Universidad Nacional de La Plata
Facultad de Humanidades y Ciencias de la Educación
Departamento de Bibliotecología

Aplicación de métodos bibliométricos a la evaluación de colecciones: el caso de la Biblioteca del Instituto Argentino de Radioastronomía

Tesina para optar por el grado de Licenciada en Bibliotecología y Ciencia de la información

Claudia E. Boeris
Legajo: 64348/1

Director: Dra. Sandra Miguel
Co-director: Prof. César Archuby

La Plata, 2010

Agradecimientos

Quisiera llevar estas palabras de agradecimiento a todas las personas que me ayudaron y me acompañaron en este camino.

A mis directores de tesina, Sandra y César. A Sandra por encontrar siempre la respuesta justa a cada una de mis consultas y por aportar su visión crítica sobre mi trabajo. A César por enseñarme que un bibliotecario no debe ver a la matemática y la informática como misterios inalcanzables solo comprensibles por unos pocos afortunados, sino como poderosas herramientas que complementan el trabajo cotidiano.

También quiero agradecer a Ricardo Morras y a Marcelo Arnal por el apoyo que me dieron desde la dirección del IAR para lograr este objetivo. A Ricardo por dejar que la Biblioteca fuera una especie de conejillo de indias durante mi carrera, y a Marcelo por interesarse y aportar su visión astronómica de la cuestión en numerosas charlas durante este último año.

Mi agradecimiento va también para Federico Bareilles que me acercó al mundo del software libre y que tuvo la paciencia de ayudarme con el LaTeX y varias de las herramientas que utilicé para escribir este texto. A Fernando Hauscarriaga por ayudarme con una parte del procesamiento de las referencias bajadas del ADS, y a Rubén Mansilla que resolvió las urgencias cuando el lenguaje de formateo de ISIS se me escapó de las manos.

A la Profesora Ana Príncipi, de la carrera de letras, con quien tuve la oportunidad de compartir un Seminario transversal de redacción de tesinas y me ayudó a delimitar qué cosa era y qué cosa no era este trabajo.

Finalmente quiero a gradecer a mi familia y a mis amigos por el apoyo y el cariño de siempre.

Resumen

Con el objetivo de determinar el grupo de publicaciones nucleares a considerar en el desarrollo de la colección de la Biblioteca del IAR, se realiza un estudio bibliométrico de la producción y del consumo de literatura científica de los investigadores de la institución a la que la biblioteca pertenece. A partir del análisis de referencias de los trabajos publicados por los investigadores se determinan la obsolescencia y la utilidad de la literatura consultada. Mediante la extracción de palabras clave y de los autores se determinan también los frentes de investigación del instituto y los grupos de investigadores que trabajan en esos frentes, aplicando los métodos de análisis de co-ocurrencia de palabras, coautorías y análisis de redes sociales. Los resultados dan cuenta de una baja obsolescencia para la literatura consultada, de una elevada preferencia para consultar y publicar en dos o tres títulos de publicaciones periódicas de la disciplina, y demuestran finalmente la existencia de dos frentes de investigación dentro de la institución.

Palabras clave: Evaluación de colecciones – Análisis bibliométrico – Ley de Brookes – Modelo de Bradford – Índice de Price – Obsolescencia de la literatura científica – Análisis de redes sociales – Publicaciones nucleares – Producción científica – Políticas de selección y adquisiciones.

Índice general

1. Introducción **1**

2. Marco teórico: evaluación de la colección mediante métodos bibliométricos **3**

2.1. Indicadores y técnicas bibliométricas para la evaluación de colecciones 4

2.2. Obsolescencia de la literatura científica: El índice de Price . . 5

2.2.1. La Ley de Brookes 7

2.3. Determinación del núcleo de publicaciones: la Ley de dispersión de Bradford 9

2.4. Métodos para determinar los frentes de investigación 11

2.4.1. Acoplamiento bibliográfico y análisis de co-citación . . 11

2.4.2. Análisis de co-autoría 13

2.4.3. Análisis de co-ocurrencia de palabras 14

2.4.4. Análisis de redes sociales 16

3. Principales antecedentes y objetivos del estudio **21**

3.1. Principales antecedentes 21

3.2. Objetivos 23

4. Metodología **25**

5. Resultados **28**

5.1. El núcleo de revistas: el modelo de Bradford aplicado al consumo y la producción de información de los científicos del IAR 28

5.2. El envejecimiento de la literatura referenciada por el IAR . . . 45

5.3. Determinación de los frentes de investigación del IAR 50

6. Discusión y conclusiones **58**

7. Anexo: Herramientas y técnicas para la recolección y preparación de los datos **65**

7.1. Obtención de referencias 65

7.2. Aplicación de utilitarios Cisis 66

7.2.1. Obsolescencia 66

7.2.2. Núcleo de revistas 67

7.3. Selección de datos relacionales 68

7.4. Análisis de redes: *Pajek* 69

7.5. Formateo del texto 69

7.6. Scripts utilizados 73

Capítulo 1

Introducción

La evaluación es una herramienta de gestión que permite a las instituciones obtener información fiable para la toma de decisiones de distinta naturaleza. En la última década los procesos evaluativos se han venido desarrollando en diferentes organizaciones públicas y privadas. Las bibliotecas no han sido ajenas a estas cuestiones y han debido llevar adelante la evaluación de su gestión con el objetivo de justificar su utilidad en las instituciones de las que dependen.

En nuestro país las bibliotecas de institutos de investigación han sido históricamente relegadas considerándoselas meros reservorios bibliográficos, por tal motivo la evaluación de la gestión en este ámbito puede constituirse en un punto de partida para mejorar la visión que tienen los gestores de las instituciones mayores con respecto a sus bibliotecas.

Existen dos enfoques para llevar a cabo la evaluación. Uno evalúa a la biblioteca como una parte de la organización a la que pertence centrándose en los costos y beneficios que la biblioteca produce a la organización. El otro enfoque se centra en la evaluación de los procesos que se desarrollan dentro de la biblioteca y que tienen que ver con sus colecciones, sus servicios, su personal, los productos que ofrece, etc.

El objetivo de una biblioteca científica es satisfacer las necesidades de información que demandan las actividades de investigación y desarrollo tecnológico de la comunidad científica y técnica a la que pertenece.

Según Lancaster (1996) el objetivo a largo plazo de una biblioteca debería ser la producción de resultados dentro de la comunidad a la que sirve, esos

resultados se optimizarían a través de la gestión adecuada de los recursos.

La colección está compuesta por todos los documentos que la biblioteca proporciona a sus usuarios y comprende tanto documentos en papel como documentos o bases de datos accesibles en línea, los cuales pueden considerarse como uno de los principales recursos con que cuenta para lograr los resultados que se esperan de ella.

El desarrollo de la colección es un proceso sistemático de creación y mantenimiento de la misma que debe darse de acuerdo a una serie de criterios establecidos de antemano. Estos criterios tienen que ver con políticas de selección, adquisición y descarte que deberían expresarse mediante documentos escritos.

Lamentablemente en nuestras instituciones de investigación estas prácticas nunca se han llevado a cabo por lo cual muchas veces se han adquirido materiales innecesarios o bien se ha dejado de comprar otros imprescindibles. Las decisiones arbitrarias de las autoridades de los institutos sumadas a la falta de presupuestos para las bibliotecas pueden verse como el producto de la falta de políticas de desarrollo de colecciones que debería darse desde el seno de las mismas.

Llevar adelante un plan de evaluación de la colección puede constituirse en una herramienta de análisis y diagnóstico que facilite la elaboración de estas políticas y que mejore la calidad de los servicios que brindan las bibliotecas científicas.

Capítulo 2

Marco teórico: evaluación de la colección mediante métodos bibliométricos

Como se mencionó anteriormente el objetivo fundamental de un servicio bibliotecario es el de satisfacer las necesidades informativas de sus usuarios mediante la gestión de recursos de información. Este proceso requiere de la planificación de los servicios de modo tal que la biblioteca ponga en marcha una política de selección, adquisición, registro, organización, disponibilidad, acceso, uso, preservación y descarte de materiales acorde con las metas de la institución. Para saber si se cumplen los objetivos propuestos la biblioteca debe realizar evaluaciones que le permitan contar con información de apoyo a la toma de decisiones en todos los aspectos que hacen a su actividad. En este trabajo nos abocaremos especialmente a la evaluación de la colección.

