

# Aplicación de técnicas de análisis de redes sociales y de co-ocurrencia de palabras en la determinación de frentes de investigación<sup>1</sup>

Claudia E. Boeris

Instituto Argentino de Radioastronomía, CCT-La Plata, Conicet

## Resumen

Se realiza un estudio bibliométrico de la producción científica de los investigadores del IAR, con el objetivo de determinar los frentes de investigación del instituto y los grupos de investigadores que trabajan en esos frentes.

Se aplican los métodos de análisis de co-ocurrencia de palabras, análisis de coautoría y análisis de redes sociales (ARS) mediante la extracción de palabras clave y los apellidos de los autores presentes en la base de trabajos publicados.

Los resultados corroboran la existencia de dos frentes de investigación dentro de la institución.

**Palabras clave:** Análisis bibliométrico, Análisis de redes sociales, Co-ocurrencia de palabras, Producción científica, Frentes de investigación.

## Introducción y marco teórico

Los indicadores bibliométricos son datos numéricos que se obtienen a partir de la información presente en los documentos publicados por los investigadores o en las bases de datos que los contienen. Estos indicadores permiten analizar diferentes características acerca de las necesidades de información de estos investigadores, como así también de su producción. La determinación de frentes de investigación lleva a explorar la literatura que produce y que consume una comunidad científica. Puede decirse que un grupo de autores que desarrolla su actividad en una determinada temática produce y consume información similar.

Para analizar los rasgos de similitud temática entre documentos pueden utilizarse diferentes metodologías. El análisis de co-ocurrencia de palabras centra su atención en el contenido de esos documentos. Se entiende por co-ocurrencia de palabras la aparición conjunta de dos términos en un corpus textual dado. Esta técnica se encarga de analizar el

---

1. Artículo enviado al Boletín de la Asociación Argentina de Astronomía y presentado como póster en la 55a Reunión Nacional de la Asociación

contenido de un texto a partir de la ocurrencia conjunta de pares de ítemes, representados por términos o palabras, que permiten identificar relaciones entre conceptos dentro de un determinado dominio. Cuanto mayor sea la frecuencia de aparición conjunta de las palabras, mayor será su vinculación conceptual. (Miguel, 2008).

Puede definirse a la colaboración científica como la interacción entre dos o más científicos dentro de un contexto social de modo de facilitar la distribución y el intercambio de tareas con el fin de lograr un objetivo determinado. (Sonnenwald, 2007).

El análisis de co-autoría da cuenta de los niveles de colaboración científica dentro de una comunidad determinada. Es posible detectar e identificar grupos de investigación analizando estas relaciones. Una elevada frecuencia de co-ocurrencia de autores daría cuenta de altos niveles de colaboración.

Las relaciones de co-ocurrencia pueden estudiarse también a partir de las técnicas de análisis de redes sociales (ARS) lo cual facilita la detección y visualización de las relaciones existentes entre las unidades, y su estudio desde un punto de vista funcional. El ARS es una nueva perspectiva conceptual y metodológica del campo de las ciencias sociales y del comportamiento, que a diferencia del enfoque tradicional de análisis atributivo de los elementos de un grupo, centra su atención en el análisis de las relaciones existentes entre esos elementos. Los datos son relacionales, entendiendo por dato relacional un vínculo específico existente entre un par de elementos. Desde esta perspectiva, la relación entre un par de elementos es una propiedad del par (es decir de la relación) y no una cuestión inherente a las características individuales de cada uno de los elementos relacionados. Una red social está conformada por al menos dos componentes: los actores, que son las unidades o grupos de unidades que se relacionan entre sí por algún motivo, finalidad o circunstancia, y las relaciones, que son los lazos o vínculos existentes entre los actores que conforman la red.

## **Metodología**

Se seleccionaron las palabras clave y los autores presentes en la base de datos de trabajos publicados del IAR correspondientes al periodo 2001-2011 realizándose dos análisis, uno para el periodo completo y otro para 2010-2011.

Se recuperó cada término con su correspondiente número de trabajo, lo cual permitió construir una tabla de frecuencias absolutas. A partir de estos datos y mediante tablas relacionales se construyó una matriz de frecuencias absolutas de co-ocurrencia en la que aparecían los términos duplicados. El proceso final consistió en realizar una consulta a la base relacional donde se podría normalizar la frecuencia de co-ocurrencias eliminando las duplicaciones.

