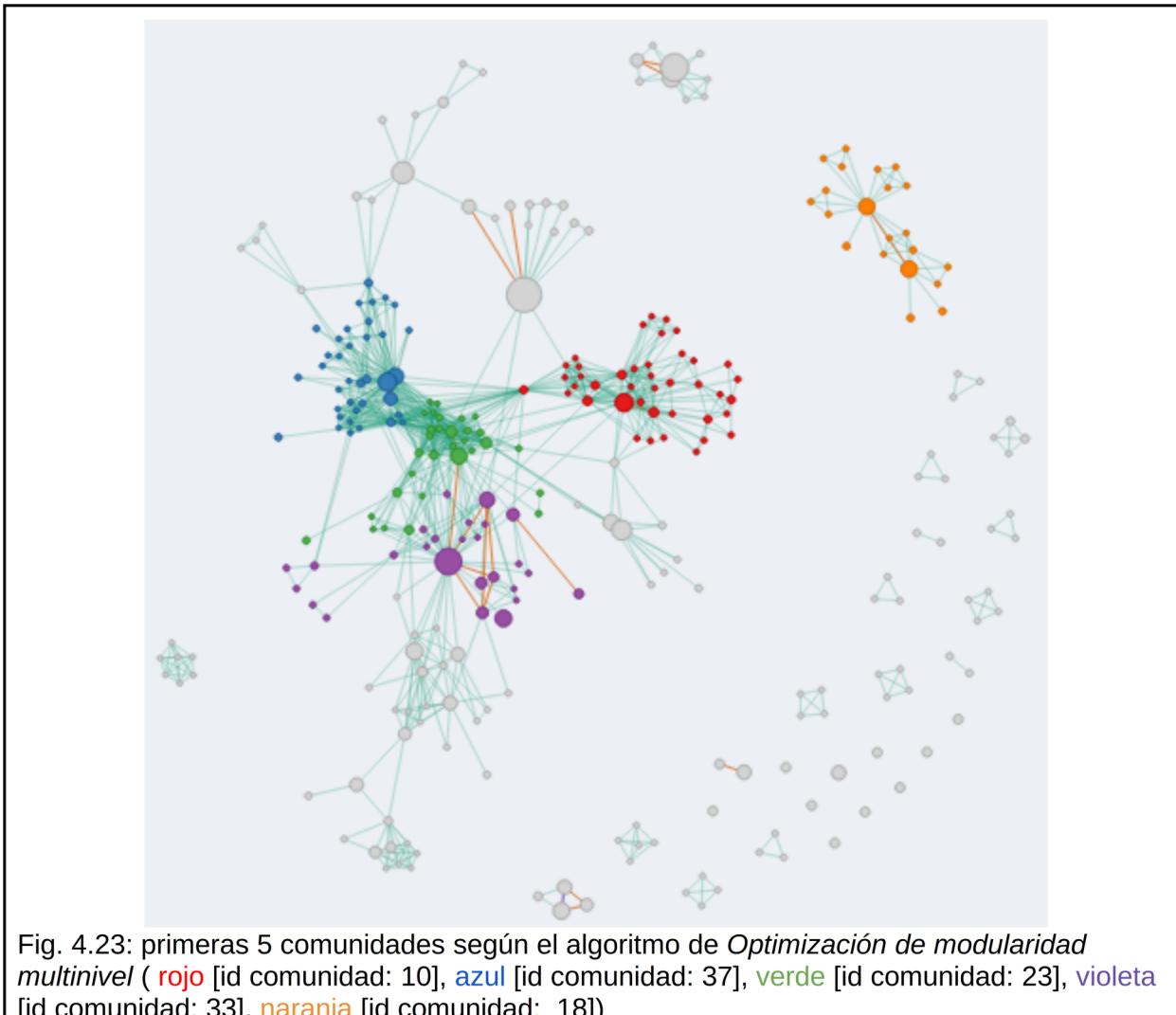


Fig. 4.22. Explicación visual del algoritmo Louvain. Imagen 9.37 del libro Network Science de Barabási (2016)

1785 Búsqueda de sentido en los grupos obtenidos

1786

1787 Estudiando los resultados del algoritmo de *Optimización de modularidad multinivel* podemos
1788 visualizar las primeras 5 comunidades (Fig. 4.23) en las cuales nos vamos a centrar para
1789 constatar un sentido más allá de métricas estructurales. Se trató de encontrar sentido a las
1790 comunidades según la información que se obtiene a partir de centrarse en sus datos, más
1791 específicamente, qué trabajos fueron elaborados por ellas.
1792



1793

1794 Explicación de criterio de búsqueda: verificación de comunidades según títulos
 1795 publicados.

1796

1797 En este análisis nos centramos en las primeras 5 comunidades según cantidad de integrantes.
 1798 Más allá de la búsqueda de la optimización de la modularidad del algoritmo, evaluamos si
 1799 podíamos entender cómo están compuestas analizando los títulos de los artículos generados
 1800 por las comunidades propuestas ([Tabla 4.9](#)). Para ello, realizamos un breve análisis de
 1801 frecuencias de palabras en los artículos publicados, tomando las primeras 10 palabras de cada
 1802 comunidad. La idea de estos análisis es tener un bosquejo de qué temas se tratan en dichas
 1803 comunidades, sin embargo si se encuentra interés, debería servir de inicio para análisis más
 1804 extensos.

1805

1806 Tabla 4.9: Artículos, autores, y palabras de mayor frecuencia en los títulos. Para las primeras 5
 1807 comunidades.

1808

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), infanto (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)

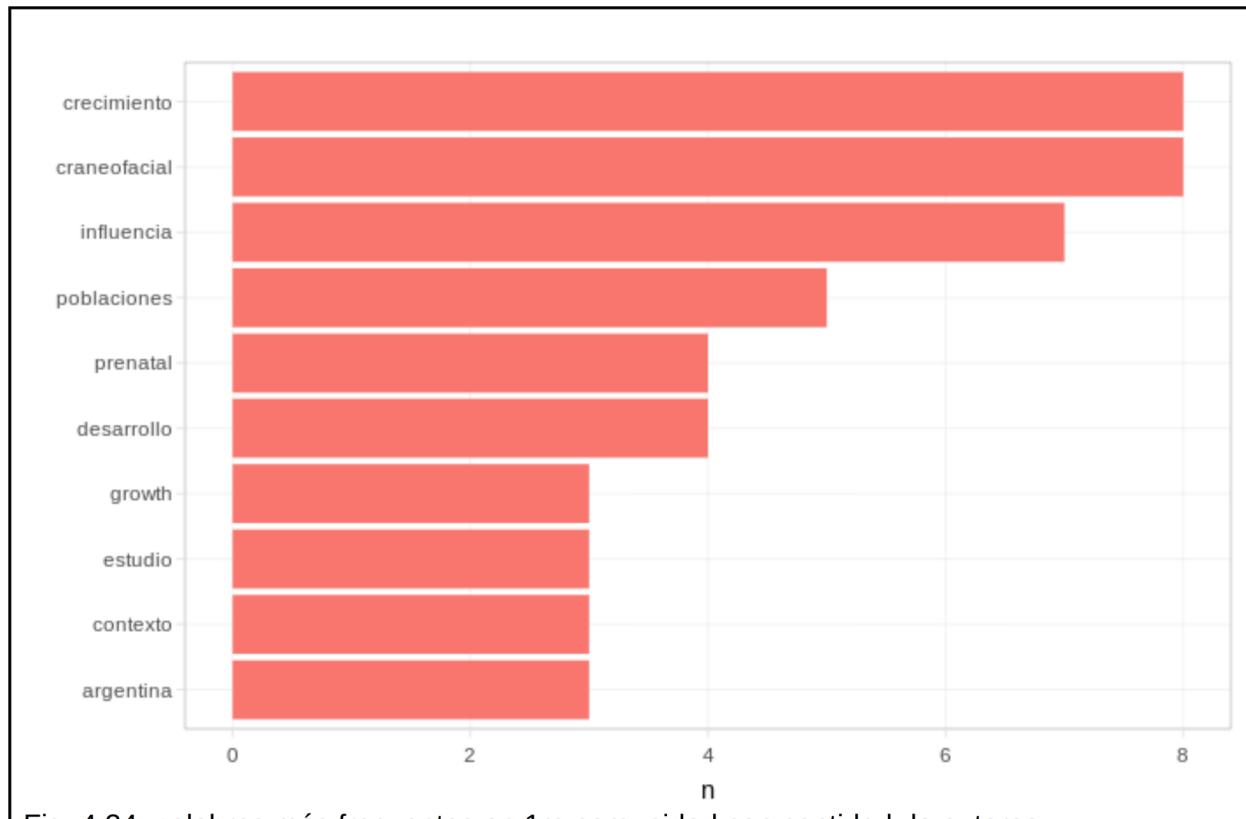


Fig. 4.24: palabras más frecuentes en 1ra comunidad con cantidad de autores.

1811 En la comunidad 1 ([Fig. 4.24](#)) se destaca como principal las palabras “craneofacial”,
 1812 “crecimiento” e “influencia”, y llama la atención que aparecen palabras en inglés. Esto último
 1813 puede deberse a que algunos artículos en la revista poseen sus títulos publicados tanto en
 1814 castellano como inglés, cosa que se repite en tres de las cinco comunidades evaluadas.
 1815 Respecto de una de las palabras que más se destacan, “influencia”, resultó difícil comprender
 1816 la razón. Sin embargo, al leer los títulos de los artículos se observó que muchos de ellos
 1817 arrancan con esa palabra (como ‘*influencia de X factor en crecimiento/nutrición de Y
 1818 población*’), lo cual podría explicar su relativamente alta frecuencia.

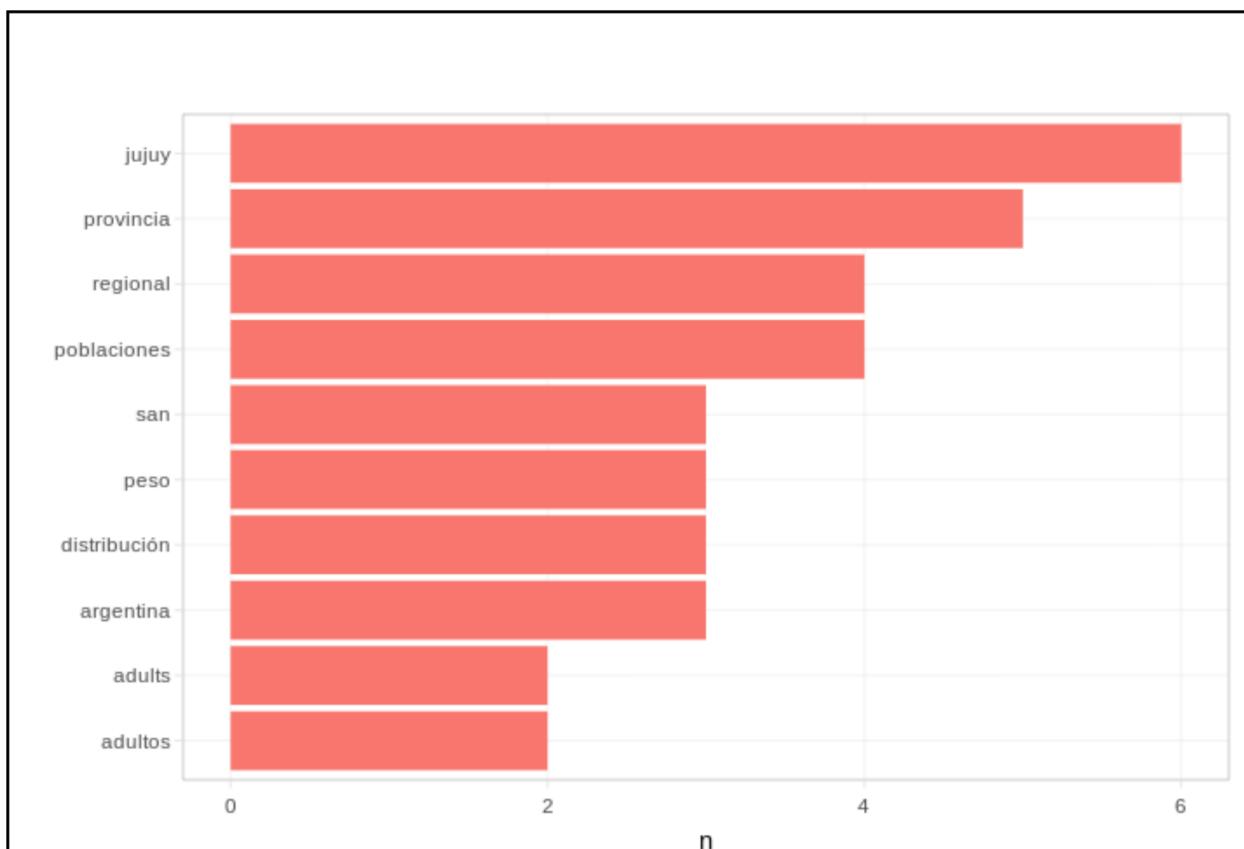


Fig. 4.25: palabras más frecuentes en 2da comunidad con mayor cantidad de autores.

1822 En la comunidad 2 ([Fig. 4.25](#)), los nodos parecen estar vinculados por un componente de
1823 localidad, asociado a la provincia de Jujuy, y por estudios poblacionales sobre dicha provincia.
1824

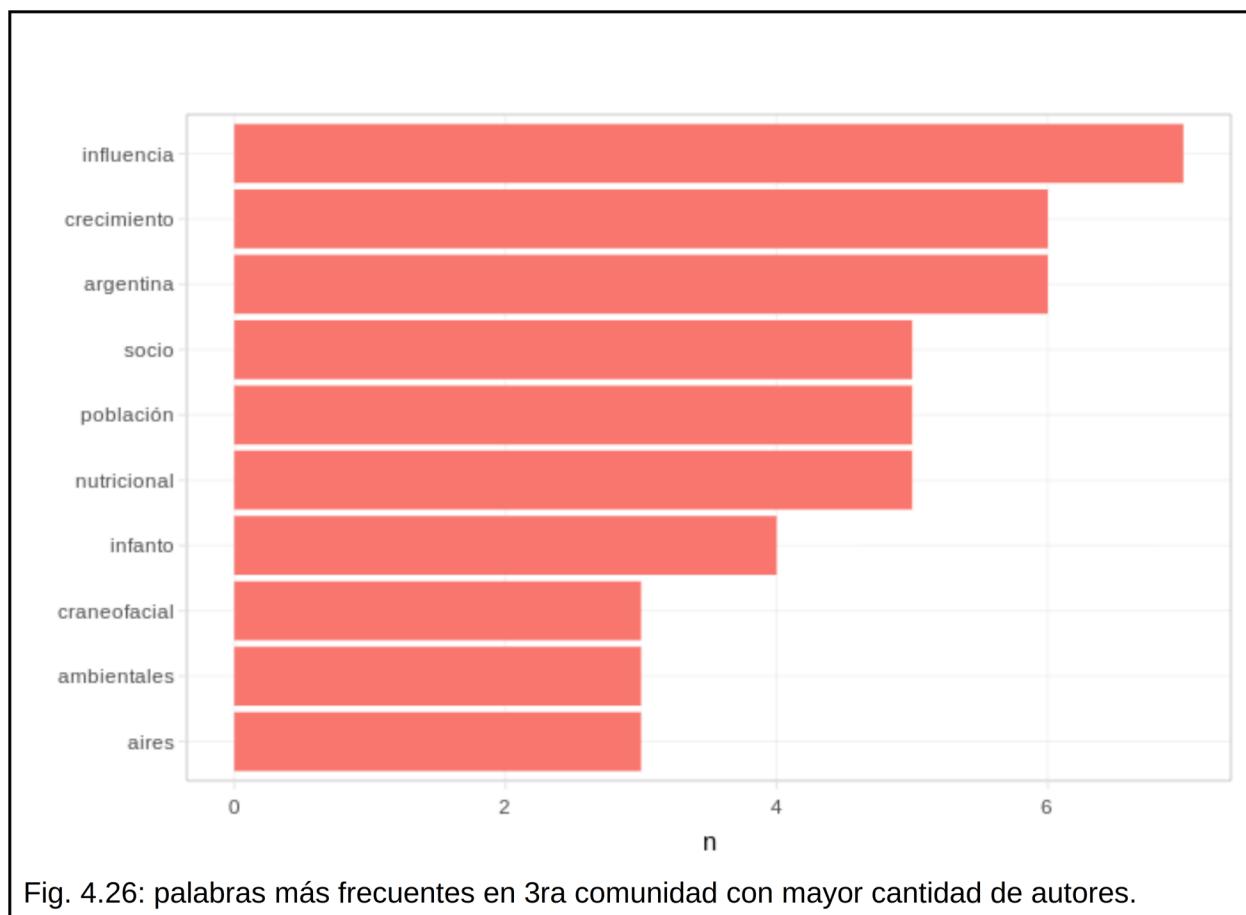


Fig. 4.26: palabras más frecuentes en 3ra comunidad con mayor cantidad de autores.