La colección puede evaluarse aplicando diferentes metodologías. Desde una perspectiva general podemos decir que la evaluación de la colección puede basarse en el estudio de la colección misma o en el análisis de su comportamiento, medido a partir de su uso real o potencial. Los datos para estos análisis pueden recabarse de diferentes maneras según sea la finalidad. En algunos casos, cuando el objetivo es conocer las opiniones que los usuarios tienen sobre la colección suele utilizarse el método de encuesta; en otros casos, cuando lo que se pretende es contar con información objetiva se recurre a otros métodos como el bibliométrico.

La aplicación de este último permite realizar un diagnóstico objetivo de la colección a partir de un conjunto de indicadores, técnicas y modelos de análisis basados en el estudio tanto de las características de la colección, de su uso, como de los patrones de producción y consumo de información que caracterizan a las diferentes displicinas y especialidades científicas, y que determinan preferencias y hábitos particulares de los usuarios .

A continuación se realiza una somera descripción de los principales indicadores y técnicas bibliométricas ampliamente utilizados para el estudio de las colecciones de las bibliotecas, con la finalidad de mostrar las posibilidades que ofrece el método y la riqueza de información tanto cuantitativa como cualitativa que aporta para la toma de decisiones en materia de desarrollo de colecciones.

Luego, en los capítulos 5 y 6 se presentará una aplicación concreta del uso de algunos de estos indicadores y técnicas para la evaluación de colecciones de una biblioteca en particular, la biblioteca del Instituto Argentino de Radiostronomía (IAR) del Conicet.[1]

2.1. Indicadores y técnicas bibliométricas para la evaluación de colecciones

Una de las estrategias ampliamente utilizadas para determinar cómo debería ser una colección que satisfaga la demanda de información de los usuarios de una biblioteca universitaria o de investigación es analizando los patrones de consumo de información de los mismos en tanto productores de conocimiento. Más específicamente nos referimos al análisis de las referencias bibliográficas que incluyen en sus publicaciones y de las citas que reciben.

El análisis de referencias es un método que consiste en el estudio de la bibliografía incluida por los autores en sus trabajos de investigación, con el fin de obtener elementos que demuestren el comportamiento de la comunidad científica estudiada en términos de la tipología de documentos que utilizan, la actualidad o vigencia de esa información, la capacidad idiomática, o las características de su modo de trabajo, entre otros.

[1] Consejo Nacional de Investigaciones Científicas y Técnicas.

Los títulos referenciados en los artículos que producen los investigadores de una institución dan cuenta de las publicaciones que éstos consumen, por tanto contar con esa información facilita planificar servicios y delinear políticas de selección y adquisiciones para esa comunidad de usuarios.

Mediante el análisis de referencias también es posible determinar la actualidad de los documentos que la comunidad de usuarios utiliza, como así también cuáles son las publicaciones nucleares de la disciplina. Para ello la bibliometría ha aportado herramientas de evaluación sustentadas en las leyes de Price y Brookes, y la ley de Bradford.

Otro tipo de estudios que aporta información para planificar servicios y llevar a cabo la gestión de la colección es el que permite conocer el comportamiento de la comunidad de usuarios en relación, no solamente con su consumo de información, sino también con su producción. La determinación de frentes de investigación posibilitaría gestionar la colección de manera más eficiente en tanto se haría sobre necesidades de grupos identificados previamente. El análisis de redes sociales asociado al análisis de coautoría y co-ocurrencia de palabras puede considerarse una estrategia metodológica válida para este tipo de estudios.

2.2. Obsolescencia de la literatura científica: El índice de Price

La información científica va perdiendo vigencia con el paso del tiempo. Este fenómeno puede medirse a través de diversos indicadores. Un estudio que refleje medianamente la realidad debe tener en cuenta tres aspectos fundamentales: la temática estudiada, los usuarios a quienes ese estudio afecta, y el tamaño de la muestra a tomar (Brookes, 1970).

Diferentes investigadores se dedicaron a estudiar la obsolescencia de la literatura científica con el objeto de evaluar las transformaciones que se producen en la misma ciencia. Estos estudios tienen también su aplicación en el ámbito de las bibliotecas desde el momento en que es posible evaluar la trayectoria de una publicación periódica, la producción científica de una institución o autor, o bien dar herramientas para llevar a cabo la gestión de colecciones.

Burton y Kebler (1960) introdujeron el concepto de semiperiodo o vida media para hacer referencia al tiempo durante el cual fue publicada la mitad de la literatura corriente o la literatura referenciada dentro de una disciplina. Si se distribuyen las referencias por año de procedencia, el semiperiodo es la mediana de esa distribución. De acuerdo a la disciplina de la que se trate la vida media o semiperiodo será mayor o menor. Si la vida media es mayor significa que la literatura publicada en esa disciplina se mantendrá vigente por un número mayor de años que para una disciplina con una vida media menor.

A partir de estos conceptos sería posible distinguir dos tipos de literatura (Burton & Kebler, 1960; Price, 1965): la literatura clásica o histórica que podría ser la referenciada más frecuentemente por trabajos de carácter teórico y la literatura corriente que poseería una vida media menor.

Por ejemplo, según el estudio de Burton & Kebler el 75 % de las referencias de ingeniería química fueron publicadas dentro de los últimos diez años de realizado el estudio y el 13 % de las referencias en los diez años anteriores. No ocurre lo mismo en disciplinas como la matemática o la botánica donde casi el 50 % de las referencias son de los primeros diez años y el resto se distribuyen en forma más pareja entre los últimos cincuenta años previos a la realización del estudio.

Price por su parte se dedicó entre otras cuestiones a determinar cuán reciente es la bibliografía usada y cuál es la proporción de literatura clásica y corriente en una disciplina; para ello desarrolló el famoso índice que lleva su nombre.

El **índice de Price** representa la relación que existe entre las referencias de no más de cinco años y el total de las referencias. En disciplinas donde la literatura corriente es más abundante el índice de Price se acerca al 50 % cuando crece la literatura histórica el índice disminuye.

Hay algunas discrepancias en la definición de los datos para calcular dicho índice. Price consideraba razonable tomar la proporción de referencias que corresponden a los últimos cinco años (Price, 1970). Otros autores hablan de tomar las referencias desde el año 0 al 5 de modo que los datos analizados en este caso corresponderían a seis años y no a cinco (Sanz Casado, 1994).

$$IP = (\sum_{k=0}^{5} NR_k)/NR_T * 100 \tag{2.1}$$

Donde:

IP = Índice de Price

NR_k = Número de referencias del año k

NR_T = Número total de referencias

2.2.1. La Ley de Brookes

El modelo matemático desarrollado por Brookes, conocido como la Ley que lleva su nombre, introduce un nuevo elemento, la utilidad de los documentos. En su ley describe la pérdida temporal de la utilidad de un conjunto de documentos.

La utilidad total o inicial **U(0)**, representa el número total de citas que se espera pueda recibir un conjunto de documentos al cabo de infinitos años, o de las referencias que posee al momento de su publicación (Ruiz Baños, 1997). En términos estadísticos esa utilidad representa el 100 % del total. Si se trabaja con frecuencias relativas su valor es 1.

A lo largo de la vida de esos documentos la utilidad inicial se va reduciendo y transformándose en utilidad residual **U(t)**, según una función exponencial decreciente hasta llegar a ser nula en un periodo de infinitos años.

Para calcular la utilidad residual o teórica Brookes propone la siguiente fórmula, que será aplicada a los datos del IAR más adelante en el estudio:

$$U(t) = U(0) * a^t$$

donde:

U(t) = utilidad residual

U(0) = utilidad total o inicial

a = factor de envejecimiento anual

t = edad

El factor de envejecimiento anual, a representa el tanto por uno al que se reduce la utilidad teórica o residual por cada año que pasa. Toma valores entre cero y uno. Cuánto más se acerca al uno la utilidad es mayor. La edad t es la diferencia, en años, entre un año determinado y la fecha en que los documentos fueron publicados.

Para calcular el factor de envejecimiento en este estudio se utilizaron dos métodos: el método de la vida media h y el método del máximo de citas anuales C.

Cálculo del factor de envejecimiento mediante la vida media h

Para aplicar este método es necesario conocer la vida media de los documentos. Para realizar los cálculos se trabajará con frecuencias relativas de citación, entonces la utilidad total o inicial es igual a uno:

$$U(0) = 1 \quad (1)$$

La vida media es la edad en la cual la utilidad, y en consecuencia, el número de citas o referencias, se reduce a la mitad, de modo que la vida media puede calcularse a partir de esta expresión:

$$a^h = 0{,}5 \quad (2)$$

Para calcular h es necesario acumular las referencias desde la edad más antigua hasta la edad cero transformándolas en frecuencias relativas. La columna de utilidades se calcula dividiendo cada referencia acumulada por el total de referencias.

h es la edad en la que la utilidad se reduce a la mitad y como veremos más adelante en nuestro estudio ese valor se da en el año 18.