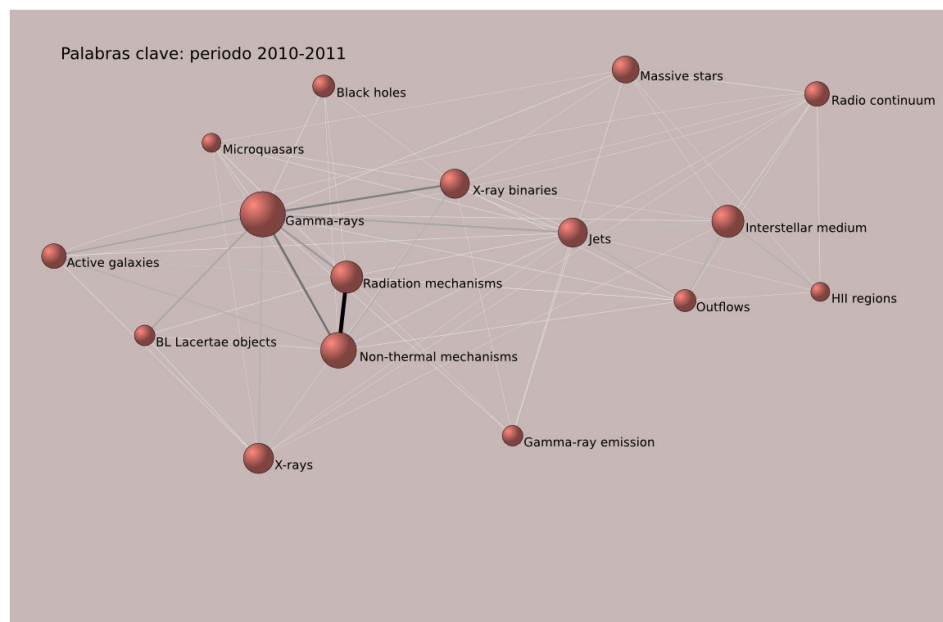
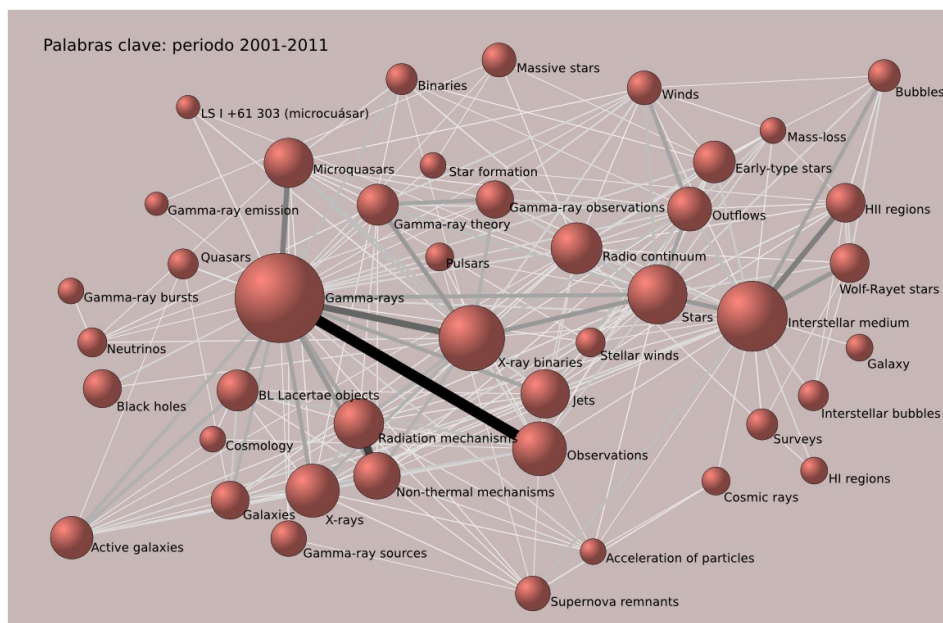
Para generar y visualizar las redes se utilizó el software *Pajek*. Esta aplicación toma como fuente de entrada las matrices de datos generadas por los procedimientos antes mencionados.