1825

1826 En la comunidad 3 se resaltan las palabras “influencia”, “argentina”, “crecimiento”, “nutricional”
 1827 y “población” ([Fig. 4.26](#)). Es posible que esta comunidad esté relacionada con el tema análisis
 1828 de poblaciones en Argentina considerando diversos factores que afectan el crecimiento.

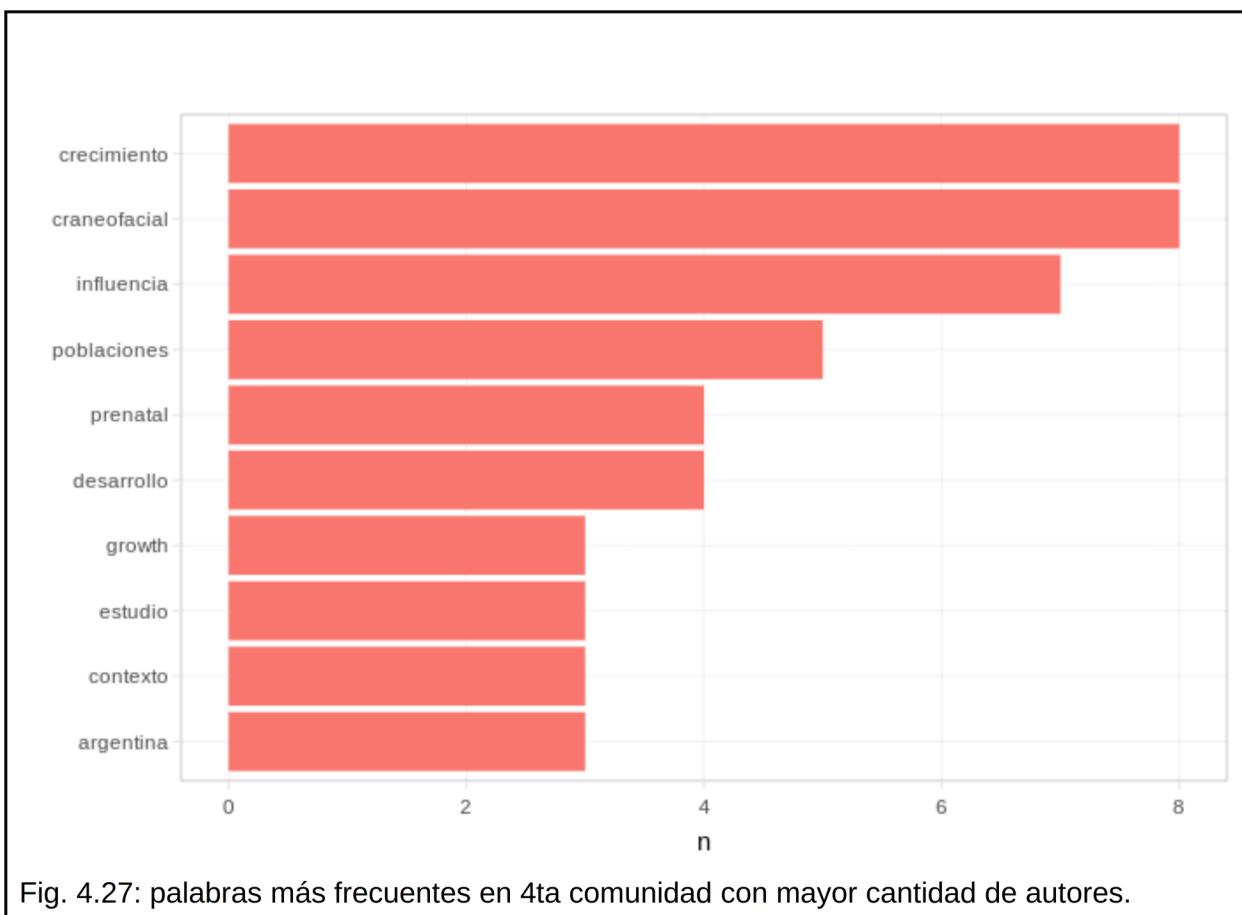


Fig. 4.27: palabras más frecuentes en 4ta comunidad con mayor cantidad de autores.

1829

1830 En la comunidad 4 se observan tanto palabras en inglés como castellano y, nuevamente, se
 1831 resalta la temática “craneofacial”, “crecimiento” y la palabra “influencia” ([Fig. 4.27](#)). Al leer
 1832 algunos títulos se puede ver que los temas desarrollados por esta comunidad se asocian a
 1833 cómo repercuten las dietas en el desarrollo humano. Cabe destacar que los temas asociados a
 1834 las comunidades 3 y 4, según el análisis de las 10 palabras más frecuentes en los títulos,
 1835 poseen similitudes por lo que queda para futuro realizar otro tipo de análisis de texto que
 1836 permita profundizar e identificar mejor las líneas de investigación posiblemente asociadas a las
 1837 comunidades.

1838

1839

1840

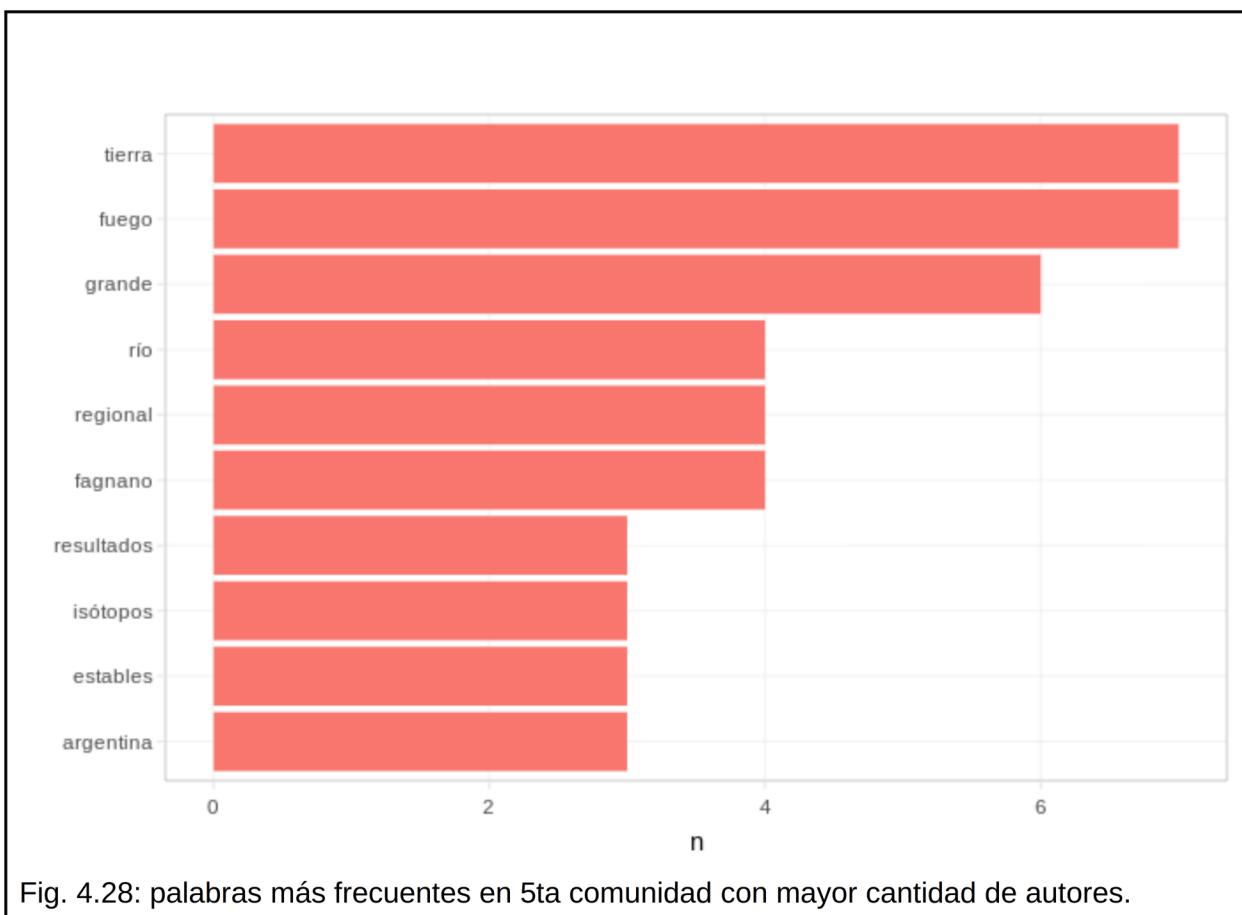


Fig. 4.28: palabras más frecuentes en 5ta comunidad con mayor cantidad de autores.

1841

1842 Finalmente, la comunidad 5 parece estar afectada principalmente por una componente de
1843 localidad asociada a la provincia Tierra del Fuego y, más específicamente, la localidad de Río
1844 Grande ([Fig. 4.28](#)). Se puede observar que la frecuencia de palabras de forma simple no
1845 resultó suficiente para este análisis, que se deben analizar compuestos de palabras, como ser
1846 “tierra del fuego” en su completitud en vez de cada palabra por separado.

1847

1848 Todo este breve análisis realizado sobre las comunidades halladas nos permitió confirmar que,
1849 a futuro, sería necesario implementar análisis de texto más complejos y detallados para poder
1850 identificar correctamente las líneas o actividades de investigación asociadas a cada comunidad.
1851 Sin embargo, a pesar de las limitaciones del análisis realizado, en muchos casos hemos podido
1852 asociar a las comunidades con diferentes temas de investigación y regiones geográficas del
1853 país.

1854

1855 Una vez concluido este análisis de la red “hacia adentro”, nos preguntamos cómo se comporta
1856 “hacia afuera”, es decir, al compararla con otras redes. Para lograr esta comparación en esta
1857 tesis realizamos dos enfoques: a través de modelos para verificar si la red tiene ciertas
1858 propiedades que se observan comúnmente en redes del mundo real, y comparando con otras
1859 redes de coautoría, de estudios previos.

1860

1861 Hipótesis 4: La red agregada total (desde 1996 a 2016) es de
 1862 mundo pequeño puesto que es una característica presente en la
 1863 mayoría de las redes de coautorías científicas

1864
 1865 En los análisis de redes es de utilidad realizar comparaciones con modelos. En palabras de
 1866 [Kolaczyk-Csárdi \(2014\)](#):

En la práctica, los modelos de redes de grafos son usados para diferentes propósitos. Estos incluyen: 1) la prueba de “significancia” de ciertas características predefinidas en una determinada red, 2) el estudio de mecanismos propuestos para generar ciertas propiedades que se observan comúnmente en redes del mundo real (como distribuciones de grado amplias o efectos de mundo pequeño) o 3) la evaluación de potenciales factores predictivos de relaciones.

1868
 1869 En este caso nos vamos a centrar en el punto 2, el estudio de mecanismos propuestos para
 1870 generar ciertas propiedades que se observan comúnmente en redes del mundo real (como
 1871 distribuciones de grado amplias o efectos de mundo pequeño).

1872 Mundo pequeño

1873 El estudio de “mundo pequeño” está asociado a evaluar ciertas propiedades de la red que
 1874 permiten que se pueda afirmar que las personas o nodos se conectan con otras a través de
 1875 una cantidad pequeña de saltos, más allá de que se hallen muy separadas entre sí en la red.

1876
 1877 Para comprender un poco mejor el estudio de “Small world”, surge de la frase en inglés ‘My it's
 1878 a small world’ / ‘Que mundo pequeño’ cuando dos personas conocen a otra y, por lo tanto,
 1879 estarían interconectados a través de una tercera. Este estudio de [Milgram \(1967\)](#) consistió en
 1880 enviar mensajes desde una persona en un punto A hasta otra en un punto B bajo ciertas reglas,
 1881 entre ellas, que solo podían mandar el mensaje a personas conocidas inmediatas.

1882
 1883 Con el pasar del tiempo se siguió analizando esta característica de las redes, en diferentes
 1884 estudios, llegando a medidas para evaluar esta propiedad. Las propiedades clave son un alto
 1885 coeficiente de *clustering*, también llamado transitividad, y un promedio bajo de camino más
 1886 corto.

1887
 1888 El coeficiente de *clustering* de un nodo mide hasta qué punto los vecinos de un nodo también
 1889 están interconectados, mientras que el coeficiente de clustering de la red es el promedio del
 1890 coeficiente de clustering de todos los nodos.

1891 El promedio de camino más corto entre nodos, mide el camino más corto de un nodo A a un
 1892 nodo B. Para esta métrica también se trabaja con el promedio para todos los nodos en la red.

1893
 1894 Para poder evaluar estas propiedades usualmente se realizan comparaciones con otras redes
 1895 de características similares. Como al momento de surgimiento de esta tesis fue difícil encontrar
 1896 redes reales con dichas características, se procedió a realizar simulaciones de modelos
 1897 aleatorios para poder comparar sus propiedades con las de la red analizada y así analizar si
 1898 cumple o no con las características de “mundo pequeño”.

1899 Libre Escala

1900 En forma complementaria evaluamos la propiedad de libre escala, que se basa en analizar la
 1901 distribución de grado de la red. Recordemos que el grado de cada nodo representa con
 1902 cuántos otros nodos se conecta directamente el nodo analizado. Una red Libre escala es una
 1903 red en la cual su distribución de grado sigue la ley de potencia ([Barabási 2016](#)), lo cual significa
 1904 que la red tiene muchos nodos con pocas conexiones y pocos nodos con muchas conexiones.

1905 Caso aplicado

1906

1907 En primer lugar, analizamos la red en su totalidad, de 1996 a 2016, para contrastar la hipótesis
 1908 general y luego por períodos, para verificar si este comportamiento varió en el tiempo.
 1909 Recordemos que los 20 años de la revista están repartidos en 18 años de publicaciones dado
 1910 que en 1997, 1998 y 2000 no hubo ediciones.

1911 Mundo pequeño para la red agregada total

1912 La [Fig. 4.29](#) muestra los parámetros utilizados para realizar las simulaciones usando el modelo
 1913 Erdos-Renyi para el caso de mundo pequeño de la red total entre 1996 y 2016.