Cálculo del factor de envejecimiento mediante el máximo de citas anuales

Otro factor que introduce Brookes es el factor de utilidad u, que está relacionado con el total (absoluto) de citas o referencias y puede calcularse a

partir de ese total y el total de citas o referencias del año con mayor cantidad de éstas.

Resulta de:

$u = 1/(1 - a)$ (4)

y

$R = u * C$ (5)

donde:

R = total de referencias

C = referencias del año con mayor cantidad

Sustituyendo la ecuación (4) en la ecuación (5):

$R = C/(1 - a)$

Al despejar se obtiene el factor de envejecimiento, que será aplicado a nuestros datos más adelante.

$a = 1 - C/R$

2.3. Determinación del núcleo de publicaciones: la Ley de dispersión de Bradford

La ley de dispersión de Bradford estudia la distribución de la literatura científica. Bradford (1934) planteó que si se consulta literatura especializada en un determinado tema, la mayor parte de los artículos sobre ese tema será publicada en un reducido número de revistas mientras que el resto de los artículos se distribuirán en un número muy superior de revistas. Su estudio

se enfocó en el análisis de la distribución de artículos en revistas especializadas en geofísica aplicada y lubricación.

Partiendo del análisis de las referencias incluidas en las bibliografías que se prepararon en su biblioteca relativas a estas temáticas, se ordenaron las fuentes contenidas en las mismas según su productividad. Al analizar las cifras obtenidas Bradford observó que en un número reducido de fuentes se reunían la mayor cantidad de artículos por título, en otro grupo intermedio de fuentes se acumulaba una producción moderada, y en un tercer grupo se concentraba una cantidad de artículos por fuente, menor, y esa cantidad disminuía notablemente.

Al analizar los datos en función de su proporcionalidad puede verse que cada uno de los grupos producía aproximadamente la misma cantidad de artículos e iban creciendo en función de un multiplicador que se acercaba a 5. La suma de referencias en un tema dado entonces, sin tener en cuenta las referencias dentro del núcleo, sería proporcional al logaritmo de la suma de fuentes analizadas si éstas fueran ordenadas de acuerdo a su productividad.

Según determinó Bradford el grueso de los artículos publicados sobre un tema dado se concentra en un número reducido de revistas especializadas de esa disciplina, mientras que el resto de artículos se distribuye en revistas que se dedican a temáticas relacionadas y otras de carácter general. Los artículos guardarían una distribución inversa y se ubicarían en zonas que irían desde un núcleo de publicaciones con la mayor cantidad de artículos por publicación hasta una periferia donde se encontraría un número reducido de artículos relacionados, por título.

Bradford formuló su ley de esta forma: *"si las revistas científicas se disponen en orden decreciente de productividad de artículos sobre un tema específico, se puede distinguir un núcleo de revistas más particularmente dedicadas al tema dado y varios grupos o zonas que contienen aproximadamente el mismo número de artículos que el núcleo, las cantidades de revistas en el núcleo y zonas sucesivas serán:* $1 : n : n^2 ....$"

O sea que, si se divide el número total de artículos de modo que el núcleo y las zonas contengan aproximadamente la misma cantidad de artículos, el número de revistas en cada zona aumenta en una progresión geométrica.

El modelo de Bradford ha sido aplicado en la evaluación de las demandas de documentos en servicios de información; del mismo modo es susceptible

de ser utilizado en el análisis de la producción y de las referencias de una determinada comunidad de usuarios.

2.4. Métodos para determinar los frentes de investigación

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 similaridad temática entre documentos pueden utilizarse diferentes metodologías. El análisis de co-citación y el acoplamiento bibliográfico se enfocan en determinar las relaciones que existen en esos documentos a partir de la aparición conjunta de las citas y referencias de los autores. Por su parte, el análisis de co-ocurrencia de palabras centra su atención en el contenido de esos documentos. Las unidades de análisis en cualquier caso son determinadas por el investigador según sea el objetivo del estudio: documentos, autores, revistas, temas, etc. El análisis de redes sociales se complementa con estas técnicas facilitando la detección y visualización de las relaciones existentes entre estas unidades y su estudio desde un punto de vista funcional.

2.4.1. Acoplamiento bibliográfico y análisis de co-citación

Otra de las vías para identificar un frente de investigación es analizando la forma en que los investigadores citan la literatura disponible. Uno de los factores que intervienen en la identificación de un frente es la similaridad que existe entre el cúmulo de documentos citantes y citados.

Desde esta perspectiva es posible identificar dos grupos de documentos. Por un lado la denominada literatura de base que estaría constituida por aquellos documentos que están altamente citados y co-citados, es decir que aparecen frecuentemente citados juntos; y por otro lado la literatura corriente de la especialidad constituida por un grupo de documentos que referencian a aquéllos.

La literatura de base da cuenta del sustrato teórico y metodológico, y los artículos citantes muestran los frentes de investigación del problema investigado al momento de realizada la investigación.

Dos de los métodos ampliamente utilizados con estos fines son el acoplamiento bibliográfico y el análisis de cocitación. Estos métodos se aplican para detectar las relaciones entre documentos, y por extensión entre autores, revistas, disciplinas, etc.) y de esta forma medir el grado de similaridad entre ellos.

Acoplamiento bibliográfico

La noción de acoplamiento bibliográfico fue introducida por Kessler en la década del sesenta (Jarneving, 2005), y fue aplicada como un método para agrupar artículos científicos y técnicos con el objeto de facilitar la provisión de documentos y la recuperación de información.

Se dice que dos artículos están bibliográficamente acoplados cuando poseen al menos una referencia en común. Esto significa que existe una relación entre ambos cuya intensidad estará dada por la cantidad de referencias en común que éstos posean. Esta intensidad es la frecuencia de acoplamiento y se mide con el número n. Cuanto mayor sea n más fuerte será la relación entre los dos documentos.

Se pueden identificar dos criterios:

A. Un número de artículos constituye un grupo **G1** si cada miembro del grupo posee al menos una referencia en común con un artículo testigo **A0**.

B. Un número de artículos constituye un grupo **G2** si cada miembro del grupo tiene al menos un acoplamiento bibliográfico con otro miembro del grupo.

Co-citación

La co-citación es la frecuencia con que dos artículos son citados juntos (Small, 1973). Entonces el grado de co-citación es definido como el número de documentos citantes idénticos.

Si poseemos un grupo de documentos denominado **A** los cuales citan a un documento **a**, y otro grupo de documentos **B** que citan a **b**; la intersección

de **A** y **B** estará dada por los documentos dentro de los grupos que citen a **a** y a **b** al mismo tiempo (Small, 1973).

El número de elementos en la intersección es la frecuencia de co-citación:

$n(A \cap B)$

La frecuencia de co-citación relativa sería:

$n(A \cap B)/n(A \cup B)$

Para que dos artículos estén fuertemente co-citados deben ser citados juntos varias veces por la literatura posterior; esto significa que las relaciones de co-citación están dadas por los autores citantes. De modo que en la medida que cambien los patrones de citación con el tiempo, también cambiarán los términos utilizados en los documentos, lo que se verá reflejado en una modificación en su temática.

A diferencia del acoplamiento bibliográfico que relaciona documentos fuente, la co-citación relaciona documentos citados, por lo tanto puede decirse que tiene una analogía con la medición de descriptores o la asociación de palabras.

La co-citación es una relación que se modifica constantemente, así como la co-ocurrencia de palabras puede variar en la medida que un campo determinado evoluciona. El acoplamiento bibliográfico es una relación fija en tanto depende de las referencias incluidas en los documentos acoplados.

2.4.2. Análisis de co-autoría

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.

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).

Durante el siglo XX después de la segunda guerra mundial los trabajos en colaboración se incrementaron notablemente, pero puede decirse que el real origen de la colaboración científica se remonta a los siglos XVII y XVIII[2] con la aparición de los primeros artículos de revistas científicas destinados a comunicar los nuevos descubrimientos de la revolución industrial. Esta característica da cuenta de la profesionalización de la ciencia por cuanto los científicos trabajan en grupos regidos por un conjunto de reglas que les permiten organizar su trabajo y a su vez ser reconocidos por la sociedad a la que pertenecen (Beaver & Rosen, 1978).

