

PRESENTACIÓN MURAL

Aplicación de técnicas de análisis de redes sociales y de co-ocurrencia de palabras en la determinación de frentes de investigación

C. E. Boeris¹

(1) *Instituto Argentino de Radioastronomía (CONICET-CCT La Plata)*

Abstract. A bibliometric study of the scientific production of the IAR researchers has been performed, with the aim of determining the institute's research fronts and groups of researchers working on these fronts. Methods of analysis of co-occurrence of words, authorship analysis and social network analysis (SNA) has been applied by extracting keywords and the names of the authors on the base of published works. The results support the existence of two research fronts within the institution.

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.

1. Introducción y marco teórico

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. Es posible encontrar rasgos de similaridad temática entre estos documentos mediante el análisis de co-ocurrencia de palabras.

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 contenido de un texto a partir de la ocurrencia conjunta de pares de ítems, 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). 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 visua-

lización de las relaciones existentes entre las unidades, y su estudio desde un punto de vista funcional. Los datos son relacionales, entendiendo por dato relacional un vínculo específico existente entre un par de elementos. 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.

2. Metodología

Se seleccionaron las palabras clave presentes en la base de datos de trabajos publicados del IAR, y los autores de los trabajos correspondientes al periodo 2001-2011. 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 normalizó la frecuencia de co-ocurrencias eliminando las duplicaciones. Para generar y visualizar las redes se utilizó el software Pajek¹. Esta aplicación toma como fuente de entrada las matrices de datos generadas por los procedimientos antes mencionados.

3. Resultados

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* (Fig. 2). 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 (Tabla 2). 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 (Fig. 1). Se observa que Arnal posee un grado nodal menor que otros autores como Benaglia o Combi, no obstante se mantiene como líder en su propio grupo (Tabla 1).

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.

¹Pajek : *Program for Large Network Analysis*. <http://pajek.imfm.si/doku.php>

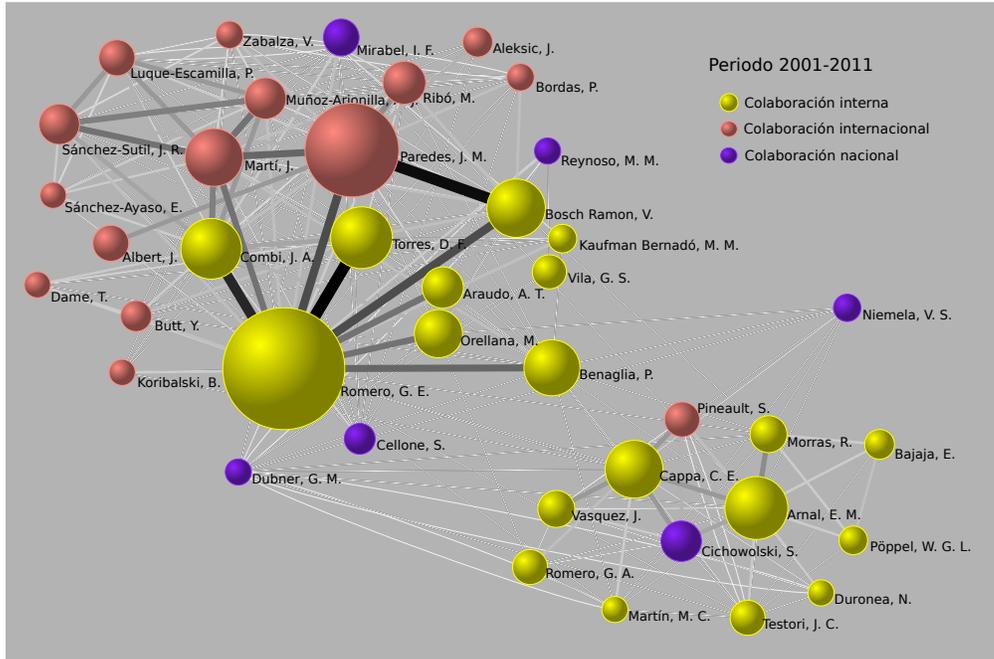


Figura 1. Red de autores

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

Tabla 1. Indicadores de la red de autores

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

Tabla 2. Indicadores de la red de palabras clave