1914

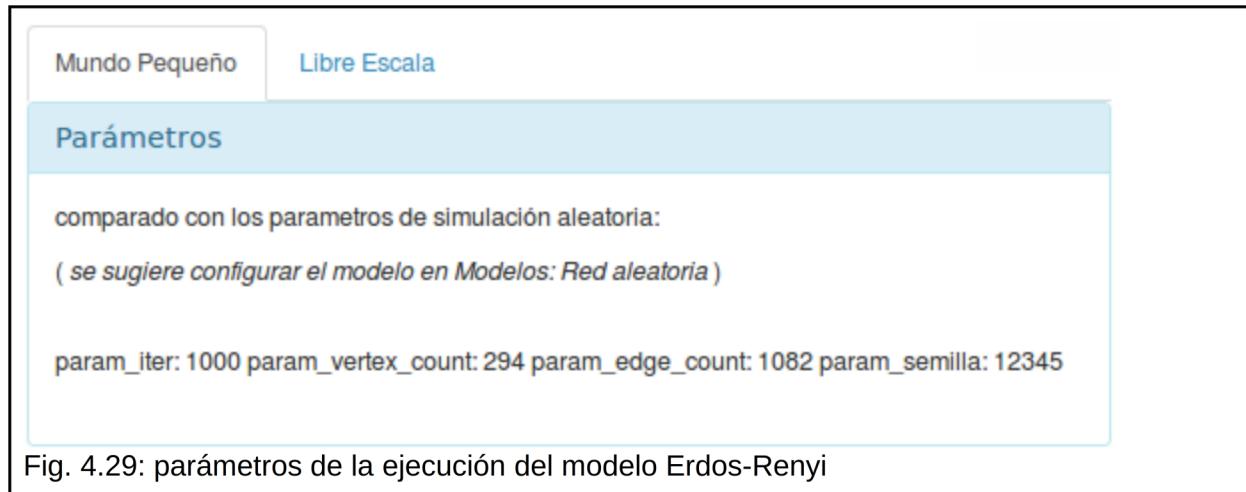


Fig. 4.29: parámetros de la ejecución del modelo Erdos-Renyi

1915

1916 La red total de los períodos desde 1996 hasta 2016, según el criterio de validación expuesto
 1917 por [Bialonski et al. \(2010\)](#) es de mundo pequeño ([Fig. 4.30](#)). Sin embargo, para el criterio de
 1918 validación utilizando la métrica *smallworldness*, no lo es ([Fig. 4.31](#)), tanto para una cota laxa de
 1919 1, como para una cota estricta de 3, según lo planteado por [Humphries-Gurney \(2008\)](#).
 1920 Viendo los resultados comparados ([Fig. 92](#)), vemos que la generación de las simulaciones a
 1921 través de la función *smallworldness* varía considerablemente con lo obtenido en las
 1922 simulaciones creadas bajo el criterio de validación de [Bialonski et al. \(2010\)](#).

<p>Validación 1:</p> <p>para validar small world se tomó el criterio de comparar la red actual contra redes aleatorias (lo seleccionado en el tab de Modelos / Aleatorio).</p> <p>la comparación se lleva a cabo en el promedio del camino mas corto VS la media del promedio del camino más corto de las N redes simuladas.</p> <p>para mas detalles ver: Bialonski (2010) donde se detalla el procedimiento.</p> <p>Mundo pequeño: si Delta ≈ 1 y Gamma > 1</p> <p>Delta = L/Lr. y Gamma = C/Cr.</p> <p>L: promedio camino más corto</p> <p>Lr: promedio de promedio camino más corto de las simulaciones</p> <p>C: transitivity</p> <p>Cr: promedio de transitivity de las simulaciones</p> <p>Delta: 1.11397526759519</p> <p>Gamma: 21.647993978366</p> <p>Es mundo pequeño: Si</p>	<p>Validación 2:</p> <p>Una red puede ser considerada 'mundo pequeño' si la 'smallworldness' es mayor a 1 (igraph::smallworldness)</p> <p>Un punto de vista más estricto dice de la red llevarla a 'smallworldness' ≥ 3 (Humphries & Gurney, 2008)</p> <p>smallworldness: 0.676971834912878</p> <p>mayor a 1: NO</p> <p>mayor a 3: NO</p>
--	---

Fig. 4.30: Criterio de validación 1: es mundo pequeño

Fig. 4.31: Criterio de validación 2: NO es mundo pequeño

1923

1924

<p>Comparación de Validaciones</p> <p>Recordar que para Validación 2, se vuelven a ejecutar simulaciones.</p> <p>Sobre las simulaciones de Validación 1, se puede ver su detalle en su pestaña</p> <p>Sobre las simulaciones de Validación 2, se puede ver su detalle en su pestaña</p> <p>para cada variable: valor validación 1 valor validación 2</p> <p>transitivity [red]: 0.542988831767417 0.542988831767417</p> <p>transitivity [promedio simulaciones]: 0.0250826396344185 0.11267629852863</p> <p>promedio distancia del camino mas corto [red]: 3.41034781280236 161.513895660653</p> <p>promedio distancia del camino más corto [promedio simulaciones]: 3.06142148035701 22.6893559007221</p>
--

Fig. 4.32: Comparación de resultados de los criterios.

1925

1926

1927

1928

1929

1930

1931

1932

1933

1934

1935

1936

Esta diferencia de valores que aparece en la comparación de métodos, está asociada al método de generación de las redes aleatorias. En el criterio 1 ([Fig. 4.30](#)) se usa el método Erdos-Renyi implementado en el método `sample_gnm` de la librería `igraph` que se basa en la publicación de [Erdős-Renyi \(1959\)](#), mientras que en el criterio 2 la función de la librería `ggraph` se utiliza la función para generar los aleatorios de la librería `igraph degree.sequence.game` pasando el método de generación "`simple.no.multiple`".

Los parámetros para el modelo de Erdos-Renyi se pueden ver en la [Fig. 4.29](#). 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. El modelo de Erdos-Renyi se basa en la cantidad de nodos y de aristas del

1937 grafo original, y se hacen 1000 iteraciones aleatorias para obtener los resultados. Por otro lado,
 1938 *degree.sequence.game* se basa en generar un grafo aleatorio dada una lista de grados tomada
 1939 del grafo original, en la implementación de la función *smallworldness*. Es importante aclarar que
 1940 en la documentación de la función *degree.sequence.game*, se recomienda ver otras funciones
 1941 por problemas de bucles que se generan al ejecutarla y, entre ellas, recomienda el modelo de
 1942 Erdos-Renyi basado en probabilidades. Para más información se recomienda ver la
 1943 documentación de la librería [igraph](#).

1944

1945 Como conclusión, podemos decir que la red de coautorías de la RAAB es de mundo pequeño,
 1946 aunque hemos visto que depende de los diferentes enfoques utilizados para analizarla.

1947 Libre escala para la red total

1948

1949 En los análisis realizados se observa claramente que la red de coautorías de la RAAB posee
 1950 muchos nodos con pocas conexiones y pocos nodos con muchas conexiones ([Fig. 4.33](#) y [Fig.](#)
 1951 [4.34](#)). Confirmamos entonces que se da el fenómeno de libre escala.

1952

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: 2.51557126350872

El exponente es mayor a 2: TRUE

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

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

El test de ajuste de Kolmogorov-Smirnov significativo?

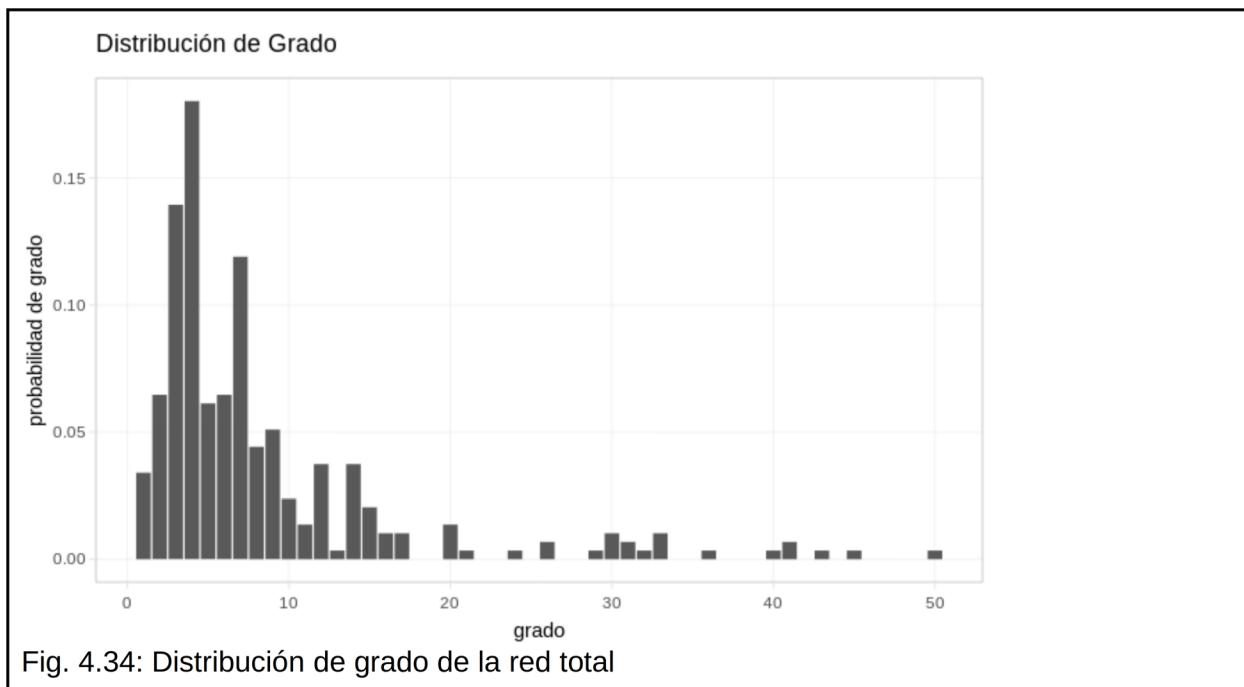
p-value test KS: 0.508187872780151

menor a 0.05: FALSE

Es de Libre Escala.

Fig. 4.33: resultado de comparación con modelos de libre escala. Es de libre escala.

1953



1954

1955

1956 Luego de realizar los análisis para la red total (período completo), evaluamos cómo se comportaba la red en períodos inferiores. Para ello, se utilizó la división especificada en la
 1957 [Tabla 4.10](#). Recordemos que los 20 años de la revista están repartidos en 18 años de
 1958 publicaciones dado que en 1997, 1998 y 2000 no hubo ediciones.

1959

1960

1961

1962

Período	desde	hasta	incluye
1	1996	2001	1996,1999,2001
2	2002	2006	2002,2003,2004,2005,2006
3	2007	2011	2007,2008,2009,2010,2011
4	2012	2016	2012,2013,2014,2015,2016

1963

1964 Mundo pequeño para el periodo 1996 - 2001

1965 Podemos ver que para el periodo 1996-2001, La [Fig. 4.35](#) muestra los parámetros utilizados
 1966 para realizar las simulaciones usando el modelo Erdos-Renyi para el caso de mundo
 1967 pequeño. ([Fig. 4.35](#))

1968

1969

1970

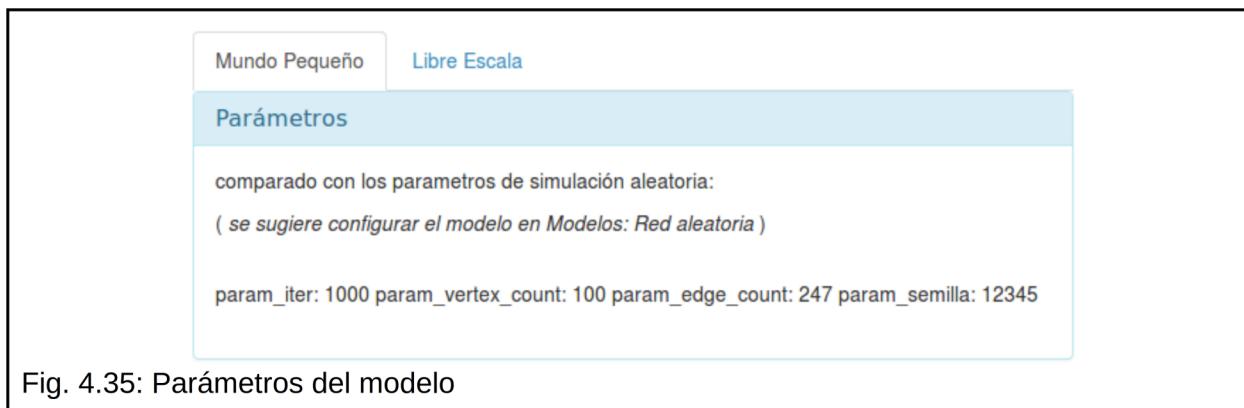


Fig. 4.35: Parámetros del modelo

1971
1972
1973
1974

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. 4.36](#) y [Fig. 4.37](#)).

<p>Validación 1</p> <p>Validación 1:</p> <p>para validar small world se tomó el criterio de comparar la red actual contra redes aleatorias (lo seleccionado en el tab de Modelos / Aleatorio).</p> <p>la comparación se lleva a cabo en el promedio del camino mas corto VS la media del promedio del camino más corto de las N redes simuladas.</p> <p>para mas detalles ver: Bialonski (2010) donde se detalla el procedimiento.</p> <p>Mundo pequeño: si Delta ~1 y Gamma > 1</p> <p>Delta = L/Lr. y Gamma = C/Cr.</p> <p>L: promedio camino más corto</p> <p>Lr: promedio de promedio camino más corto de las simulaciones</p> <p>C: transitivity</p> <p>Cr: promedio de transitivity de las simulaciones</p> <p>Delta: 0.821270386979065</p> <p>Gamma: 10.912232185799</p> <p>Es mundo pequeño: SI</p> <p>Fig. 4.36: Criterio de validación 1: es mundo pequeño</p>	<p>Validación 2</p> <p>Validación 2:</p> <p>Una red puede ser considerada 'mundo pequeño ' si la 'smallworldness' es mayor a 1 (qgraph::smallworldness)</p> <p>Un punto de vista más estricto dice de la red llevarla a 'smallworldness' >= 3 (Humphries & Gurney, 2008)</p> <p>smallworldness: 0.609533223585713</p> <p>mayor a 1: NO</p> <p>mayor a 3: NO</p> <p>Fig. 4.37: Criterio de validación 2: NO es mundo pequeño</p>
--	---

1975
1976
1977

Como se mencionó anteriormente, puede evaluarse la diferencia entre ambos enfoques de validación de mundo pequeño ([Fig. 98](#)).

Comparación de Validaciones

Recordar que para Validación 2, se vuelven a ejecutar simulaciones.

Sobre las simulaciones de Validación 1, se puede ver su detalle en su pestaña

Sobre las simulaciones de Validación 2, se puede ver su detalle en su pestaña

para cada variable: valor validación 1 | valor validación 2

transitivity [red]: 0.54233409610984 | 0.54233409610984

transitivity [promedio simulaciones]: 0.0496996477783549 | 0.0980560640732265

promedio distancia del camino mas corto [red]: 2.48183442525313 | 66.9226262626263

promedio distancia del camino más corto [promedio simulaciones]: 3.02194559136879 | 7.3752696969697

Fig. 4.38: Comparación de resultados de los criterios.

1978

1979 Libre escala 1996 - 2001

1980 Podemos ver que para el periodo 1996-2001, la red es de libre escala ([Fig. 4.39](#)). Además, vemos que hay mayor probabilidad en grados menores a 10 y menor en grados mayores a 10

1982 ([Fig. 4.40](#)).

1983

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 distribucion de ley de potencia obtenida.