Hay diversos factores que dan origen a esta colaboración y pueden relacionarse con cuestiones puramente científicas como la resolución de un problema complejo difícil de ser abordado por un solo investigador, o con factores socieconómicos como la disponibilidad de instrumental en países menos desarrollados. Las redes sociales y los factores personales intervienen también en la conformación de grupos de investigación.

Disciplinas tales como la ciencia de la información, la psicología, la computación o la sociología han abordado el tema desde perspectivas diferentes. La cienciometría y la bibliometría han estudiado la colaboración científica mediante el análisis de ciertos patrones presentes en la literatura científica.

Es dificultoso cuantificar este fenómeno dado que es producto de una actividad social donde se ponen en juego relaciones humanas; sin embargo los datos presentes en las publicaciones relativos a autoría o afiliación son el sustrato para producir análisis estadísticos e indicadores bibliométricos que den cuenta del comportamiento de la comunidad científica que se quiere estudiar.

Más adelante en este trabajo se abordará el análisis de redes sociales (ARS) como un marco conceptual y metodológico que aporta soluciones para este análisis.

2.4.3. Análisis de co-ocurrencia de palabras

Se entiende por co-ocurrencia de palabras la aparición conjunta de dos términos en un texto dado. Esta técnica se encarga de analizar el contenido de

[2]Es interesante señalar que una de las áreas que han mostrado un mayor nivel de colaboracion durante el siglo XVIII ha sido la astronomía.

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).

Michael Callon realizó el primer estudio sobre análisis de co-palabras, seguido luego por Bruno Latour. Para Callon y Latour los textos ocupaban un lugar preponderante en el desarrollo de la ciencia, con el fin de identificar los actores y explicar la dinámica global de una disciplina (He, 1999). Estos métodos comenzaron a ser utilizados más tarde como métodos de recuperación de información. Recientemente fueron propuestos por algunos autores como medios para representar mapas temáticos de las colecciones de bibliotecas digitales (Miguel, 2008).

A partir del análisis de co-ocurrencia de términos o palabras, es posible identificar las relaciones existentes entre los conceptos de un determinado dominio, o los individuos de un determinado grupo. Esta técnica puede aplicarse a cualquier corpus textual, como puede ser una base de datos.

Si se extraen las palabras clave, los descriptores o las palabras de los títulos de los documentos contenidos en la base es posible establecer por un lado temáticas tratadas por los autores, y por el otro las relaciones que existen entre esas temáticas. Del mismo modo pueden analizarse las veces que un autor determinado aparece mencionado conjuntamente con otros.

Para realizar el análisis de co-ocurrencia hay que seguir una serie de pasos que consisten en extraer las palabras representativas del texto o los términos a analizar de la base de datos, para luego confeccionar una matriz de co-ocurrencia que permita realizar una análisis posterior. Este análisis se basa en el rol que ocupa cada palabra según su jerarquía, su peso relativo o la dinámica con la que ésta se relaciona con otras palabras de la matriz.

La calidad de la indización tanto como la cobertura de la base de datos utilizada son determinantes para la obtención de una red que refleje la realidad. Es necesario contar con palabras que reflejen fielmente el contenido de los documentos.

Para realizar la extracción de términos se presentan dos posibilidades con sus ventajas y sus desventajas. Por un lado la utilización de términos de indización asignados por el indizador, como palabras clave o descriptores; y por otro la utilización de palabras presentes en los títulos.

El proceso de indización es subjetivo y sujeto al análisis y la conceptualización que el indizador haga del texto analizado. Con respecto al uso de palabras de los títulos, es una técnica que asegura la representatividad de las palabras, no obstante se debe realizar un proceso de normalización de las mismas eliminando variantes en formas verbales, nominales, sinónimos, homónimos, etc.

2.4.4. Análisis de redes sociales

Las relaciones de co-ocurrencia pueden estudiarse también a partir de las técnicas de análisis de redes sociales (ARS).

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 (Wasserman and Faust, 1998).

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.

Los actores de una red pueden desempeñar diferentes roles en función de su centralidad, prestigio, poder en relación a los demás. De este modo, los vínculos entre ellos pueden ser directos o indirectos, direccionales o no-direccionales, tener diferentes intensidades; aspectos todos que van determinado el tipo de relación existente y el tipo de estructura de red que conforman (Rodriguez, 1995).

El análisis de redes sociales se basa en la teoría de grafos. En un grafo cada actor es representado como un nodo y cada relación como un enlace entre ellos. Cada uno de esos enlaces es un episodio, un evento, algo que sucede dentro de la red, por lo cual puede decirse que la red es un conjunto de episodios o eventos. En el caso de dos autores que escriben juntos un artículo, ese trabajo en común (ese enlace) constituiría un episodio de relación

social. Cada uno de los enlaces tiene un valor que demuestra la intensidad de la relación entre los nodos entonces será más fuerte la relación entre dos autores que poseen un alto número de trabajos compartidos.

Un grafo puede representar un solo tipo de relación entre los actores (relación simple) o más de una relación (relación múltiple). Las relaciones pueden ser también dirigidas y no dirigidas o de co-ocurrencia. Las relaciones dirigidas son representadas mediante flechas, las de co-ocurrencia mediante segmentos.

Los actores en la red pueden ser personas, instituciones, términos o cualquier elemento de un universo que necesite ser estudiado. En los siguientes ejemplos pueden observarse las relaciones mencionadas:

Para los autores 1 y 2:

- Relación dirigida

  citado por

  1 es citado por 2

  2 puede no ser citado por 1

- Relación de co-ocurrencia

  publica con

  1 publica con 2

  2 publica con 1

Para los términos A, B y C:

A, B y C aparecen en el mismo documento

Término A Término B Término C

Sumada a la teoría de grafos el álgebra de matrices se utiliza también para colectar los datos de la red. Se asignan valores a pares de actores de modo de representar las relaciones entre ellos mediante una matriz. Una matriz es una tabla donde se ubican los actores en las filas y las columnas de manera tal que cada celda de la tabla represente una relación entre ellos. Las matrices utilizadas en el análisis de redes sociales son cuadradas pues tienen el mismo número de filas que de columnas.

El tamaño de la red estará dado por la cantidad de nodos y de enlaces que la conformen (Hanneman & Riddle, 2005).

Se incluyen a continuación ejemplos de matrices:

Cuadro 2.1: Publica con

	Autor 1	Autor 2	Autor 3	Autor 4
Autor 1				
Autor 2				
Autor 3				
Autor 4				

Cuadro 2.2: Publica N trabajos con

	Autor 1	Autor 2	Autor 3	Autor 4
Autor 1				
Autor 2				
Autor 3				
Autor 4				

Medidas de centralidad

Existen ciertos indicadores que permiten analizar la posición que ocupa cada actor dentro de la red, en relación con el resto de los actores. Estas medidas son: el grado nodal, la cercanía y la intermediación.

Grado nodal

En una red no dirigida el grado nodal representa el número de enlaces relacionados con un nodo dado. En redes dirigidas hay que distinguir entre los enlaces entrantes y los enlaces salientes que representan el grado de influenciar o de ser influenciado dentro de la red.

Cercanía (closeness)

La cercanía se mide en el conjunto de la red. El grado de cercanía se mide en función de las relaciones entre los nodos.

Grado de intermediación (betweeness)

Se dice que un nodo cumple una función de intermediación cuando funciona como mediador entre dos actores. Por ejemplo, A se relaciona con C a través de B.

Las técnicas hasta aquí descriptas permiten conocer los hábitos de producción o de consumo de información de los investigadores, la vigencia de

la literatura que citan, y las temáticas en las que investigan. Estos datos brindan una serie de pautas a considerar en la gestión de la colección. Por ejemplo, una elevada frecuencia de referencias que pertenecen a una determinada revista es un indicador indiscutible de que esa publicación debe estar en el grupo nuclear a adquirir. Por otro lado un alto nivel de concentración de determinadas palabras clave presentes en los documentos daría cuenta de las temáticas predominantes en la producción y el consumo de información y en función de ellas podría realizarse la selección de bibliografía sobre esas temáticas. El conocimiento de los grupos de investigación que trabajan en un tema, así como la determinación de los temas en los que trabajan a partir del análisis de coautorías y de co-ocurrencia de palabras son estrategias que facilitan la organización de recursos y servicios orientados a grupos específicos de usuarios, favoreciendo la racionalidad y la eficiencia de las actividades bibliotecarias.