## Resultados

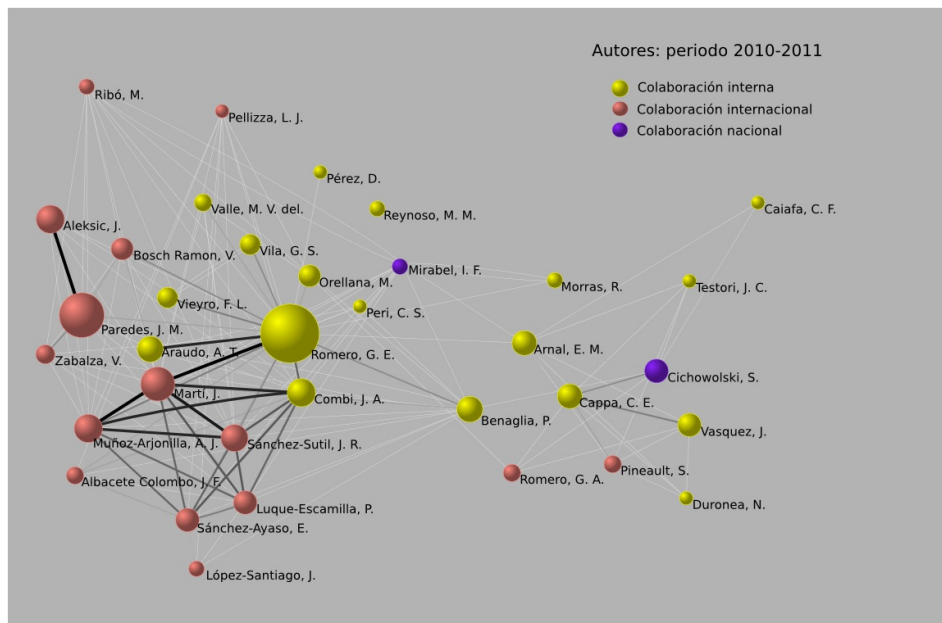
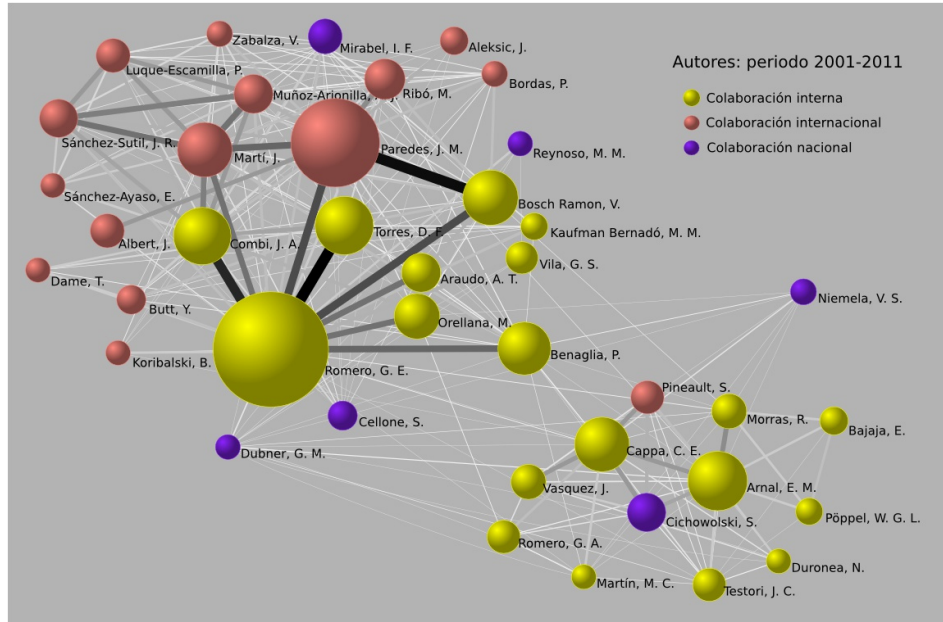
A partir del análisis de redes sociales es posible estudiar si existe alguna correspondencia entre los distintos grupos de investigadores y determinadas temáticas.

Pudo determinarse que las palabras clave con mayor peso y mayor poder de intermediación y cercanía son *Gamma-rays*, *Interstellar medium*, *Radiation and Nonthermal mechanisms*, *Jets*, *Stars*, *X-rays*, *Radio continuum*, *X-ray binaries* y *Early-type stars*. Ello significa que son las palabras clave más representativas de la producción del IAR, al menos desde la perspectiva de su frecuencia de aparición y de su rol en el contexto de la red de palabras claves basada en el análisis de co-ocurrencias.

Estos datos permiten distinguir a grandes rasgos dos grandes grupos de palabras clave que se asocian con las dos grandes líneas de investigación actualmente vigentes en el Instituto: astrofísica de altas energías y medio interestelar.