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

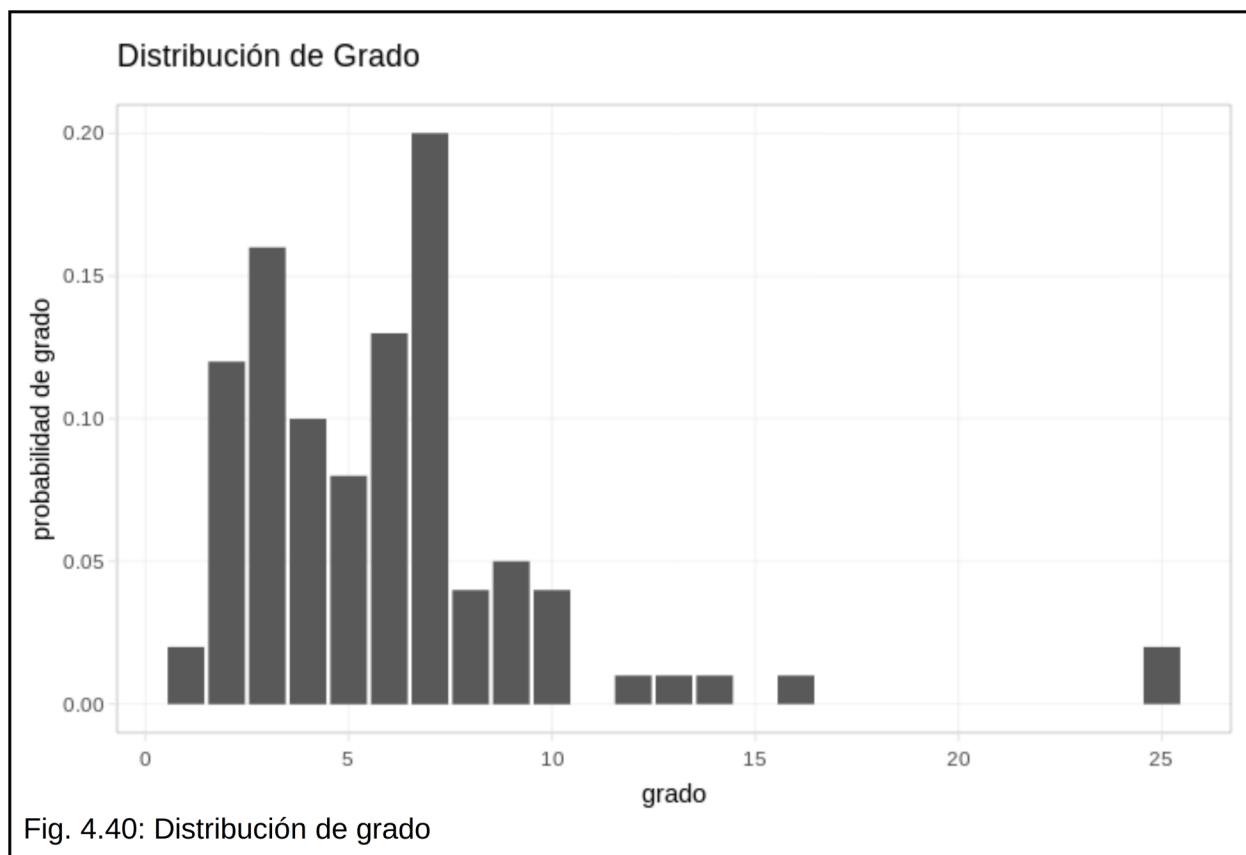
El test de ajuste de Kolmogorov-Smirnov significativo?

p-value test KS: 0.999544718619725

menor a 0.05: FALSE

Es de Libre Escala.

Fig. 4.39: resultado de comparación con modelos de libre escala. Es de libre escala.

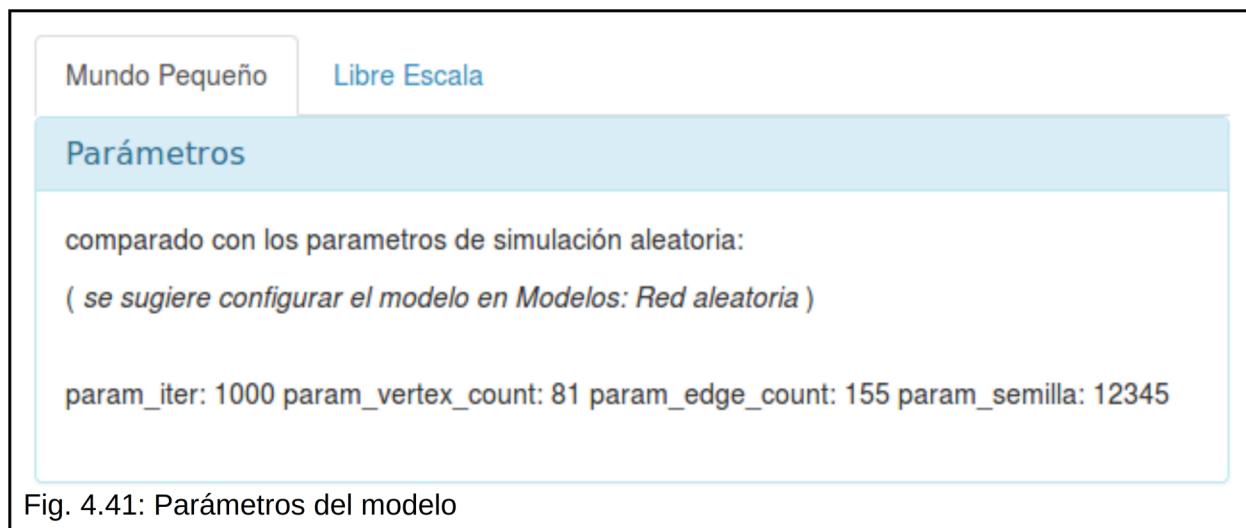


1984

1985 Mundo pequeño para el periodo 2002 - 2006

1986 Los parámetros utilizados para esta ejecución se pueden verificar en la Fig. 4.41.

1987



1988

1989 Podemos ver que para el periodo 2002-2006, la red no es de mundo pequeño según el criterio 1 (Fig. 4.42). Sin embargo, para el criterio 2 sí cumple con esta característica bajo el criterio laxo pero no bajo el estricto (Fig. 4.43).

Validación 1	Validación 2
<p>Validación 1:</p> <p>para validar small world se tomó el criterio de comparar la red actual contra redes aleatorias (lo seleccionado en el tab de Modelos / Aleatorio). La comparación se lleva a cabo en el promedio del camino más corto VS la media del promedio del camino más corto de las N redes simuladas.</p> <p>para mas detalles ver: Bialonski (2010) donde se detalla el procedimiento.</p> <p>Mundo pequeño: si Delta ~=1 y Gamma > 1</p> <p>Delta = L/Lr. y Gamma = C/Cr.</p> <p>L: promedio camino más corto</p> <p>Lr: promedio de promedio camino más corto de las simulaciones</p> <p>C: transitivity</p> <p>Cr: promedio de transitivity de las simulaciones</p> <p>Delta: 0.385343531190216</p> <p>Gamma: 18.8186344724456</p> <p>Es mundo pequeño: NO</p>	<p>Validación 2:</p> <p>Una red puede ser considerada 'mundo pequeño' si la 'smallworldness' es mayor a 1 (qgraph::smallworldness)</p> <p>Un punto de vista más estricto dice de la red llevarla a 'smallworldness' >= 3 (Humphries & Gurney, 2008)</p> <p>smallworldness: 1.26399469471737</p> <p>mayor a 1: SI</p> <p>mayor a 3: NO</p>

Fig. 4.42: Criterio de validación 1: **NO** es Mundo pequeño

Fig. 4.43: Criterio de validación 2: Es de mundo pequeño en criterio laxo, no lo es en estricto.

1994 La diferencia entre ambos enfoques de validación de mundo pequeño se observa en la [Fig. 4.44](#).

Comparación de Validaciones
Recordar que para Validación 2, se vuelven a ejecutar simulaciones.
Sobre las simulaciones de Validación 1, se puede ver su detalle en su pestaña
Sobre las simulaciones de Validación 2, se puede ver su detalle en su pestaña
para cada variable: valor validación 1 valor validación 2
transitivity [red]: 0.879736408566722 0.879736408566722
transitivity [promedio simulaciones]: 0.046748153265575 0.0526457990115321
promedio distancia del camino mas corto [red]: 1.28365384615385 75.8824074074074
promedio distancia del camino más corto [promedio simulaciones]: 3.33119344754278 5.73980987654321

Fig. 4.44: Comparación de resultados de los criterios.

1998 Libre escala 2002 - 2006

1999 Podemos ver que para el periodo 2002-2006, la red es de libre escala ([Fig. 4.45](#)), hay mayor

2000 probabilidad de grados bajos y bajas probabilidades de grados elevados. Sin embargo, en la
 2001 distribución vemos que el grado 7 se destaca bastante ([Fig. 4.46](#)).
 2002

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: 8.80267059423706

El exponente es mayor a 2: TRUE

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

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

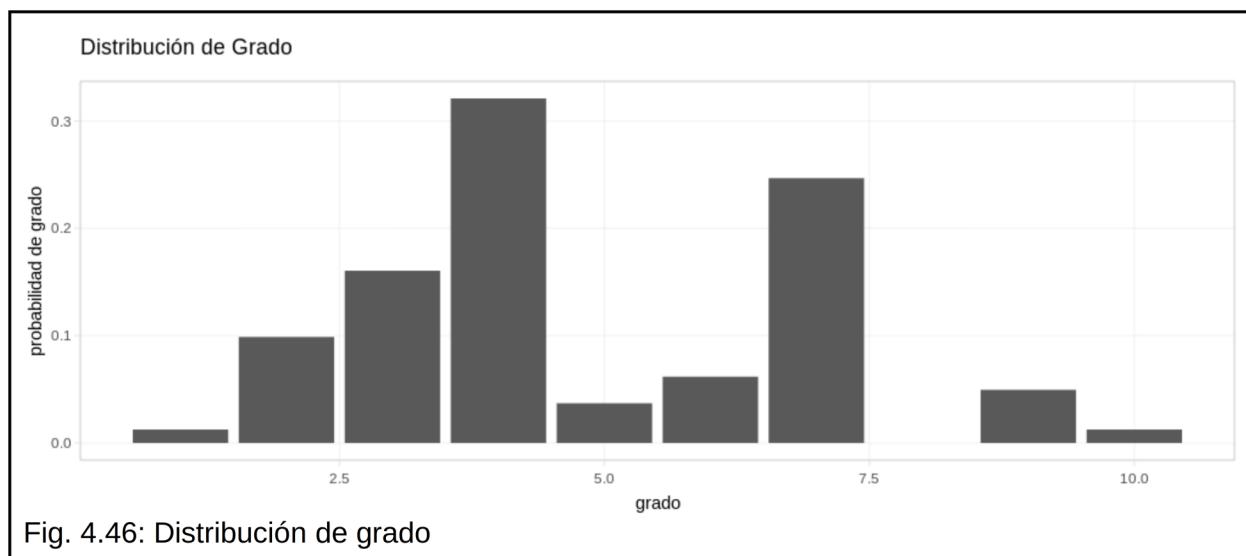
El test de ajuste de Kolmogorov-Smirnov significativo?

p-value test KS: 0.935237399412154

menor a 0.05: FALSE

Es de Libre Escala.

Fig. 4.45: resultado de comparación con modelos de libre escala. Es de libre escala.

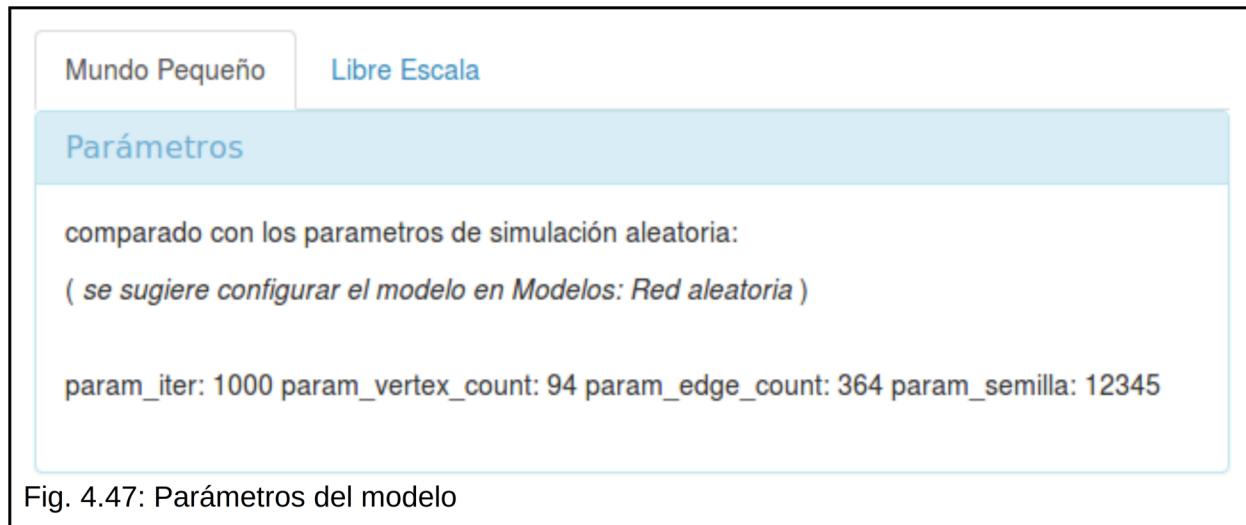


2003

2004 Mundo pequeño para el periodo 2007 - 2011

2005 Los parámetros utilizados para esta ejecución se pueden verificar en la [Fig. 4.47](#).

2006



2007

2008 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. 4.48](#) y [Fig. 4.49](#)).

2009

2010

2011

<p>Validación 1</p> <p>Validación 1:</p> <p>para validar small world se tomó el criterio de comparar la red actual contra redes aleatorias (lo seleccionado en el tab de Modelos / Aleatorio).</p> <p>la comparación se lleva a cabo en el promedio del camino mas corto VS la media del promedio del camino más corto de las N redes simuladas.</p> <p>para mas detalles ver: Bialonski (2010) donde se detalla el procedimiento.</p> <p>Mundo pequeño: si Delta ~=1 y Gamma > 1</p> <p>Delta = L/Lr. y Gamma = C/Cr.</p> <p>L: promedio camino más corto</p> <p>Lr: promedio de promedio camino más corto de las simulaciones</p> <p>C: transitivity</p> <p>Cr: promedio de transitivity de las simulaciones</p> <p>Delta: 0.553873119154232</p> <p>Gamma: 10.7826956093515</p> <p>Es mundo pequeño: SI</p>	<p>Validación 2</p> <p>Validación 2:</p> <p>Una red puede ser considerada 'mundo pequeño' si la 'smallworldness' es mayor a 1 (qgraph::smallworldness)</p> <p>Un punto de vista más estricto dice de la red llevarla a 'smallworldness' >= 3 (Humphries & Gurney, 2008)</p> <p>smallworldness: 0.594685437804652</p> <p>mayor a 1: NO</p> <p>mayor a 3: NO</p>
--	---

Fig. 4.48: Criterio de validación 1: es Mundo pequeño

Fig. 4.49: Criterio de validación 2: NO es Mundo pequeño

2012

2013 La diferencia entre ambos enfoques de validación de mundo pequeño se observa en la [Fig. 4.50](#).

2014

2015

<p>Comparación de Validaciones</p> <p>Recordar que para Validación 2, se vuelven a ejecutar simulaciones.</p> <p>Sobre las simulaciones de Validación 1, se puede ver su detalle en su pestaña</p> <p>Sobre las simulaciones de Validación 2, se puede ver su detalle en su pestaña</p> <p>para cada variable: valor validación 1 valor validación 2</p> <p>transitivity [red]: 0.887767220902613 0.887767220902613</p> <p>transitivity [promedio simulaciones]: 0.082332586680151 0.284945961995249</p> <p>promedio distancia del camino mas corto [red]: 1.34274952919021 82.7437657286662</p> <p>promedio distancia del camino más corto [promedio simulaciones]: 2.42429083982374 15.7937764813544</p>
--

Fig. 4.50: Comparación de resultados de los criterios.