Capítulo 3

Principales antecedentes y objetivos del estudio

3.1. Principales antecedentes

La literatura publicada sobre la producción y el consumo de información en astronomía data de los últimos veinticinco años y se ocupa de analizar fundamentalmente la distribución de las referencias por tipo de publicación, y de realizar estudios específicos por títulos de revistas. La bibliografía consultada permite deducir que los astrónomos prefieren citar publicaciones periódicas sometidas a evaluación por pares, no obstante esto, las citas a proceedings de congresos han aumentado notablemente: Vagiswari & Louis (1995) refieren un aumento del 150 % en la citación a proceedings de conferencias entre 1980 y 1990. El nivel de citación a estos tipos de documentos no llega a compararse con las recibidas por las publicaciones periódicas, dado que muchas veces los proceedings, por ser menores en extensión, suelen no incluir los datos completos de la investigación y los autores los descartan.

Según Abt, H. (1995) las citas a artículos de publicaciones periódicas representan un 78 % del total en la literatura, de las cuales el 65,8 % se concentra en las publicaciones nucleares de la disciplina (*Astronomy & Astrophysics* y *Astrophysical Journal* se llevan la mayor parte de los resultados). Le siguen en importancia las monografías y los proceedings de congresos con el 12,5 %. El resto de las citas se divide entre publicaciones de observatorios con el 1,8 %, preprints con el 2,9 %, tesis y disertaciones con el 1,3 y otras publicaciones.

Un estudio posterior de Henneken et al. (2007) demuestra que ha aumentado notablemente el uso de servicios de preprints, dado que la mayor parte de los artículos publicados en astronomía, han sido enviados previamente a *e-prints* o a *ArXiv.org*

Los estudios de Abt son posteriormente reafirmados por Gómez (2003), quien detecta que el 78 % de las citas se concentran en publicaciones periódicas, el 11,2 % en libros, el 2 % en *newsletters* y publicaciones de observatorios, el 2,6 % en datos astronómicos, el 4 % en material no publicado, y en fuentes no identificadas el 2 %. La mayor parte de las publicaciones periódicas citadas son revistas nucleares de la disciplina.

Isaksson (2007) por su parte sostiene que el 43 % de las citas en Finlandia se concentran en *Astronomy & Astrophysics*, revista referida en los estudios anteriores como la que más aparece en las referencias.

Con respecto a los índices de obsolescencia, Abt (1996), basándose en un estudio sobre dos revistas nucleares como son *Astrophysical journal y Astrophysical journal Supplement* , afirma que la vida media de los artículos observacionales es de 35 años, mientras que la de los artículos teóricos baja a 22 años, para esa publicación.

No se han encontrado en la bibliografía consultada estudios específicos relativos al análisis bibliométrico por áreas temáticas, por lo cual este relevamiento podría resultar un antecedente para futuros trabajos. No obstante, Abt incluye algunos datos en su estudio que demuestran la tendencia en el consumo de información en relación a este tema. Según estos datos, en 1954 los artículos relativos a sistema solar representaban el 28 % mientras que los relativos a medio interestelar eran del 16 % y del 6 % para objetos extragalácticos. En 1996, fecha del estudio, el sistema solar está representado por un 7 %, el medio interestelar por un 18 % y los objetos extragalácticos por un 42 %.

Finalmente algunos artículos analizan la producción en relación con los instrumentos utilizados (tipos de telescopios), y métodos utilizados (observacionales y/o teóricos), como así también estudios acerca de la longitud de los artículos y de la cantidad de autores.

Por otro lado, con respecto a las herramientas utilizadas para llevar a cabo los estudios, se puede observar un crecimiento en la utilización del *Astrophysical Data System (ADS)* como fuente de referencia utilizada por los

astrónomos, y por los autores que han realizado estudios métricos en la disciplina.

De lo expuesto puede verse que hay varias líneas de investigación no exploradas dentro de los estudios bibliométricos realizados en la disciplina, especialmente en el análisis de citación por áreas temáticas. También es importante mencionar que se han encontrado escasos ejemplos de trabajos publicados en nuestro país.

Para concluir, y tomando como base los estudios mencionados, puede considerarse válido el análisis basado en el consumo de publicaciones periódicas de esta comunidad de usuarios.

3.2. Objetivos

La Biblioteca del IAR posee dos grupos de usuarios bien definidos: investigadores científicos, y tecnólogos. Cada grupo posee características y comportamientos propios relacionados con el consumo de información.

Los investigadores científicos en general se conducen en forma independiente, realizando sus búsquedas bibliográficas por sí solos y solicitando a la biblioteca referencias concretas que no han obtenido por sus propios medios. Están acostumbrados a consultar bases de datos por lo cual rara vez solicitan ayuda del bibliotecario. Como parte de su trabajo cotidiano tienen la obligación de publicar, motivo por el cual la cantidad de artículos de su autoría en los registros del IAR es notablemente superior a la del grupo de tecnólogos, que curiosamente es mayor en número que el de científicos.

Por su parte, los tecnólogos[1] están particularmente abocados al desarrollo de dispositivos y sistemas, por lo que rara vez producen artículos para publicar. Sus trabajos generalemnte forman parte de la literatura gris que integra la colección de informes técnicos internos de la institución. Por este motivo se ha seleccionado al grupo de científicos como comunidad interna a estudiar.[2]

El trabajo se basa en la identificación de los patrones de consumo y producción de información bibliográfica por parte de los científicos de la institución para contar con información objetiva de apoyo a la toma de decisiones en

[1] técnicos, ingenieros e informáticos.

[2] Sería deseable realizar un estudio similar para el grupo de tecnólogos.

el desarrollo de la colección especializada en astronomía. Para ello se realizará un análisis de las referencias y de los trabajos publicados por dicho grupo de científicos con el objeto de determinar la actualidad de la bibliografía que utilizan, las revistas que consultan y los frentes de investigación que puedan estar asociados a dichas revistas.

Capítulo 4

Metodología

Para realizar el estudio se trabajó con los artículos publicados por los investigadores del IAR publicados entre 1966 y 2008.

Como fuente de datos se utilizó el *NASA Astrophysics Data System (ADS)*[1], una base de datos de cobertura internacional especializada en Astronomía, Astrofísica e Instrumentación, que cubre prácticamente toda la producción científica de la disciplina; y la base de datos de Trabajos publicados[2] que posee el IAR, de cobertura local.[3]

Se bajaron del ADS un total de 12.470 referencias presentes en esos trabajos, las cuales fueron convertidas en una base de datos ISIS[4].

Se eliminaron actas de congresos, monografías y cuatro publicaciones periódicas que contenían solamente los resúmenes de los trabajos: el Boletín de la AAA, la Revista mexicana de A&A, el *Bulletin of the AAA* y el *Boletim da Sociedade Astronomica Brasileira*. Fueron eliminados además 184 registros cuyas referencias no estaban presentes en la base de datos (ADS) pues no era un número significativo en el total de artículos. No obstante se observó que estos registros se distribuían en forma pareja a lo largo del periodo estudiado y pertenecían a diferentes publicaciones cuya mayoría eran proceedings de congresos o abstracts del Boletín de la Asociación Argentina de Astronomía.

[1] http://adswww.harvard.edu/

[2] también denominada base de contribuciones.

[3] http://www.iar.unlp.edu.ar/biblio/cgi-bin/opacmarc/wxis?IsisScript=opac/xis/opac.xis&db=contri

[4] http://www.unesco.org/webworld/isis/isis.htm

Del conjunto de registros seleccionados se extrajeron las palabras clave, los autores y las referencias presentes en esos artículos. Se analizó el total de referencias contenidas en los trabajos publicados por el IAR en ese periodo. De las referencias se trabajó con los títulos de las revistas fuente, la fecha de publicación de cada referencia, el tipo de publicación, y mediante un script se relacionó cada referencia con el número del trabajo correspondiente en la base del IAR.

Se determinó el grupo nuclear de publicaciones periódicas a través del modelo de Bradford, el grado de obsolescencia de la literatura consultada aplicando el Método de Brookes y el Índice de Price.

Para determinar la obsolescencia se listaron los títulos de las revistas que aparecían en la base de datos creada a partir de las referencias, luego se recuperaron en una planilla de cálculo y se determinaron las frecuencias de cada título. Se ordenaron los títulos de acuerdo a su productividad (tablas 5.1 y 5.2). Para calcular el índice de Price se aislaron las fechas de publicación de las referencias.