En el caso de los autores esta correspondencia se reitera, dado que los que tienen una mayor presencia en la base de datos, son justamente los investigadores que encabezan esos grupos como es el caso de G. Romero y M. Arnal. Puede observarse que Arnal posee un grado nodal menor que otros autores como Benaglia o Combi, no obstante se mantiene como líder en su propio grupo.



Nodo	Autor	Grado nodal	Intermediación	Cercanía
33	Romero, G. E.	31	0.2823445	0.8095485
6	Benaglia, P.	24	0.1451954	0.704607
13	Combi, J. A.	19	0.0196313	0.5853659
20	Martí, J.	19	0.015978	0.5853659
27	Paredes, J. M.	18	0.0253621	0.5853659
22	Mirabel, I. F.	17	0.0360827	0.61369
4	Arnal, E. M.	16	0.0818703	0.61369
15	Dubner, G. M.	16	0.044985	0.5853659
24	Muñoz Arjonilla, A. J.	16	0.0069029	0.595409
31	Ribí, M.	16	0.0067106	0.595409
19	Luque Escamilla, P.	15	0.0040728	0.543554
35	Sánchez Sutil, J. R.	15	0.0040717	0.5514316
38	Torres, D. F.	15	0.0110073	0.595409
8	Besch Ramon, V.	14	0.0072296	0.543554
10	Cappa, C. E.	14	0.043687	0.5045122
37	Testori, J. C.	13	0.0477291	0.5764967
7	Bordas, P.	12	0.0031642	0.5212162
11	Cellone, S.	11	0.0027234	0.5073171
23	Morras, R.	11	0.00296198	0.5678922
41	Zabalza, V.	11	0.0002331	0.5006418
9	Butt, Y.	10	0.002296	0.5141727
21	Martín, M. C.	10	0.0062973	0.5514316
26	Orellana, M.	10	0.0321502	0.5358983
25	Niemela, V. S.	8	0.0081638	0.5141727
34	Sánchez Ayaso, E.	8	0	0.5006418
12	Cichowolski, S.	7	0.0070574	0.4816301
16	Duronea, N.	7	0.0025328	0.4181185
17	Kaufman Bernadó, M. M.	7	0.0004396	0.49414
18	Koribalski, B.	7	0.0028826	0.5212162
28	Pineault, S.	7	0.0006838	0.4047743
39	Vasquez, J.	7	0.0024158	0.4135737
2	Aleksic, J.	6	0	0.4756098
3	Araudo, A. T.	6	0.0006284	0.4878049
14	Darne, T.	6	0	0.4878049
32	Romero, G. A.	6	0.0055951	0.4373423
30	Reynoso, M. M.	5	0	0.4816301
5	Bajaja, E.	4	0	0.3922555
29	Pöppel, W. G. L.	4	0	0.3922555
40	Vila, G. S.	4	0	0.4816301
1	Albert, J.	2	0	0.4529617

Nodo	Keyword	Grado nodal	Intermediación	Cercanía
1	Gamma rays	36	0.0523942	0.875
2	Interstellar medium	32	0.0427521	0.8076923
8	Radiation mechanisms	31	0.0220202	0.7924528
10	Jets	31	0.0313921	0.7924528
11	Non thermal mechanisms	31	0.0234955	0.7924528
4	Stars	30	0.030004	0.7777778
6	X rays	30	0.0405612	0.7777778
7	Radio continuum	30	0.0370022	0.7777778
3	X ray binaries	28	0.0158651	0.75
14	Early type stars	27	0.0211957	0.7368421
5	Observations	25	0.0171613	0.7118644
12	Outflows	25	0.0140232	0.7118644
21	Gamma ray observations	25	0.0105591	0.7118644
9	Microquasars	24	0.0101654	0.7
15	Gamma ray theory	24	0.0092633	0.6885246
28	Binaries	24	0.0081228	0.7
32	Neutrinos	22	0.0087803	0.6774194
13	Active galaxies	21	0.0109517	0.6666667
25	Winds	21	0.0074684	0.6666667
22	Gamma ray sources	20	0.0076053	0.65625
38	Gamma ray bursts	20	0.0091472	0.6461538
24	Massive stars	19	0.0146071	0.6461538
36	Galaxy	19	0.0202795	0.6363636
31	Stellar winds	18	0.0067816	0.6363636
33	Cosmic rays	18	0.0082652	0.6268657
41	Acceleration of particles	18	0.0052376	0.6363636
23	Supernova remnants	17	0.0042177	0.6268657
34	Pulsars	17	0.0034376	0.6176471
17	Wolf Rayet stars	16	0.0050015	0.6086957
20	Black holes	16	0.0030411	0.6176471
16	BL Lacertae objects	15	0.003398	0.6
18	HII regions	15	0.0048941	0.5915493
19	Galaxies	15	0.0041548	0.6086957
26	Bubbles	15	0.005727	0.6
29	Quasars	15	0.0042942	0.6086957
40	Star formation	15	0.0094705	0.6
42	LS I +61 303	15	0.0004463	0.6086957
30	Interstellar bubbles	14	0.0050626	0.5915493
39	Mass loss	13	0.0014496	0.5833333
43	Gamma ray emission	12	0.0016927	0.5675676
35	HI regions	11	0.0025594	0.5675676
27	Surveys	10	0.0019286	0.56
37	Cosmology	6	0	0.5185185