2016 Libre escala 2007 - 2011

2017 Para el periodo 2007 - 2011 vemos que la red no es de libre escala ([Fig. 111](#)).

2018

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: 1.97458618149325

El exponente es mayor a 2: FALSE

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

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

El test de ajuste de Kolmogorov-Smirnov significativo?

p-value test KS: 0.163574343489645

menor a 0.05: FALSE

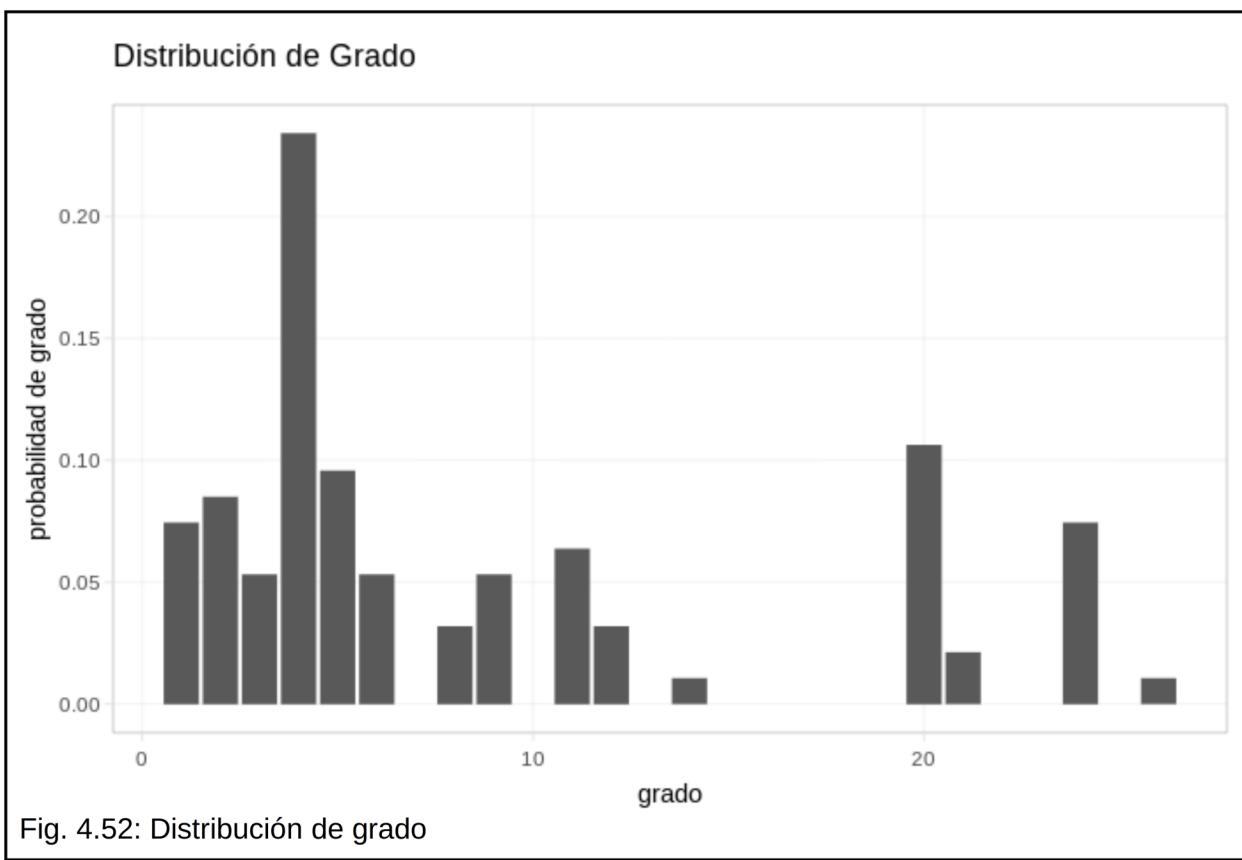
No es de Libre Escala.

Fig. 4.51: resultado de comparación con modelos de libre escala. No es de libre escala.

2019

2020
2021

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. 4.52](#)).

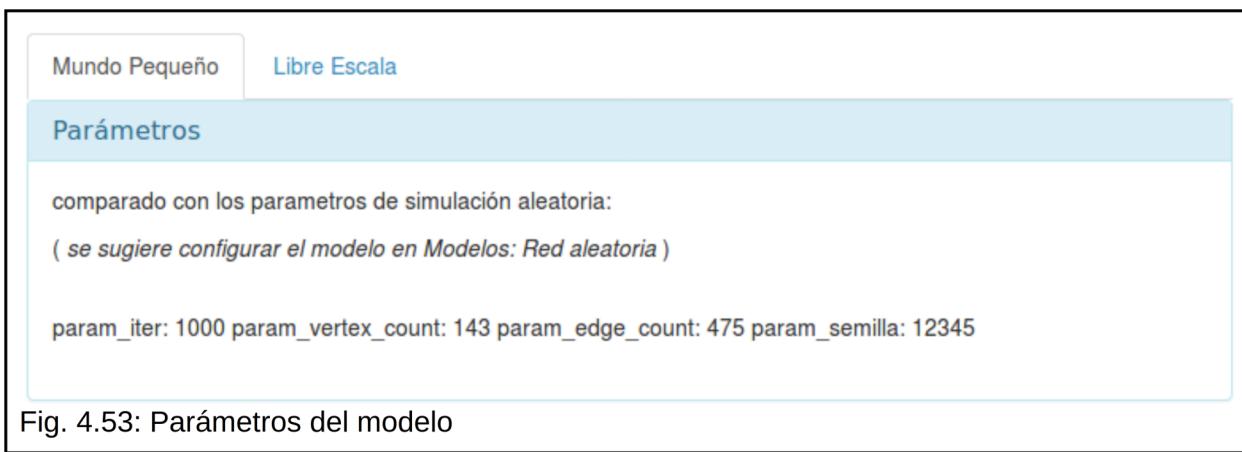


2022

2023 Mundo pequeño para el periodo 2012 - 2016

2024 Los parámetros utilizados para esta ejecución se pueden verificar en la [Fig. 4.53](#).

2025



2026

2027 Podemos ver que para el periodo 2012-2016, la red es de mundo pequeño para el criterio 1
2028 ([Fig. 4.54](#)). Sin embargo, para el criterio 2 sí cumple esta característica para el umbral laxo pero
2029 no para el estricto ([Fig. 4.55](#)).

2030

2031

<p>Validación 1</p> <p>para validar small world se tomó el criterio de comparar la red actual contra redes aleatorias (lo seleccionado en el tab de Modelos / Aleatorio). la comparación se lleva a cabo en el promedio del camino mas corto VS la media del promedio del camino más corto de las N redes simuladas. para mas detalles ver: Bialonski (2010) donde se detalla el procedimiento.</p> <p>Mundo pequeño: si Delta ~=1 y Gamma > 1</p> <p>Delta = L/Lr. y Gamma = C/Cr. L: promedio camino más corto Lr: promedio de promedio camino más corto de las simulaciones C: transitivity Cr: promedio de transitivity de las simulaciones Delta: 1.03650983289379 Gamma: 17.5491265217923 Es mundo pequeño: SI</p> <p>Fig.4.54: Criterio de validación 1: es Mundo pequeño</p>	<p>Validación 2</p> <p>Una red puede ser considerada 'mundo pequeño' si la 'smallworldness' es mayor a 1 (qgraph::smallworldness) Un punto de vista más estricto dice de la red llevarla a 'smallworldness' >= 3 (Humphries & Gurney, 2008)</p> <p>smallworldness: 1.09313691569536 mayor a 1: SI mayor a 3: NO</p>
--	--

Fig. 4.55: Criterio de validación 2: Es de Mundo pequeño en modo laxo, sin embargo no lo es en modo estricto.

2032

2033 Nuevamente, la diferencia entre ambos enfoques de validación de mundo pequeño se pueden observar en la [Fig. 4.56](#).

2034

2035

<p>Comparación de Validaciones</p> <p>Recordar que para Validación 2, se vuelven a ejecutar simulaciones. Sobre las simulaciones de Validación 1, se puede ver su detalle en su pestaña Sobre las simulaciones de Validación 2, se puede ver su detalle en su pestaña</p> <p>para cada variable: valor validación 1 valor validación 2</p> <p>transitivity [red]: 0.818181818181818 0.818181818181818 transitivity [promedio simulaciones]: 0.0466223670543265 0.10008561341571</p> <p>promedio distancia del camino mas corto [red]: 2.91586538461538 108.561902885847 promedio distancia del camino más corto [promedio simulaciones]: 2.81315747528867 14.516898453659</p>

Fig. 4.56: Comparación de resultados de los criterios.

2036

2037 Libre escala 2012 - 2016

2038 En el periodo de 2012-2016, la red es de libre escala ([Fig. 4.57](#)). Sin embargo, en la distribución se observa que en los grados 14 y 17 hay probabilidad alta ([Fig. 4.58](#)).

2039 Si vemos la referencia es de 0.05, y el máximo manejado para este periodo es de 0.20.

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: 15.8287203221457

El exponente es mayor a 2: TRUE

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

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

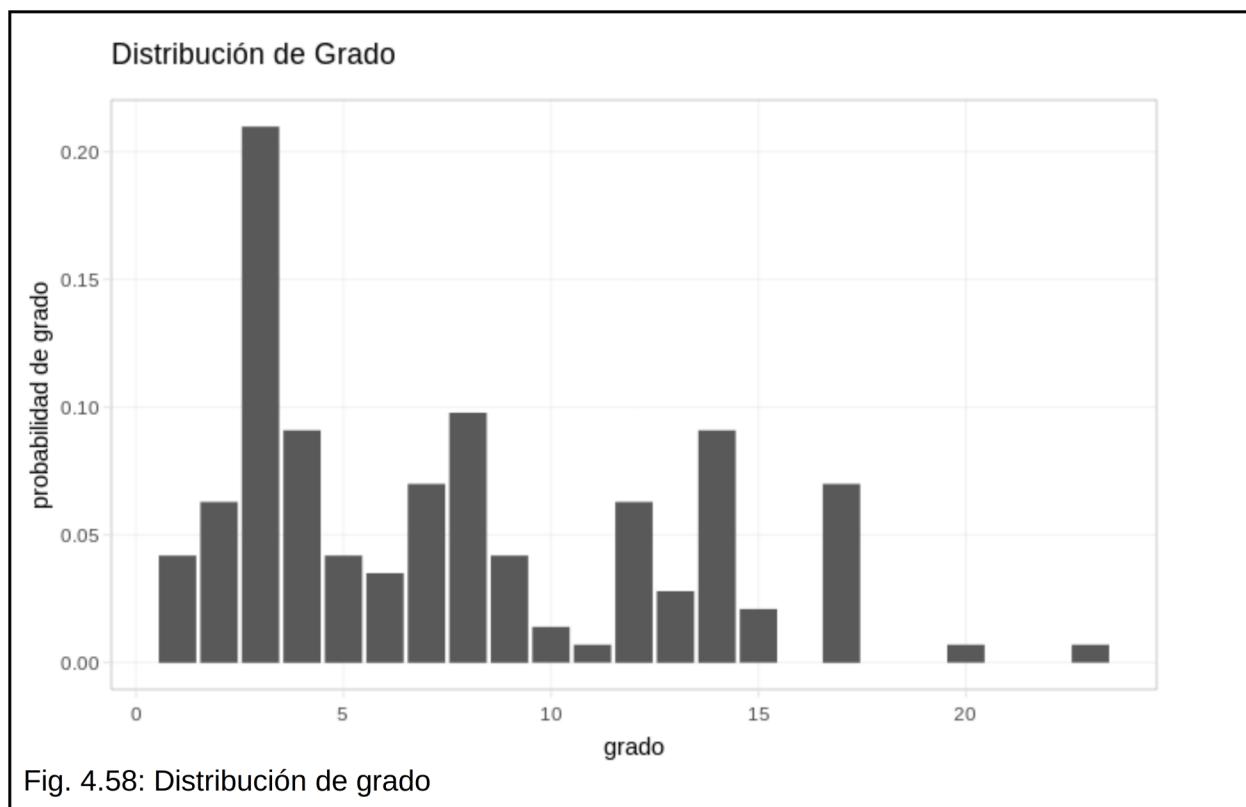
El test de ajuste de Kolmogorov-Smirnov significativo?

p-value test KS: 0.999942853828528

menor a 0.05: FALSE

Es de Libre Escala.

Fig. 4.57: resultado de comparación con modelos de libre escala. Es de libre escala.



2045

2046 Resumen de análisis de las propiedades de mundo pequeño y libre escala

2047

2048 El resumen de la comparación de las redes en el tiempo para las propiedades de mundo pequeño y libre escala se ve en [Tabla 4.11](#).

2049

2050 Tabla 4.11: Resumen de los análisis de Mundo pequeño y Libre Escala para todos los períodos considerados.

2051

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

2052

2053 Podemos ver 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 criterios de evaluación. Según el criterio 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, cuando se aplicó el umbral laxo, tenemos más variación y, para algunos períodos, si se comporta como mundo pequeño.

2054

2055

2056

2057

2058

2059

2060

2061

2062 Luego de esta comparación de la red contra redes generadas según diversos modelos, el
 2063 siguiente paso fue evaluar cómo se compara con otras redes de coautoría reales.
 2064

2065 Comparación con otras redes de coautoría

2066

2067 Así como comparamos la red con modelos similares generados en base sus propiedades, lo
 2068 más parecido que podríamos obtener en el mundo real son otras redes de coautoría de
 2069 escritura científica. De esta forma, comparamos las métricas con las obtenidas en otros
 2070 trabajos, tales como: (a) artículos por autor, (b) coeficiente de clustering (transitividad), (c)
 2071 tamaño del componente principal, (d) componente principal representado en porcentaje, (e)
 2072 distancia media. La siguiente Tabla fue obtenida del trabajo de [Koseoglu \(2016\)](#) y modificada
 2073 para agregar el caso de la RAAB ([Tabla 4.12](#)).

2074

2075 Tabla 4.12: Métricas obtenidas para la RAAB y otras redes de coautorías de revistas científicas.
 2076

métricas	Biomedical (Newman, 2001)	Computer science (Newman, 2001)	Management and organization (Acedo et al.,2006a)	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

2077

2078 Más allá de las métricas, hay que recordar que para hacer una mejor comparación se debe
 2079 tratar de entender el contexto de las otras revistas con las cuales se compara, sin embargo
 2080 este nivel de detalle quedará para trabajos futuros. Algunas preguntas que pueden ayudar a
 2081 entender el contexto son las siguientes: ¿de qué temáticas tratan?, ¿se acercan más a lo
 2082 técnico o a lo humano?, ¿cuál es la cota de artículos por edición ?, ¿con qué periodicidad se
 2083 publica?, ¿publican de todo el mundo?, ¿publican solo artículos locales?, ¿en qué idioma
 2084 publican?, ¿qué políticas de publicación tienen? Cada una de estas preguntas contestadas en
 2085 el contexto de publicación del paper en el cual se ven los resultados. Y siempre que las cosas
 2086 son “en el tiempo” estas preguntas pueden variar de tiempo T a tiempo T+1. Dicho esto,
 2087 pasamos a comparar solamente valores específicos.

2088

2089 En cuanto a cantidad de autores RAAB se posiciona entre Computer science ([Newman, 2001](#))
 2090 y Management and organization ([Acedo et al.,2006a](#)). Esta métrica representa que en el tiempo
 2091 analizado, los autores participan en promedio con 2.14 artículos a la revista para el caso de la
 2092 RAAB.

2093

2094 Respecto de autores por artículo, se posiciona cerca de Biomedical ([Newman, 2001](#)). En

2095 cuanto a qué puede interpretarse, esta métrica indica qué tanto trabajan en conjunto los
 2096 autores. Podemos ver que para el caso de RAAB trabajar en un trabajo original en equipo es
 2097 algo común.

2098
 2099 Al revisar transitividad, vemos que lo más próximo es Computer science ([Newman, 2001](#)). La
 2100 transitividad es la probabilidad media de que dos coautores también sean coautores con otro
 2101 en común. Esta métrica refuerza lo observado en el punto anterior de que los autores que
 2102 publican en la RAAB tienden a trabajar en equipo.
 2103

2104 Respecto de cuánto representa el componente principal en porcentaje de la red, vemos que lo
 2105 más próximo es Strategic Management ([Koseoglu 2016](#)). En cuanto a la métrica podemos
 2106 expresar que representa qué tan conectada está la red total. Cuanto más conectada esté
 2107 significa que hay menos equipos aislados, por trabajar en trabajos originales en conjunto por lo
 2108 menos alguna vez a lo largo de la historia de la revista.
 2109

2110 Observamos que según la distancia media está más próxima a Biomedical ([Newman, 2001](#)).
 2111 Recordemos que la distancia media es el promedio de la distancia entre nodos conectados en
 2112 la red. Vemos que la RAAB es la que da valor más bajo respecto de las otras redes.
 2113

2114 Finalmente para el tamaño del componente mayor observamos que la RAAB es la red más
 2115 pequeña respecto a las demás.