Para determinar el grupo nuclear de publicaciones se trabajó sobre dos segmentos de registros de la base de datos del IAR: los trabajos publicados durante los primeros 15 años desde la fundación del Instituto, y los publicados durante los últimos 10 años. Se eliminaron las referencias de actas de congresos y monografías. Con el objetivo de determinar si había diferencias significativas entre los patrones de consumo de información y los patrones de producción, se aplicó el modelo de Bradford por un lado a los trabajos publicados de los dos periodos mencionados y por otro a las referencias incluidas en dichos trabajos publicados durante los dos periodos.

Finalmente se identificaron los frentes de investigación combinando el análisis de co-ocurrencia de palabras con el análisis de redes sociales.

Para realizar el análisis de co-ocurrencia de términos se seleccionaron las palabras clave presentes en la base de datos de trabajos publicados y los autores de los trabajos correspondientes al periodo 1999-2008. Con respecto a los autores se trabajó con dos listas, una que incluía a los investigadores del IAR exclusivamente y otra que incluía a todos los investigadores presentes en los trabajos, pertenecieran o no al IAR.

Para el análisis de co-términos se recuperó cada término con su correspondiente número de contribución, 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 podía normalizar la frecuencia de co-ocurrencias eliminando las duplicaciones.

Por otra parte se normalizaron las palabras clave eliminando sinónimos y asimilando términos muy específicos con poca frecuencia de aparición a términos genéricos, pero sin desvirtuar la indización original. Por ejemplo, *ISM* fue reemplazado por *Medio interestelar* y *Active galactic nuclei* fue reemplazado por *Active galaxies*.

Para lograr una mejor visualización de las redes resultantes de los análisis anteriores fue necesario reducir la cantidad de palabras clave y autores, respectivamente; para ello se seleccionaron los términos, fuesen autores o palabras clave, que poseían una frecuencia mayor que 9 dentro de la base de datos y se trabajó con ese subconjunto.

Capítulo 5

Resultados

5.1. El núcleo de revistas: el modelo de Bradford aplicado al consumo y la producción de información de los científicos del IAR

En las tablas 5.1, 5.2, 5.5 y 5.6 se incluyen los datos utilizados para aplicar el Modelo de Bradford al consumo y a la producción de información de los investigadores del IAR.

En cada una de las tablas la columna 1 registra la cantidad de revistas que incluyen una determinada cantidad de referencias, las cuales se ubican en la celda correspondiente de la columna 2. La columna 3 es la suma acumulada de la cantidad de revistas (Col. 1). La columna 4 es la suma acumulada de las referencias de la columna 2, que se calculan previo multiplicar cada valor de la columna 2 por el correspondiente de la columna 1. La columna 5 da el logaritmo común de los números de la columna 3.

Según los datos obtenidos puede verse claramente que la mayor parte de las referencias en ambos periodos pertenecen a dos revistas: *Astrophysical journal* y *Astronomy and Astrophysics*, ubicándose en un tercer lugar *Monthly Notices of the R.A.S.* Nótese que en los primeros 15 años hay una importante cantidad de referencias del *Bulletin of the Astronomical Institutes of the Netherlands*, revista que se fusionó para formar *Astronomy and Astrophysics* (Ver Tabla 5.8).

Cuadro 5.1: Primeros 15 años (Referencias)

Revistas	Refer.	Acum rev	Acum refer.	Log acum rev
1	124	1	124	0
1	105	2	229	0.3
1	48	3	277	0.48
1	41	4	318	0.6
1	40	5	358	0.7
1	30	6	388	0.78
1	27	7	415	0.85
1	20	8	435	0.9
1	13	9	448	0.95
1	12	10	460	1
1	8	11	468	1.04
2	7	13	482	1.11
2	6	15	494	1.18
2	5	17	504	1.23
1	4	18	508	1.26
2	3	20	514	1.3
1	2	21	516	1.32
15	1	36	531	1.56

Cuadro 5.2: Ultimos 10 años(Referencias)

Rev	Refer.	Acum rev	Acum refer.	Log acum rev	Log acum refer.
1	2114	1	2114	0	3.33
1	1886	2	4000	0.3	3.6
1	617	3	4617	0.48	3.66
1	458	4	5075	0.6	3.71
1	321	5	5396	0.7	3.73
1	310	6	5706	0.78	3.76
1	204	7	5910	0.85	3.77
1	151	8	6061	0.9	3.78
1	111	9	6172	0.95	3.79
2	80	11	6332	1.04	3.8
1	74	12	6406	1.08	3.81
1	68	13	6474	1.11	3.81
1	58	14	6532	1.15	3.82
1	47	15	6579	1.18	3.82
1	46	16	6625	1.2	3.82
1	35	17	6660	1.23	3.82
1	33	18	6693	1.26	3.83
1	32	19	6725	1.28	3.83
1	29	20	6754	1.3	3.83
1	28	21	6782	1.32	3.83
1	25	22	6807	1.34	3.83
1	24	23	6831	1.36	3.83
2	22	25	6875	1.4	3.84
1	21	26	6896	1.41	3.84
1	20	27	6916	1.43	3.84
1	18	28	6934	1.45	3.84
2	17	30	6968	1.48	3.84
2	16	32	7000	1.51	3.85
2	13	34	7026	1.53	3.85
2	11	36	7048	1.56	3.85
2	10	38	7068	1.58	3.85
2	9	40	7086	1.6	3.85
3	8	43	7110	1.63	3.85
3	7	46	7131	1.66	3.85
6	5	52	7161	1.72	3.85
9	4	61	7197	1.79	3.86
9	3	70	7224	1.85	3.86
17	2	87	7258	1.94	3.86
26	1	113	7284	2.05	3.86

Cuadro 5.3: Trabajos publicados (contribuciones)

Revistas donde publicó Arnal	Art.	Revistas donde publicó Romero	Art.
Astronomical Journal	15	Astronomy and astrophysics	42
Astronomy and Astrophysics	12	Astrophysical journal	11
Astronomy and Astrophysics Supplement Series	12	Astrophysics and space science	8
Monthly Notices of the Royal Astronomical Society	4	Monthly notices of the Royal Astronomical Society	7
		Revista mexicana de astronomía y astrofísica Serie de conferencias	7
		Modern physics letters A	6
		Physical review D: particles, fields, gravitation, and cosmology	5
		The astrophysical journal letters	4
		Chinese journal of astronony and astrophysics	3
		Ciencia hoy	3
		International journal of modern physics D	3
		The astronomical journal	3
		Memorie della Societa Astronomica Italiana	2
		New astronomy reviews	2
		Astronomische nachrichten	1
		Astronomy and astrophysics letters	1
		Astronomy and astrophysics Supplement series	1
		Astroparticle physics	1
		General relativity and gravitation	1
		Invenio: revista de investigación académica	1
		Nature	1
		Physics letters B	1
		Physics reports	1
		Science	1

Cuadro 5.4: Referencias

Referencias de Arnal	Art.	Referencias de Romero	Art.
Astronomy and astrophysics	10	Astronomy and Astrophysics	310
Revista mexicana de astronomía y astrofísica Serie de conferencias	10	Astrophysical Journal	60
Monthly notices of the Royal Astronomical Society	5	Science	22
The astronomical journal	3	Physical Review D	14
Astronomy and astrophysics Supplement series	2	Astronomical Journal	11
Ciencia hoy	1	Monthly Notices of the Royal Astronomical Society	11
Revista mexicana de astronomía y astrofísica	1	Physics Reports	11
Unión: revista iberoamericana de educación matemática	1	Astronomische Nachrichten	10
		Modern Physics Letters A	10
		Astrophysics and Space Science	9
		Astronomy and Astrophysics Supplement Series	7
		Chinese Journal of Astronomy and Astrophysics	4
		Nature	4
		Chinese Astronomy and Astrophysics	2
		Astroparticle Physics	1