Puede afirmarse entonces que se han encontrado ciertas correspondencias en los agrupamientos de palabras clave y de autores, lo cual nos ayuda a corroborar la validez de la metodología empleada.

## Conclusiones

A partir de estas técnicas es posible conocer en forma indirecta los hábitos de una determinada comunidad científica y en función de ellos planificar servicios de información o bien detectar características de comportamiento de esa comunidad que permitan describirla o evaluarla.

## Bibliografía

Pajek : Program for Large Network Analysis.  
<http://pajek.imfm.si/doku.php>

Abt, H. A. (1996). How long are astronomical papers remembered? *Publications of the Astronomical Society of the Pacific*, 108(729), 1059-1061.

Astrom, F. (2007). Changes in the LIS research front : time-sliced cocitation analyses of LIS journal articles, 1990-2004. *Journal of the American Society for Information Science and Technology*, 58(7), 947-957.

Beaver, D. d., & Rosen, R. (1978). Studies in scientific collaboration. Part I: teh profesional origins of scientific coautorship. *Scientometrics*, 1(1978), 65-84.

Boeris, C. E. Aplicación de métodos bibliométricos a la evaluación de colecciones : El caso de la Biblioteca del Instituto Argentino de Radioastronomía, 2011.  
<http://www.iar.unlp.edu.ar/biblio/htdocs/artic/tesis/eval%5Fboeris%5F11.pdf>

Bordons, M. (2001 ). Aspectos metodológicos en la obtención de indicadores bibliométricos. *Cuadernos De Indicios*, (Jun), 17-26.

Börner, K., Chen, C., & Boyack, K. W. (2005). Visualizing knowledge domains. *Annual eview of Information Science and Technology*, 37(1), 179-255.

Eck, N. J. v. , & Waltman, L. (2009). How to normalize cocurrence data?: an analysis of some wellknown similarity measures. *Journal of the American Society for Information Science and Technology*, 60(8), 1635-1651.

Hanneman, R. A., & Riddle, M. (2005) Introduction to social network methods.  
<http://www.faculty.ucr.edu/~hanneman/nettext/>

He, Q. (1999). Knowledge discovery through co-word analysis. *Library Trends*, 48(1), 133-159.

Jarneving, B. (2005). A comparison of two bibliometric methods for mapping of the research front. *Scientometrics*, 65(2), 245-263.

Katz J. Sylvan, & Martin, B. R. (1997). What is research collaboration? *Research Policy*, 26, 1-18.

Lancaster, F. W. (1996). *Evaluación de la biblioteca*. Madrid: Anabad.

Miguel, S., Caprile, L., & Jorquera-Vidal, I. (2008). Análisis de co-términos y de redes sociales para la generación de mapas temáticos. *El Profesional De La Información*, 17( 6), 637-646.

Perianes Rodríguez, A., Olmeda Gómez, C., & Moya Anegón, F. (2009). Detecting, identifying and visualizing research groups in co-authorship networks. *Scientometrics* .

Persson, O. (1994). The intellectual base and research fronts of JASIS 1986-1990. *Journal of the American Society for Information Science*, 45(1), 31-38.

Price, D. J. d. S. (1965). Networks of scientific papers. *Science*, 149, 510-515.

Sanz Casado, E., & Marín Moreno, C. (1998). Aplicación de técnicas bibliométricas a la gestión bibliotecaria. *Investigación Bibliotecológica*, 12(24), 24-39.

Sonnenwald, D. (2007). Scientific Collaborations. *Annual Review of Information Science and Technology* (Vol. 41pp. 643-681).

Spinak, E. (1998). Indicadores bibliométricos. *Ciencia Da Informação*. 27(maio/ago), 141-148.