2116 Resumen de capítulo

2117 En este capítulo nos focalizamos en verificar las hipótesis planteadas en el capítulo 1 a través
 2118 de los datos obtenidos en el capítulo 2, utilizando la aplicación web generada en el capítulo 3.
 2119

2120 La hipótesis 1 “*La cantidad de autores aumenta en el tiempo debido al crecimiento de la
 2121 disciplina en el país, lo que se refleja en un aumento en la cantidad de nodos en la red*” nos
 2122 interesó para verificar si la disciplina de Antropología Biológica crece en el país luego de
 2123 verificar revisando la cantidad de autores a lo largo del tiempo en la Revista Argentina de
 2124 Antropología Biológica, acotamos el estudio a dicha publicación dado que buscamos
 2125 representación de la materia a nivel local, entendemos que esta hipótesis puede ser expandida
 2126 en publicaciones internacionales por autores locales y en otras revistas, llegamos a que no
 2127 podemos rechazar la hipótesis. De manera complementaria se analizó la cantidad de autores
 2128 por artículo por año como potencial indicador del trabajo en equipo.
 2129

2130 Luego de estar más familiarizados con los autores y artículos asociados a la RAAB, nos
 2131 interesó revisar si el avance de la revista depende de los autores fundadores, o si está
 2132 impulsada por nuevos autores. Para esto planteamos la hipótesis 2 “*Los autores fundadores se
 2133 refuerzan en el tiempo, lo cual puede ser medido por grado, intermediación y fuerza de
 2134 colaboración*” que luego de verificar a través del análisis de las métricas de grado,
 2135 intermediación y fuerza de colaboración, llegamos a que no podemos rechazar la hipótesis y
 2136 que la revista depende en gran medida de los autores participantes en las primeras ediciones.
 2137

2138 Para tratar de entender en mayor profundidad el comportamiento interno de la red revisamos
 2139 cómo se comportan los componentes llegando a plantear la hipótesis 3: “*Los componentes
 2140 más grandes absorben a los más pequeños a lo largo del tiempo, dado que autores periféricos
 2141 pasan a trabajar en conjunto con autores principales*” analizamos la evolución temporal del
 2142 tamaño del componente principal identificando momentos específicos de absorción de

2143 componentes. Luego detallamos qué componentes eran integrados al principal. Sin embargo a
2144 este enfoque le faltó el detalle de comprender qué estudios o temas fueron los que crearon
2145 puentes entre los componentes, y se decidió dejar para un futuro dado que podría ser de
2146 interés revisar con profundidad los temas tratados por estos textos y su relación con la creación
2147 de puentes entre componentes. De forma complementaria al análisis de componentes
2148 verificamos si existe relación con las actividades de investigación, para validar que el
2149 componente mayor es la rama principal de investigación, mientras que los otros componentes
2150 son ramas especializadas. Y complementamos con la búsqueda de comunidades, para dar otro
2151 enfoque a la búsqueda de la rama principal de investigación y ramas especializadas, ambas en
2152 una profundidad superficial analizando solo los títulos de los trabajos. Se podría profundizar
2153 analizando los textos completos en trabajos futuros.

2154

2155 Luego de tener un panorama interno de la revista, nos llamó la atención comprar con puntos de
2156 vista externos, inicialmente validación de propiedades comunes que se dan en redes del
2157 mundo real como Mundo pequeño y libre escala dando origen a la hipótesis 4: “*La red*
2158 *agregada total (desde 1996 a 2016) es de mundo pequeño puesto que es una característica*
2159 *presente en la mayoría de las redes de coautorías científicas*” que dependiendo del criterio
2160 adoptado puede rechazarse no. Finalmente, de forma complementaria realizamos la
2161 comparación con otras redes de coautoría, para complementar la vista hacia afuera, no solo
2162 quedarnos en los modelos de la hipótesis 4, sino constatar contra redes de coautoría existentes
2163 en trabajos previos.

2164

2165 Logrados los objetivos planteados en el capítulo 1, y con complementos a estos, damos por
2166 cerrado el capítulo 4 pasando al capítulo de cierre de la tesis.

2167

2168

2170 Capítulo 5: Conclusiones

2171 Resumen

2172 En la presente tesis se evaluaron las características de la red de coautorías de la Revista
 2173 Argentina de Antropología Biológica (RAAB), como caso de aplicación para el estudio de redes
 2174 de coautorías mediante el desarrollo de una aplicación web que facilitó el análisis a través del
 2175 cálculo automatizado de grafos, métricas y modelos, entre otras herramientas.

2176

2177 En el capítulo uno vimos qué es la RAAB y qué son las redes de coautoría, a modo de contexto
 2178 y antecedentes. Además, se presentaron los objetivos, preguntas e hipótesis que se trataron en
 2179 esta tesis.

2180

2181 En el capítulo dos vimos cómo se obtuvo el conjunto de datos. Definimos las fuentes de
 2182 entrada, las transformaciones realizadas y las salidas de esas transformaciones, y también nos
 2183 relacionamos con las herramientas para realizar estas tareas.

2184

2185 En el capítulo tres vimos cómo se realizó el proceso de armado de la herramienta para explorar
 2186 los datos que nos ayudó a investigar las inquietudes iniciales de forma reproducible, interactiva
 2187 y accesible por otros pares. Además, se detalló qué herramientas usamos para su creación y
 2188 diferentes consideraciones de armado.

2189

2190 El capítulo cuatro se focalizó en utilizar esta herramienta para analizar el caso puntual de la
 2191 RAAB. Este análisis se realizó contextualizando y armando un hilo conductor entre las
 2192 preguntas y sus respuestas.

2193

2194 Finalmente en el capítulo cinco vemos un repaso de lo realizado, discutimos posibles
 2195 aplicaciones, y planteamos algunos potenciales trabajos futuros.

2196 Productos asociados

2197 Esta tesis dio como resultado los siguientes productos que pueden ser de utilidad en otras
 2198 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 del armado de la base de datos.
- **Proceso de armado de la aplicación:** El detalle del proceso del armado de la aplicación.

2208 Consideraciones sobre el conjunto de datos y la variación de 2209 estructura del sitio en el tiempo

2210 Durante el desarrollo de la presente tesis hemos encontrado algunos desafíos y dificultades
2211 que vale la pena mencionar. Uno de ellos fue que los datos fueron obtenidos en 2017 pero a la
2212 hora de la escritura final de la tesis, la estructura de la página de la revista cambió y, por
2213 ejemplo, los archivos bibTex ya no son más accesibles.

2214

2215 Además, actualmente las referencias que tiene cada artículo son incluidas en el HTML. Sin
2216 embargo, en etapas tempranas de este estudio se descartó hacer análisis de referencias dado
2217 que, en aquel momento, estaban embebidas en los PDF y estos no eran accesibles de no ser
2218 por tareas manuales. A partir del cambio de estructura del sitio, las referencias de cada artículo
2219 están disponibles en la página de cada uno pero esto no era así al momento de extracción de
2220 los datos.

2221 Trabajos futuros

2222 Más allá de las ideas propuestas en la presente tesis, existe una amplia bibliografía en el
2223 campo de análisis de redes de coautorías y, tomadas en su conjunto, pueden expandirse a
2224 futuros trabajos, ya sean de desarrollo de herramientas para facilitar tareas o de aplicación de
2225 las técnicas para el soporte en la toma de decisiones fundamentadas en un mapa estratégico
2226 más complejo. Como menciona [Kumar \(2015\)](#), el análisis de redes de coautoría es una parte
2227 de un análisis que se debe complementar con otros estudios para lograr mejores resultados.

2228

2229 Finalmente el concepto de nodos que se asocian para generar un producto, y el conjunto de
2230 herramientas utilizadas para analizar este fenómeno puede abrir otras implementaciones en
2231 otras ramas diferentes de coautoría de artículos científicos.

2232

2233 A partir de los desafíos y los análisis realizados en la presente tesis, es posible sugerir varias
2234 líneas para explorar a futuro. A continuación mencionamos algunas de ellas.

2235 Actualización de la base de datos

2236 Un producto deseable podría ser expandir la base de datos y quitar la cota temporal. De esta
2237 forma, sería posible reutilizar la aplicación para expandir el análisis a más períodos de
2238 publicación en la RAAB.

2239

2240 También podría replicarse con la misma estructura para otras revistas, logrando que los
2241 análisis sean comparables a través del mismo proceso. Esto no es un trabajo trivial dado que
2242 una nueva base de datos implica el proceso de desambiguación de autores, la carga de datos
2243 en la aplicación, y luego sacarle la cota temporal en los lugares donde está filtrado. Sin
2244 embargo, el trabajo realizado en esta tesis agilizaría mucho la realización de estudios similares.

2245 Aspectos técnicos

2246 La aplicación puede mejorarse desde su estructura, procesos y performance. Así como también
2247 aspectos estéticos y de calidad de la herramienta, como puede ser navegabilidad, y
2248 sugerencias de trabajo para realizar análisis específicos. También pueden incluirse mejoras en
2249 evaluaciones y métricas. Por este motivo se optó por liberar el código fuente para que sea
2250 accesible y pueda ir evolucionando como comunidad para poder aceptar propuestas de

2251 interesados en el proyecto.

2252 Proyección sobre artículos

2253 Dada la red de naturaleza bipartita (autores - escriben - artículos), el trabajo actual se basó en
 2254 la proyección sobre autores, pudiéndose leer: “*autor - trabaja con - autor*”. La otra proyección
 2255 basada sobre los nodos artículos, que se puede leer como “*artículo - comparte autor con -*
 2256 *artículo*”, representa un interesante análisis para tener en cuenta en trabajos futuros.

2257 Análisis de textos y tópicos en el tiempo

2258 En etapas muy tempranas del desarrollo de la presente tesis otro objetivo de interés era la
 2259 realización de análisis de texto en base a los artículos publicados en la RAAB en el período
 2260 estudiado. Sin embargo, se decidió no incluir estos análisis, dado que extenderían
 2261 considerablemente los tiempos de trabajo y forman parte de otro tipo de hipótesis. Además,
 2262 otro aspecto que ayudó a tomar la decisión es que al momento de escritura de la tesis se está
 2263 desarrollando una “revolución” con respecto a la interpretación de imágenes, que avista que se
 2264 obtendrán mejores resultados facilitando esta tarea en trabajos futuros.

2265

2266 Por lo tanto, consideramos que, para trabajos futuros, sería interesante realizar análisis textual
 2267 de los PDF para la extracción de tópicos, evaluar su evolución en el tiempo y su relación con la
 2268 red de coautoría. En particular, resulta de interés investigar tópicos a través de un análisis
 2269 textual y verificar si esos tópicos resultantes cuadran con las descripciones de las líneas de
 2270 investigación en Antropología Biológica planteadas por [Cesani et al.\(2013\)](#). Además, se podría
 2271 validar si los grupos de coautores identificados en la red están asociados a los tópicos
 2272 determinados.

2273 Análisis de conectividad / puentes

2274 Este análisis permite plantear varias preguntas estratégicas. Para entender mejor el contexto
 2275 vamos a contextualizar lo que es un puente según [Barabási \(2016\)](#):

2276

Si una red consiste de dos componentes una arista bien puesta puede conectar ambos componentes, transformando la red en una red conectada. Esas aristas se llaman Puentes. En general un puente es cualquier link que si se corta, desconecta la red.

2277

2278 La herramienta desarrollada en esta tesis ayuda a analizar mejor cómo se fueron dando los
 2279 puentes en el tiempo, así como qué elementos generaron los mismos. También puede
 2280 utilizarse para analizar componentes candidatos para incentivar a generar nuevos “Puentes”
 2281 entre los componentes.

2282 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?

2288

2289 Así como se generan puentes para componentes, se puede tratar de fomentar la conectividad
 2290 entre comunidades, fomentando nuevas áreas de investigación o complemento entre equipos.
 2291 En este sentido, algunas preguntas que pueden surgir son:

2292 - ¿Quiénes son los puentes entre las comunidades?

2293 - ¿Qué trabajos se encuentran asociados a los puentes?

2294 - ¿Qué temáticas incluyen estos trabajos?

2295

2296 Dentro del análisis de relaciones se puede evaluar el vínculo entre los autores en el tiempo,

2297 conocer quiénes están vinculados, qué temas tratan, en qué época, si existe una frecuencia,

2298 entre otras.

2299

2300

2301

2302

2303

2304 Bibliografía

2305

2306 [1] AABA, Asoc. Antrop. Biolog. Arg. Revista Argentina de Antropología Biológica. URL:
 2307 <https://revistas.unlp.edu.ar/raab/>. Accessed: 2020-01-06.

2308

2309 [2] Barabási, A. L. (2016). *The power of network science, the beauty of network visualization*.
 2310 Cambridge University Press. ISBN: 978-1107076266. URL: <http://networksciencebook.com/>.

2311

2312 [3] Bialonski, S., M. Horstmann, and K. Lehnertz (2010). "From brain to earth and climate
 2313 systems: Small-world interaction networks or not?" In: *Chaos: An Interdisciplinary Journal of*
 2314 *Nonlinear Science* 20.1, p. 013134. DOI: 10.1063/1.3360561. URL:
 2315 <https://doi.org/10.1063/1.3360561>

2316

2317 [4] Blischak, J. D., E. R. Davenport, and G. Wilson (2016). "A Quick Introduction to Version
 2318 Control with Git and GitHub". In: *PLOS Computational Biology* 12.1, pp. 1-18. DOI:
 2319 <https://doi.org/10.1371/journal.pcbi.1004668>. URL: <https://doi.org/10.1371/journal.pcbi.1004668>.

2320

2321 [5] Blondel, V. D., J. Guillaume, R. Lambiotte, et al. (2008). "Fast unfolding of communities in
 2322 large networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, p.
 2323 P10008. DOI: 10.1088/1742-5468/2008/10/p10008.
 2324 URL: <https://doi.org/10.1088/1742-5468/2008/10/p10008>.

2325

2326 [6] Bourgeois, D. T. (2014). *Information Systems for Business and Beyond. Open Textbook*
 2327 *Challenge by the Saylor Academy*. URL: <https://bus206.pressbooks.com/>.

2328

2329 [7] Cesani Rossi, M. F., M. Sardi, S. E. Colantonio, et al. (2013). "Líneas de investigación
 2330 actuales de la antropología biológica argentina/Current researches in Argentine biological
 2331 anthropology". In: *Revista Argentina de Antropología Biológica* 16.1, pp. 31-37.
 2332 URL: <https://revistas.unlp.edu.ar/raab/article/view/749>.

2333

2334 [8] Clauset, A., M. E. J. Newman, and C. Moore (2004). "Finding community structure in very
 2335 large networks". In: *Phys. Rev. E* 70 (6), p. 066111. DOI: 10.1103/PhysRevE.70.066111.
 2336 URL: <https://link.aps.org/doi/10.1103/PhysRevE.70.066111>.

2337

2338 [9] Clauset, A., C. R. Shalizi, and M. E. J. Newman (2009). "Power-Law Distributions in
 2339 Empirical Data". In: *SIAM Review* 51.4, pp. 661-703. DOI: 10.1137/070710111.
 2340 URL: <https://doi.org/10.1137/070710111>.

2341

2342 [10] ERDÖS, P. and A. RÉNYI (1959). "On random graphs I". In: *Publ. Math. Debrecen*.

2343

2344 [11] Fatt, C. K., E. A. Ujum, and K. Ratnavelu (2010). "The structure of collaboration in the
 2345 Journal of Finance". In: *Scientometrics* 85.3, pp. 849-860. ISSN: 1588-2861. DOI:
 2346 10.1007/s11192-010-0254-0. URL: <https://doi.org/10.1007/s11192-010-0254-0>.

2347

2348 [12] Freeman, L. C. (1978). "Centrality in social networks conceptual clarification". In: *Social*
 2349 *Networks* 1.3, pp. 215 - 239. ISSN: 0378-8733.

2350 DOI: [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7).

2351 URL: <http://www.sciencedirect.com/science/article/pii/0378873378900217>.

2352

[13] 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>.

2356

[14] 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>.

2359

[15] 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>.

2360

[16] 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>.

2369

[17] 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>.