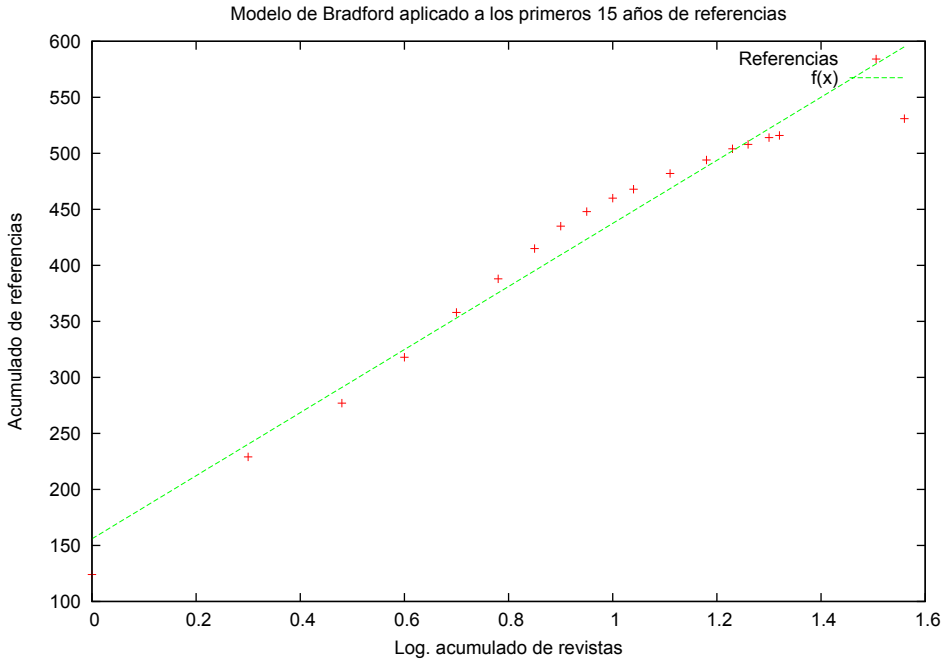


Figura 5.1:

En los gráficos puede advertirse que la dispersión de la distribución de referencias para los trabajos del IAR es opuesta a la dispersión comprobada por Bradford en su estudio. De hecho el 57 % de las referencias de los últimos diez años se concentra en dos títulos, para luego repartirse en forma más o menos homogénea entre el resto de las publicaciones.

Se analizaron además las referencias y los trabajos publicados de los dos autores que encabezan las dos líneas de investigación del IAR (Romero y Arnal), donde pudo reafirmarse la alta presencia de las revistas mencionadas. También aparecen en el caso de Romero, especialmente en las referencias, títulos que corresponden al área de física, hecho que se corresponde con la naturaleza de las investigaciones en las que Romero participa donde hay un fuerte componente de estudios teóricos (Tablas 5.3 y 5.4).

Para comprobar si existen similitudes entre los patrones de consumo de información y los patrones de producción se aplicó también el modelo de Bradford a los artículos publicados por el IAR en los mismos periodos (Tablas 5.5 y 5.6).

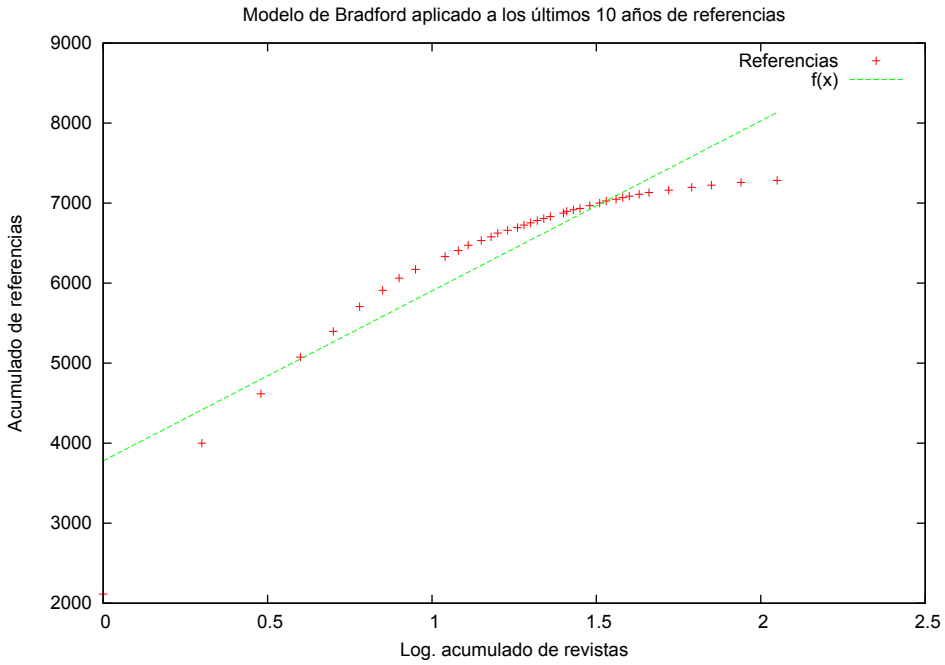


Figura 5.2:

Cuadro 5.5: Primeros 15 años (contribuciones)

Rev	Contrib	Acum rev	Acum contrib	Log acum rev
1	17	1	17	0
1	12	2	29	0.3
1	9	3	38	0.48
1	7	4	45	0.6
1	3	5	48	0.7
4	2	9	56	0.95
10	1	19	66	1.28

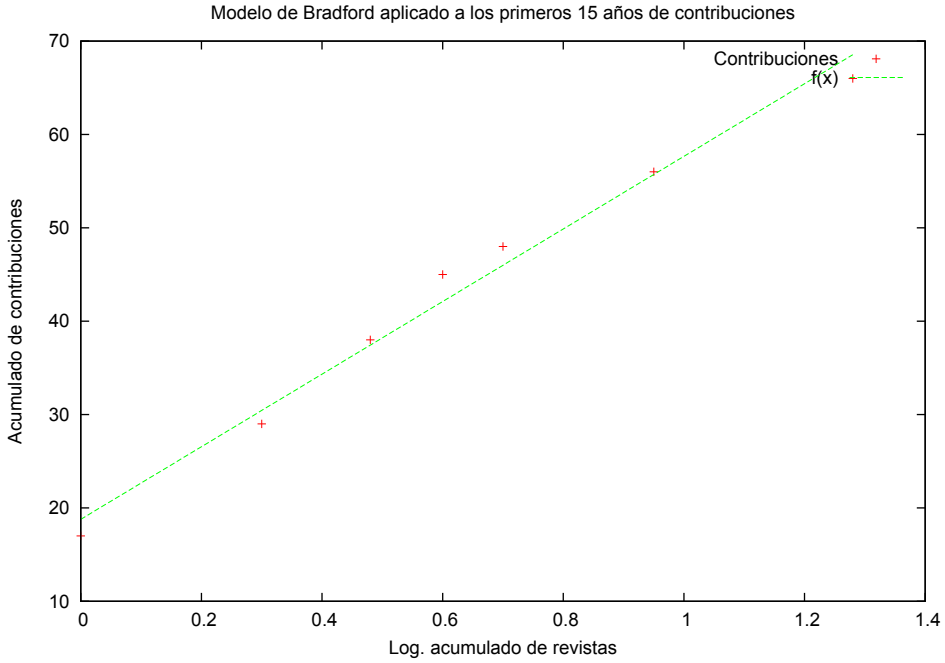


Figura 5.3:

Cuadro 5.6: Ultimos 10 años(contribuciones)

Rev	Contrib	Acum rev	Acum contrib	Log acum rev
1	71	1	71	0
1	34	2	105	0.3
1	19	3	124	0.48
1	14	4	138	0.6
1	12	5	150	0.7
1	8	6	158	0.78
2	7	8	172	0.9
1	6	9	178	0.95
1	5	10	183	1
4	4	14	199	1.15
1	3	15	202	1.18
4	2	19	210	1.28
20	1	39	230	1.59

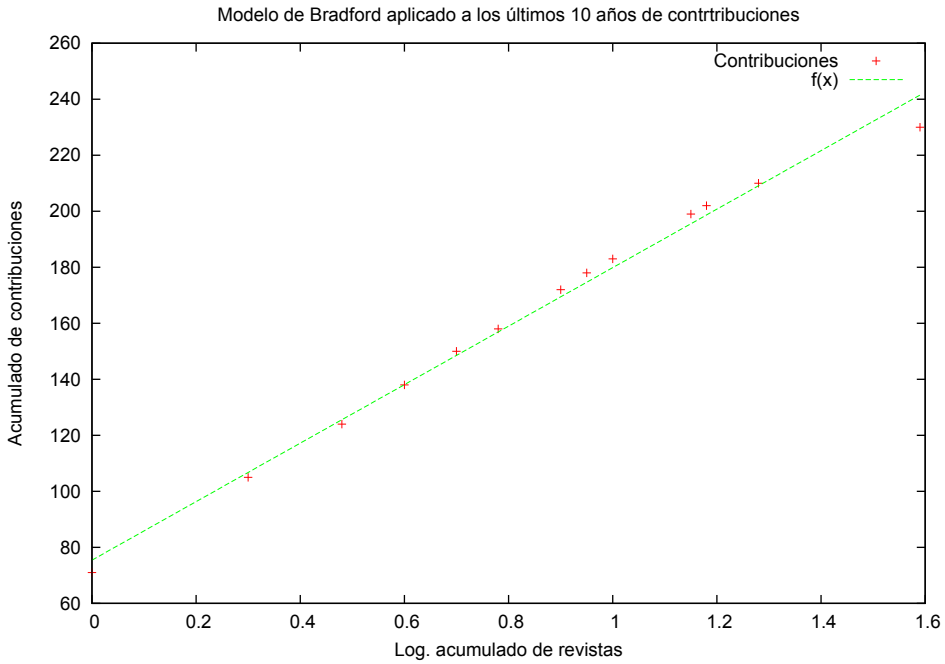

Modelo de Bradford aplicado a los últimos 10 años de contrtribuciones
Contribuciones
f(x)
Acumulado de contribuciones
Log. acumulado de revistas
260
240
220
200
180
160
140
120
100
80
60
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6