2373

[18] 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>.

2377

[19] 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.

2381

[20] Milgram, S. (1967). "The small world problem". In: Psychol. Today 2.1, pp. 60-67.

2383

[21] 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>.

2388

[22] 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>.

2393

[23] 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.

2398

[24] 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>.

2402

- 2403 [25] Newman, M. E. J. (2006). "Finding community structure in networks using the eigenvectors
 2404 of matrices". In: *Phys. Rev. E* 74 (3), p. 036104. DOI: 10.1103/PhysRevE.74.036104.
 2405 URL: <https://link.aps.org/doi/10.1103/PhysRevE.74.036104>.
- 2406
- 2407 [26] Newman, M. E. J. (2016). "Mathematics of Networks". In: *The New Palgrave Dictionary of
 2408 Economics*. London: Palgrave Macmillan UK, pp. 1-8. ISBN: 978-1-349-95121-5.
 2409 DOI: 10.1057/978-1-349-95121-5_2565-1.
 2410 URL: https://doi.org/10.1057/978-1-349-95121-5_2565-1.
- 2411
- 2412 [27] Newman, M. E. J. and M. Girvan (2004). "Finding and evaluating community structure in
 2413 networks". In: *Phys. Rev. E* 69 (2), p. 026113. DOI: 10.1103/PhysRevE.69.026113.
 2414 URL: <https://link.aps.org/doi/10.1103/PhysRevE.69.026113>.
- 2415
- 2416 [28] Ordoñez, G, A. Hernández Badillo, C. Hernández, et al. (2009). "Análisis bibliométrico de la
 2417 Revista de Economía Institucional en sus primeros diez años". In: *Revista de Economía
 2418 Institucional* 11.20. URL: <https://revistas.uexternado.edu.co/index.php/ecoins/article/view/354>.
- 2419
- 2420 [29] Pons, P. and M. Latapy (2005). Computing communities in large networks using random
 2421 walks (long version). arXiv: physics/0512106 [physics.soc-ph].
- 2422
- 2423 [30] Raghavan, U. N, R. Albert, and S. Kumara (2007). "Near linear time algorithm to detect
 2424 community structures in large-scale networks". In: *Physical Review E* 76.3. ISSN: 1550-2376.
 2425 DOI: 10.1103/physreve.76.036106. URL: <http://dx.doi.org/10.1103/PhysRevE.76.036106>.
- 2426
- 2427 [31] Rosvall, M, D. Axelsson, and C. T. Bergstrom (2009). "The map equation". In: *The
 2428 European Physical Journal Special Topics* 178.1, p. 13–23. ISSN: 1951-6401.
 2429 DOI: 10.1140/epjst/e2010-01179-1. URL: <http://dx.doi.org/10.1140/epjst/e2010-01179-1>.
- 2430
- 2431 [32] Saurkar, A. V., K. G. Pathare, and S. A. Gode (2018). n Overview on Web Scraping
 2432 Techniques and Tools.
 2433 URL:
http://www.ijfrcsce.org/download/browse/Volume_4/April_18_Volume_4_Issue_4/1524638955_25-04-2018.pdf.
- 2436
- 2437 [33] Savić, M., M. Ivanović, M. Radovanović, et al. (2015). "Exploratory Analysis of
 2438 Communities in Co-authorship Networks: A Case Study". In: *ICT Innovations 2014*. Springer
 2439 International Publishing, pp. 55-64. DOI: https://doi.org/10.1007/978-3-319-09879-1_6.
 2440 URL: https://link.springer.com/chapter/10.1007/978-3-319-09879-1_6.
- 2441
- 2442 [34] Wilson, G, J. Bryan, K. Cranston, et al. (2017). "Good enough practices in scientific
 2443 computing". In: *PLOS Computational Biology* 13.6, pp. 1-20. DOI: 10.1371/journal.pcbi.1005510.
 2444 URL: <https://doi.org/10.1371/journal.pcbi.1005510>.
- 2445
- 2446 [35] Wojciechowski, J, A. M. Hopkins, and R. N. Upton (2015). "Interactive Pharmacometric
 2447 Applications Using R and the Shiny Package". Eng. In: *CPT: pharmacometrics & systems
 2448 pharmacology* 4.3. 26225240[pmid], pp. e00021-e00021. ISSN: 2163-8306.
 2449 DOI: 10.1002/psp4.21. URL: <https://pubmed.ncbi.nlm.nih.gov/26225240>.
- 2450
- 2451 [36] Acedo, F. J., Barroso, C., Casanueva, C., & Galan, J. L. (2006). Co-Authorship in
 2452 Management and Organizational Studies: An Empirical and Network Analysis*. *Journal of
 2453 Management Studies*, 43(5), 957–983. doi:10.1111/J.1467-6486.2006.00625.X

2454

2455

2456

2457

2458

2459

2460

2461

2462

2463

2464

2465

2466 ANEXO: software y librerías utilizadas

2467 Dado que las librerías son parte esencial de cualquier sistema desarrollado hoy en día, se
 2468 especifica qué librerías se utilizaron para cada sección de la aplicación y con qué propósito. Si
 2469 como también sus citas correspondientes.

2470

libreria / software	sección	proposito	citas
R	Generación y aplicación	Lenguaje de programación	R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/ .
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. https://CRAN.R-project.org/package=tidyR
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. https://CRAN.R-project.org/package=dplyr
purrr	Generación y aplicación	Transformación datos	Lionel Henry and Hadley Wickham (2019). purrr: Functional Programming Tools. R package version 0.3.2. https://CRAN.R-project.org/package=purrr
stringr	Generación y aplicación	Transformación datos	Hadley Wickham (2019). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. https://CRAN.R-project.org/package=stringr
stringi	Generación y aplicación	Transformación datos	Gagolewski M. and others (2019). R package stringi: Character string processing facilities. http://www.gagolewski.com/software/stringi/ .
lubridate	Generación y aplicación	Transformación datos	Garrett Grolemund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL http://www.jstatsoft.org/v40/i03/ .
RSSQLite	Generación y aplicación	Operaciones Base de datos	Kirill Müller, Hadley Wickham, David A. James and Seth Falcon (2018). RSSQLite: 'SQLite' Interface for R. R package version 2.1.1. https://CRAN.R-project.org/package=RSSQLite
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. https://CRAN.R-project.org/package=DBI
igraph	Aplicación	Operaciones grafos	Csardi G, Nepusz T: The igraph software package for complex network research, InterJournal, Complex Systems 1695. 2006. http://igraph.org
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 http://www.jstatsoft.org/v48/i04/ .
RColorBrewer	Aplicación	Selección colores	Erich Neuwirth (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. https://CRAN.R-project.org/package=RColorBrewer
visNetwork	Aplicación	Visualización grafos	Almende B.V., Benoit Thieurmel and Titouan Robert (2019). visNetwork: Network Visualization using 'vis.js' Library. R package version 2.0.6. https://CRAN.R-project.org/package=visNetwork

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. https://CRAN.R-project.org/package=shiny
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. https://CRAN.R-project.org/package=DT
shinyWidgets	Aplicación	Funcionamiento Aplicación	Victor Perrier, Fanny Meyer and David Granjon (2019). shinyWidgets: Custom Inputs Widgets for Shiny. R package version 0.4.8. https://CRAN.R-project.org/package=shinyWidgets
bsplus	Aplicación	Funcionamiento Aplicación	Ian Lyttle (2018). bsplus: Adds Functionality to the R Markdown + Shiny Bootstrap Framework. R package version 0.1.1. https://CRAN.R-project.org/package=bsplus
plotly	Aplicación	Gráficos interactivos	Carson Sievert (2018) plotly for R. https://plotly-r.com
shinydashboard	Aplicación	Funcionamiento Aplicación	Winston Chang and Barbara Borges Ribeiro (2018). shinydashboard: Create Dashboards with 'Shiny'. R package version 0.7.1. https://CRAN.R-project.org/package=shinydashboard
shinydashboardPlus	Aplicación	Funcionamiento Aplicación	David Granjon (2019). shinydashboardPlus: Add More 'AdminLTE2' Components to 'shinydashboard'. R package version 0.7.0. https://CRAN.R-project.org/package=shinydashboardPlus
ndtv	Aplicación	Animaciones grafos	Skye Bender-deMoll (2018). ndtv: Network Dynamic Temporal Visualizations. R package version 0.12.3. https://CRAN.R-project.org/package=ndtv
intergraph	Aplicación	Transformación grafos	Bojanowski, Michal (2015) intergraph: Coercion Routines for Network Data Objects. R package version 2.0-2. http://mbojan.github.io/intergraph
assertive	Aplicación	Pruebas de funciones	Richard Cotton (2016). assertive: Readable Check Functions to Ensure Code Integrity. R package version 0.3-5. https://CRAN.R-project.org/package=assertive
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. https://CRAN.R-project.org/package=igraphdata
forcats	Aplicación	Transformación datos	Hadley Wickham (2019).forcats: Tools for Working with Categorical Variables (Factors). R package version 0.4.0. https://CRAN.R-project.org/package=forcats
visdat	Aplicación	Gráficos	Tierney N (2017). "visdat: Visualising Whole Data Frames." _JOSS_, *2*(16), 355. doi:10.21105/joss.00355 (URL: https://doi.org/10.21105/joss.00355), <URL: http://dx.doi.org/10.21105/joss.00355 >.
ggridges	Aplicación	Gráficos	Claus O. Wilke (2018). ggridges: Ridgeline Plots in 'ggplot2'. R package version 0.5.1. https://CRAN.R-project.org/package=ggridges
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. https://CRAN.R-project.org/package=skimr
ggrepel	Aplicación	graficos	Kamil Slowikowski (2019). ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.8.1.

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

			Butts C (2008). "network: a Package for Managing Relational Data in R." <i>Journal of Statistical Software</i> , *24*(2). <URL: http://www.jstatsoft.org/v24/i02/paper >.
2471	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. https://CRAN.R-project.org/package=networkDynamic

2473 ANEXO: Detalle de algoritmos para búsqueda de 2474 comunidades

2475 Cada uno de los algoritmos disponibles representa diferentes formas de considerar la
2476 agrupación y su interpretación depende de que representan los nodos y las aristas. Los
2477 algoritmos utilizados están disponibles a través de la [librería igraph](#). Un breve resumen se
2478 puede ver en [Tabla 21](#).

2479
2480 Tabla 21: Resumen algoritmos de búsqueda de comunidades.

Función en Herramienta	Tipo Comunidad	Referencia de implementación	Descripción algoritmo
Cluster_edge_betweenness	Edge Betweenness	http://igraph.org/r/doc/cluster_edge_betweenness.html	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 Newman-Girvan (2004)
Cluster_label_prop	Propagating Labels	http://igraph.org/r/doc/cluster_label_prop.html	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 Raghavan,Albert, et al. (2007)
Cluster_leading_eigen	Leading Eigenvector	http://igraph.org/r/doc/cluster_leading_eigen.html	Utiliza la matriz de modularidad, y luego utiliza los autovectores y autovalores de la misma. Para más detalles ver Newman (2006)
Cluster_louvain	Multi-level Optimization of Modularity	http://igraph.org/r/doc/cluster_louvain.html	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 Blondel, Guillaume, et al.(2008)
Cluster_walktrap	Short Random Walks	http://igraph.org/r/doc/cluster_walktrap.html	La idea es que caminatas aleatorias cortas se quedan en la misma comunidad. Para más detalles ver Pons-Latapy (2005)
Cluster_infomap	Infomap community finding	http://igraph.org/r/doc/cluster_infomap.html	Encuentra estructuras de comunidad que minimizan el largo de la descripción de la trayectoria de un caminador aleatorio. Rosvall, Axelsson, et al.(2009)

2483 ANEXO: Detalle procesamiento bibTEX

2484 Los archivos bibtex fueron generados de la siguiente manera:

- 2485
- 2486 1) Se accedió al link de javascript generado por revista
 - 2487 2) Se guardó la página HTML que tenía un elemento textarea con el texto bibtex correspondiente
 - 2488 3) Se extrajo el texto bibtex del elemento textarea
 - 2489 4) Se guardó en un archivo bibtex
 - 2490 5) Se trató de parsear el archivo bibtex con la [librería RefManageR](#)
 - 2491 6) Se encontraron problemas generalmente relacionados a mala escritura del formato bibtex
 - 2494 a) Tema de escribir autor1 Y autor2, mientras que en bibtex es autor1 **AND** autor2.
 - 2495 7) Se encontraron problemas relacionados con el encoding de los archivos.
 - 2496 a) ASCII vs UTF8 (acentos, ñ, etc)
 - 2497 8) Se resolvieron ambos problemas
 - 2498 a) El primero identificando los casos de forma manual
 - 2499 b) El 2do utilizando sistema operativo Linux para realizar la transformación y guardar en UTF-8
 - 2500 9) Se encontraron inconsistencias en los datos presentados en el formato bibtex y lo que se encontraba presentado en los HTML de la revista.
 - 2503 a) Falta de dato volumen
 - 2504 b) Falta de abstract
 - 2505 10) Se terminó extrayendo la información que se consideraba complementaria y se almacenó en la base de datos de la revista.

2508 Problemas de transformación y encoding bibtex

2509 Detalles técnicos de los sistemas operativos utilizados:

2510 Linux: LInux Mint 17.3 Rosa (basado en Ubuntu 14.04)

2511 Windows: Windows 10 y Windows 7

2512 Luego de resolver estos temas en Linux (con caracteres no ASCII), se pudo generar el archivo de texto y luego levantarlos sin problemas.

2515 Sin embargo cuando utilice 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

2520 De esta forma se optó por hacer las tareas de parseo de archivos en Linux, y luego de tener generados los archivos bases, se pudo seguir trabajando en ambos sistemas operativos.

2523 Problemas autores

2524 Apellidos con apóstrofe.

2525 <https://revistas.unlp.edu.ar/raab/article/view/229> Gabriela L'Heureux

2526 <https://revistas.unlp.edu.ar/raab/article/view/1513> Manuel D. D'Angelo del Campo

2527

2528 Solución: reemplazar el apóstrofe por comilla simple.

2529 Problemas con el volumen de la revista

2530

Link bibtex	
https://revistas.unlp.edu.ar/raab/article/cite/612/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/425/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/2541/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/554/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/2505/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/458/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/1172/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/258/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/2506/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/259/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/574/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/2514/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/260/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/561/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/2516/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/261/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/557/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/2515/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/268/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/2507/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/269/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/2177/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/270/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/2513/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/271/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/272/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/273/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/274/BibtexCitationPlugin	Falta campo volumen
https://revistas.unlp.edu.ar/raab/article/cite/275/BibtexCitationPlugin	Falta campo volumen

2531 Problema Abstracts Vacíos en bibtex

https://revistas.unlp.edu.ar/raab/article/cite/800/BibtexCitationPlugin	Comillas dobles y simples
https://revistas.unlp.edu.ar/raab/article/cite/173/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/2860/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/747/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/2223/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/2489/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/612/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1090/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/352/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/377/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1510/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/675/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/863/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/3047/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/874/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1172/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/3722/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1513/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1514/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1515/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/2516/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/3036/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/228/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/269/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/2177/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/272/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/274/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1706/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1709/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1710/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1720/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1721/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1722/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1730/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1735/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1745/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1746/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1747/BibtexCitationPlugin	

https://revistas.unlp.edu.ar/raab/article/cite/1748/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1813/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1818/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1826/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1829/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1831/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1837/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1839/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1840/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1842/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1852/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1861/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1804/BibtexCitationPlugin	
https://revistas.unlp.edu.ar/raab/article/cite/1869/BibtexCitationPlugin	

2532

2533

2534

2535

2536

2537

2538 ANEXO: Detalle consideraciones PDF

2539 Los archivos PDF al final no fueron considerados. Se eliminó la sección de text mining para
 2540 tópicos, quedando para potenciales trabajos futuros. Sin embargo existieron las siguientes
 2541 lecciones aprendidas:

2542 Obtención de los pdf

2543 Se pudieron bajar sin problema, se agregó una demora entre archivo y archivo, para no saturar
 2544 el servidor de la revista.

2545 Parseo de los pdf

2546 Conversión de los pdf a texto; Existían diferentes formatos de los pdf;
 2547 Se probó:

- 2548 • Bajar un artículo de la edición más antigua: es un escaneo
- 2549 • Bajar un artículo de una edición intermedia: es un escaneo
- 2550 • Bajar un artículo de una edición moderna: es un texto

2551

2552 Potenciales soluciones:

- 2553 • OCR (*optical character recognition*) para los escaneos
- 2554 • Extracción de texto para los pdf más modernos.

2556 Extracción de texto (Se utilizó el lenguaje R , con la [librería pdftools](#))

- 2557 • Errores con formato de más de 1 columna.
- 2558 • Errores de extracción de texto por márgenes de página.
- 2559 • Se decidió convertir cada hoja a una imagen y luego pasarlas por OCR.

2561 Aplicación de OCR (Se utilizó el lenguaje R , con las [librerías pdftools y tesseract](#))

- 2562 • Extracción de las imágenes de cada página a través de la [librería pdftools](#)
- 2563 • Procesar cada imagen por artículo ejecutando OCR a través de la [librería tesseract](#), se aplicó el uso del motor de idioma en español.
- 2564 • Compaginar todos los textos en un archivo txt final que contenga todos los resultados
- 2566 OCR de cada una de las páginas.

2567

2568 Con esto se logró superar considerablemente el problema de márgenes, sin embargo seguían
 2569 existiendo problemas con los textos, y extracción de los datos de los mismos.

2570

2571 Se decidió no incluir la parte de text mining en la tesis, dado que sino se extendería
 2572 considerablemente. Además dado que hoy en día (2017-2020) hay una revolución con la
 2573 interpretación de imágenes, que tal vez en breve tengamos mejores resultados pudiendo
 2574 realizar de forma más fácil en trabajos futuros.

2575 ANEXO: Detalle estandarización autores

2576 Se juntaron los conjuntos de datos de los orígenes de HTML , y los archivos de citas BIBtex.

2577 Se exportaron los datos resultantes a un archivo en formato texto separado por tabulaciones.

2578

2579 Este archivo se importó en la herramienta [OpenRefine](#) la cual resulta muy práctica a la hora de
2580 buscar similitudes entre registros de cada columna. La versión utilizada para realizar este
2581 trabajo fue OpenRefine 2.8.

2582

2583 Una vez cargado el archivo , se procedió a repetir la columna autor , y a la nueva ponerle de
2584 nombre “autor_original”. Esta columna se agrega para cumplir 2 funciones, una de copia de
2585 respaldo, en caso de que llegase a haber inconvenientes. Y la segunda, que es para mapear el
2586 valor original de la columna con el valor obtenido luego de procesar la columna “autor”
2587 repetidas veces. ([Fig. anx_est_aut_1](#))

2588

	articulo_html_url	autor	autor_original
5.	https://revistas.unlp.edu.ar/raab/article/view/1090	Sandra C. Scala	Sandra C. Scala
7.	https://revistas.unlp.edu.ar/raab/article/view/1091	Marina Laura Sardi	Marina Sardi
13.	https://revistas.unlp.edu.ar/raab/article/view/1172	Sandra Assis	Sandra Assis
29.	https://revistas.unlp.edu.ar/raab/article/view/151	Susana Blanca Ocampo	Susana Blanca Ocampo
34.	https://revistas.unlp.edu.ar/raab/article/view/1512	Marina Laura Sardi	Marina L. Sardi

Fig. anx_est_aut_1: 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.

2589

2590 Sobre la columna “autor” las tareas realizadas fueron para buscar similitudes a través de
2591 métodos de agrupamiento (*clustering*). A continuación se pueden ver los métodos que se
2592 ejecutaron, y en qué orden. Más allá de este resultado se exportó un archivo con las
2593 transformaciones realizadas; sin embargo este archivo no conserva los métodos de clustering
2594 usados, ni los parámetros utilizados; por lo cual se pasó a detallar en formato textual dentro de
2595 este anexo. (ver [Iteraciones de clustering](#))

2596

2597 Revisión y validación con expertos

2598 Luego de depurar empíricamente con diferentes formas de agrupar, validamos los nombres y

2599 casos con los expertos del área. Surge una lista de autores a unificar, que no aparecieron en el
 2600 análisis de clustering o que había que modificar. (ver [Tabla 22](#))

2601
 2602 Se cambió el formato de excel enviado por los expertos a un formato más amigable para la
 2603 automatización. Se armó una utilidad intermedia que dado un excel, transforma a el formato de
 2604 importación del openrefine, que es un archivo con un formato específico. Gracias a la
 2605 herramienta se pudieron distinguir errores que se pasaron, en la parte de la unificación de
 2606 autores.

2607 Iteraciones de clustering

2608 Clustering - Iteración 1

2609 **Método y parámetros**

2611 Clustering > columna autor > método: keycollision > keying function: fingerprint

2612

2613 **Criterio de selección**

2614 Empírico. Es la primera opción al seleccionar la opción de clustering en openrefine en la
 2615 versión utilizada (version: 2.7).

2616

2617 **Observación**

2618 Dado que son varios los cambios se procede a detallar reglas generales aplicadas, para ver
 2619 detalles de cuáles fueron exactamente ver el archivo exportado de reglas del openrefine.

2620

2621 **Resultado**

2622 Al realizar el primer clustering en nombres de autores se puede ver que existen varios Temas
 2623 con escritura del nombre:

- 2624 - Nombres que consideran el nombre intermedio (María Antonia Luis)
- 2625 - Nombres que al intermedio lo ponen con una abreviación (María A. Luis)
- 2626 - Abreviación del nombre intermedio con punto (María A. Luis)
- 2627 - Abreviación del nombre intermedio sin punto (María A Luis)
- 2628 - Nombres con acentos o caracteres especiales (María A. Luis)
- 2629 - Nombres en los que se obvian los acentos o caracteres especiales. (Maria A. Luis)
- 2630 - Nombres que cambian en el tiempo > ej: Cesani > Cesani Rossi

2631

2632 Clustering - Iteración 2

2633

2634 **Método y parámetros**

2635 Clustering > columna autor > método: keycollision > keying function: fingerprint

2636

2637 **Criterio de selección**

2638 Empírico. Volver a tratar de ejecutar el mismo método para ver que no existan más duplicados
 2639 con los parámetros utilizados.

2640

2641 **Resultado**

2642 Se trata de dar un segundo clustering bajo la misma metodología y no se encuentran
 2643 resultados

2644

2645 Clustering - Iteración 3

2646

2647 **Método y parámetros**

2648 Se procede a probar con otro tipo de configuración de clustering

2649 método: keycollision > keying function: ngram-fingerprint > ngram-size: 2

2650

2651 **Criterio de selección**

2652 Empírico. Dentro de las opciones de parámetros dentro del método de key collision es la 2da

2653 opción. El tamaño del enagrama fue seleccionado porque es el que viene por defecto.

2654

2655 **Observación**

2656 Dados que son varios los cambios se procede a detallar reglas generales aplicadas, para ver
2657 detalles de cuáles fueron exactamente ver el archivo exportado de reglas del openrefine.

2658

2659 **Resultado**

2660 Se encuentran temas con:

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

2664

2665 **Clustering - Iteración 4**

2666

2667 **Método y parámetros**

2668 Se procede a ejecutar nuevamente bajo la misma metodología.

2669 método: keycollision > keying function: ngram-fingerprint > ngram-size: 2

2670

2671 **Criterio de selección**

2672 Empírico. Validación de lo aplicado en Iteración 3. Se espera que no haya resultados.

2673

2674 **Observación**

2675 Dados que son varios los cambios se procede a detallar reglas generales aplicadas, para ver
2676 detalles de cuáles fueron exactamente ver el archivo exportado de reglas del openrefine.

2677

2678 **Resultado**

2679 No se encuentran resultados.

2680 Clustering - Iteraciones 5-9

2681

2682 **Método y parámetros**

2683 Se procede a ejecutar nuevamente bajo la misma metodología , variando el tamaño del
2684 enagrama con los valores 3,4,5,6,7

2685 método: keycollision > keying function: ngram-fingerprint > ngram-size: 2

2686

2687 **Criterio de selección**

2688 Empírico. Probando expandir el tamaño del enagrama base para el clustering.

2689

2690 **Resultado**

2691 No se encuentran resultados. Para los parámetros 3,4,5,6.

2692 En el parámetro con valor 7 si, pero no se utilizaron las sugerencias planteadas, dado que
2693 afecta los nombres y apellidos cortos.

2694

2695 **Clustering - Iteración 10**

2696

2697 **Método y parámetros**

2698 Se procede a pasar a otra estrategia. Método: keycollision , keying function: metaphone3

2699

Criterio de selección

2701 Empírico. Probando otra función.

2702

Resultado

2704 Dando nuevos resultados esta vez relacionados con:

- 2705 - Casos de nombres compuestos o apellidos compuestos (Fernando Ramírez Rozzi)
 - 2706 - 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
 - 2709 - Hay casos dudosos como pro ejemplo: Rolando González-José , vs Rolando González
 - 2710 - Al ser apellidos tan comunes es difícil de realizar elecciones; a constatar con expertos.

2712

2713 Respecto de consideraciones con apellidos del estilo

- 2714 - 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.
- 2717 - 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.

2720

Clustering - Iteración 11

2722

Método y parámetros

2724 Se procede a validar la estrategia anterior. Método: keycollision , keying function: metaphone3

2725

Criterio de selección

2727 Empírico. Verificando método utilizado en iteración 10, para búsqueda de más resultados.

2728

Resultado

2730 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.

2733

Clustering - Iteración 12

2735

Método y parámetros

2737 Se procede a pasar a otra estrategia. Method: keycollision , keying function: cologne-phonetic

2738

Criterio de selección

2740 Empírico. Siguiente opción dentro de keycollition.

2741

Resultado

- 2743 - 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
- 2748 - Se siguen encontrando casos de apellidos compuestos

2749 - Se siguen encontrando casos de nombres con iniciales

2750

2751 Clustering - Iteración 13

2752 Método y parámetros

2753 Se procede a validar la estrategia anterior.

2754 Metodo: keycollision , keying function: cologne-phonetic

2755

2756 Criterio de selección

2757 Empírico. Verificando método utilizado en iteración 13, para búsqueda de más resultados.

2758

2759 Resultado

- 2760 - Se volvió a ejecutar esa estrategia, y solo aparecieron los casos de dejean y otro que no
- 2761 cuadran los apellidos, solo los nombres.

2762

2763 Clustering - Iteración 14

2764

2765 Método y parámetros

2766 Se procede con otra estrategia:

2767 Method: nearest neighbour , distance function: levenshtein , Radius: 1.0, Block Chars: 6

2768

2769 Criterio de selección

2770 Empírico. Siguiente combinación dentro de la herramienta.

2771

2772 Resultado

- 2773 - En los casos de :
- 2774 - Emma E. Alfaro(20 rows) / Emma L. Alfaro(14 rows)
- 2775 - María C. Barboza(11 rows) / María J. Barboza(7 rows)
- 2776 - No se si es la misma personas, hay que ver con los expertos.
- 2777 - No se hace ese cambio

2778

2779 Clustering - Iteración 15

2780

2781 Método y parámetros

2782 Se procede con otra estrategia: aumentamos el radio a 2

2783 Method: nearest neighbour , distance function: levenshtein , Radius: 2.0, Block Chars: 6

2784

2785 Criterio de selección

2786 Empírico. Variando los parámetros del algoritmo probado en iteración anterior.

2787

2788 Resultado

- 2789 - Generico, Aparecieron casos que eran de segundos nombres.

2790

2791 Clustering - Iteraciones 16,17,18,19,20,21,22

2792

2793 Método y parámetros

2794 Se procede con otras estrategias de variar el block chars: entre 3 y 10

2795 Method: nearest neighbour, distance function: levenshtein , Radius: 2.0, Block Chars: 3, 10.

2796

2797 Criterio de selección

2798 Empírico. Variando los parámetros del algoritmo probado en iteración anterior.

2799

2800 **Resultado**

- Sin resultados favorables.

2802

2803 **Clustering - Iteraciónes 23**

2804

2805 **Método y parámetros**

2806 Se procede con otras estrategia de aumentar el radio nuevamente por 1 punto.

2807 Method: nearest neighbour , distance function: levenshtein , Radius: 3.0, Block Chars: 6

2808

2809 **Criterio de selección**

2810 Empírico. Variando los parámetros del algoritmo probado en iteración anterior.

2811

2812 **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.

2813

2822 **Clustering - Iteración 24**

2823

2824 **Método y parámetros**

2825 Se procede a repetir la estrategia

2826 Method: nearest neighbour , distance function: levenshtein , Radius: 3.0, Block Chars: 6

2827

2828 **Criterio de selección**

2829 Empírico. Repetimos los parámetros anteriores para ver cómo afecta.

2830

2831 **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.

2832

2839 **Clustering - Iteración 25**

2840

2841 **Método y parámetros**

2842 Se procede con otra estrategia: para reducir los casos problemáticos y ver si se encuentran otros mejores, bajamos los Block Chars a 3.

2843 Method: nearest neighbour , distance function: levenshtein , Radius: 3.0, Block Chars: 3

2844

2846 **Criterio de selección**

2847 Empírico. Variando los parámetros del algoritmo probado en iteración anterior.

2848

2849 **Resultado**

- Resultados genericos.

- 2851 - Se encontraron casos como el de Ana M. Inda vs Ana Inda
 2852 - Siguen existiendo casos con los que tener mucho cuidado dada la poca variabilidad
 2853 pero evidente diferencia.
 2854 - Queda validar con los expertos.

2855
 2856 **Clustering - Iteración 27**

2857

Método y parámetros

2859 Se procede con otra estrategia: para reducir los casos problemáticos y ver si se encuentran
 2860 otros mejores, bajamos los Block Chars a 2.

2861 Method: nearest neighbour , distance function: levenshtein , Radius: 3.0, Block Chars: 2

2862

Criterio de selección

2864 Empírico. Variando los parámetros del algoritmo probado en iteración anterior.

2865

Resultado

- 2867 - Siguen existiendo casos con los que tener mucho cuidado dada la poca variabilidad
 2868 pero evidente diferencia.
 2869 - Queda validar con los expertos.

2870

Clustering - Iteración 28

Método y parámetros

2873 Se procede con otra estrategia, cambiar la función de distancia.

2874 Method: nearest neighbour , distance function: PPM , Radius: 3.0, Block Chars: 3

2875

Criterio de selección

2876 Empírico. Variando la función de distancia.

2877

Resultado

- 2879 - Se encuentran muchos resultados sin embargo sin valor dado que no toma en cuenta
 2880 los apellidos, agrupa casos de nombres muy comunes como por ejemplo: los nombres
 2881 que comienzan con 'María X yyyy'
 2882 - Sirve mucho para identificar los segundos nombres de varias personas, como por
 2883 ejemplo > cristina b dejean > cristina beatriz dejean

Clustering - Iteración 29

2886

Método y parámetros

2887 Se procede nuevamente con la estrategia

2888 Method: nearest neighbour , distance function: PPM , Radius: 3.0, Block Chars: 3

2889

Criterio de selección

2890 Empírico. Validando la estrategia de iteración anterior.

2891

Resultado

- 2892 - Mismas observaciones anteriores, un caso a revisar específico:
 2893 - Héctor Hugo Varela(14 rows) / Hugo H. Varela(8 rows)

2897 Resultado de Revisión y validación con expertos

2898 La [Tabla anx_estandarizacion_autores_1](#) tiene el mapeo de los autores revisados por expertos

2899 y su valor original.

2900

2901 Tabla anx_estandarizacion_autores_1: 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. Zangrado	Francisco A. Zangrado
Francisco A. Zangrado	A. Francisco Sangrado
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 Muñé	Maria C. Muñé
María Cristina Muñé	María C. Muñé
María Cristina Muñé	María Cristina Muñé
María Cristina Muñé	Maria Muñé
María Cristina Muñé	María Muñé
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	Maria Diaz
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	Maria 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