Figura 5.4:

Los datos demuestran que la distribución de los trabajos publicados coincide con el modelo de Bradford, no sucede del mismo modo con las referencias. Esto demuestra una particularidad de los usuarios de la biblioteca del IAR, los cuales parecen preferir como marco de referencia un reducido número de publicaciones, pero por otro lado su producción se distribuye según el modelo de Bradford.

Cuadro 5.7: Revistas donde se publicaron artículos producidos por el IAR

	Primeros 15 años		Ultimos 10 años
1	Anales de la Comisión de Investigaciones Científicas de la Prov de Buenos Aires	1	Advances in space research
1	Astrophysical journal Supplement series	1	Astroparticle physics
1	Carnegie Institution of Washington, Year book	1	Astrophysical Letters
1	Chemical Physics Letters	1	Bulletin Tokyo Gakugei University Sect
1	Nature physical science	1	Ciencia e investigación
1	Physics Letters	1	Classical and quantum gravity
1	Revista brasileira de fisica	1	Clumping in hot star winds
1	Revista de matemática y física teórica	1	General relativity and gravitation
1	Science	1	Invenio: revista de investigación académica
1	Theoretica chimica acta	1	Nature
2	Astronomical journal	1	Physical review letters
2	Boletín de la Academia Nacional de Ciencias, Córdoba	1	Physics Letters A
2	Revista astronómica	1	Physics letters B
2	Revista telegráfica electrónica	1	Physics reports
3	Astrophysical Letters	1	Publications of the Astronomical Society of Australia
7	Astrophysical journal	1	Publications of the Astronomical Society of Japan

Cuadro 5.7: Revistas donde se publicaron artículos producidos por el IAR (Continuación)

	Primeros 15 años		Ultimos 10 años
9	Astronomy and astrophysics Supplement series	1	Space science reviews
12	Astrophysics and space science	1	The atrophysical journal letters
17	**Astronomy and astrophysics**	1	The european physical journal B
		1	The old timer's bulletin
		2	Astronomische nachrichten
		2	Chinese journal of astronomy and astrophysics
		2	New astronomy reviews
		2	Physica A: statistical mechanics and its applications
		3	Science
		4	astrophysical journal letters
		4	Ciencia hoy
		4	International journal of modern physics D
		4	Nueva telegráfica electrónica
		5	Astronomy and astrophysics Supplement series
		6	Revista telegráfica electrónica
		7	Modern physics letters A
		7	Physical review D: particles, fields, gravitation, and cosmology
		8	Tube collector
		12	astronomical journal
		14	Astrophysics and space science
		19	**Monthly notices of the Royal Astronomical Society**
		34	**Astrophysical journal**
		71	**Astronomy and astrophysics**

Cuadro 5.8: Referencias contenidas en los trabajos publicados por el IAR

	Primeros 15 años		Ultimos 10 años
1	Bulletin of the Astronomical Institutes of Czechoslovakia	1	Annals of Physics
1	Australian Journal of Scientific Research A Physical Sciences	1	Arkiv for Astronomi
1	Bulletin of the Astronomical Institutes of Czechoslovakia	1	Astronomical and Astrophysical Transactions
1	Highlights of Astronomy	1	Baltic Astronomy
1	Journal of Chemical Physics	1	Brazilian Journal of Physics
1	Journal of the Royal Astronomical Society of Canada	1	Bull. Special Astrophys. Obs.
1	Memoirs Faculty of Sciences University of Kyoto	1	Carnegie Institute Washington D.C. Publication
1	Optics and Spectroscopy	1	Earth Moon and Planets
1	Physical Review Letters	1	European Physical Journal B
1	Proceedings of the Astronomical Society of Australia	1	International Journal of Modern Physics B
1	Reviews of Modern Physics	1	Irish Astronomical Journal
1	Royal Greenwich Observatory Bulletin	1	Journal des Observateurs
1	Royal Society of London Proceedings Series A	1	Journal of Computational and Applied Mathematics
1	Scientific American	1	Journal of High Energy Physics
1	Space Science Reviews	1	Journal of the Royal Astronomical Society of Canada
2	Annales d'Astrophysique	1	New Zealand Journal of Science
3	Proceedings of the National Academy of Science	1	Nuovo Cimento B Serie

Cuadro 5.8: Referencias contenidas en los trabajos publicados por el IAR (Continuación)

	Primeros 15 años		Ultimos 10 años
3	Zeitschrift fur Astrophysik	1	Nuovo Cimento C Geophysics Space Physics C
4	Astrophysical Letters	1	Physica B Condensed Matter
5	Astronomische Nachrichten	1	Physica Scripta Volume T
5	Publications of the Astronomical Society of Japan	1	Physical Review A
6	Nature	1	Progress in Particle and Nuclear Physics
6	The Observatory	1	Reviews of Geophysics and Space Physics
7	Memoirs of the Royal Astronomical Society	1	Sky and Telescope
7	Soviet Astronomy	1	Soviet Journal of Experimental and Theoretical Physics Letters
8	Publications of the Astronomical Society of the Pacific	1	Vistas in Astronomy
12	Astrophysics and Space Science	2	Annalen der Physik
13	Annual Review of Astronomy and Astrophysics	2	Annales d'Astrophysique
20	Astrophysical Journal Supplement Series	2	Astrophysical Letters Communications
27	Australian Journal of Physics	2	Astrophysics
30	Bulletin of the Astronomical Institutes of the Netherlands	2	Astrophysics and Space Physics Reviews
40	Astronomical Journal	2	European Physical Journal C
41	**Monthly Notices of the Royal Astronomical Society**	2	Geochimica et Cosmochimica Acta

Cuadro 5.8: Referencias contenidas en los trabajos publicados por el IAR (Continuación)

	Primeros 15 años		Ultimos 10 años
48	**Astronomy and Astrophysics Supplement Series**	2	Hamburger Sternw. Warner
105	**Astronomy and Astrophysics**	2	IEEE Proceedings
124	**Astrophysical Journal**	2	IEEE Transactions on Antennas and Propagation
		2	Journal of Geophysical Research
		2	Journal of Korean Astronomical Society
		2	Journal of Mathematical Physics
		2	Journal of Physics A Mathematical General
		2	Physics of Fluids
		2	Publications of the Royal Observatory of Edinburgh
		2	Radio Science
		3	Astrophysical Letters
		3	Bulletin d'Information du Centre de Donnees Stellaires
		3	Bulletin of the Astronomical Society of India
		3	Chinese Astronomy and Astrophysics
		3	Computer Physics Communications
		3	International Journal of Modern Physics A
		3	Planetary and Space Science

Cuadro 5.8: Referencias contenidas en los trabajos publicados por el IAR (Continuación)

	Primeros 15 años		Ultimos 10 años
		3	Publications of the Astronomical Society of Australia
		3	Reports on Progress in Physics
		4	Advances in Space Research
		4	Astronomy Reports
		4	Fundamentals of Cosmic Physics
		4	Journal of Astrophysics and Astronomy
		4	Nuclear Physics B
		4	Proceedings of the Astronomical Society of Australia
		4	Publications of the Warner
		4	Soviet Journal of Nuclear Physics
		4	Zeitschrift fur Astrophysik
		5	Acta Astronomica
		5	Chinese Journal of Astronomy and Astrophysics
		5	International Journal of Modern Physics D
		5	Progress of Theoretical Physics
		5	Quarterly Journal of the Royal Astronomical Society
		5	Soviet Astronomy Letters
		7	General Relativity and Gravitation
		7	Highlights of Astronomy
		7	Journal of Physics Conference Series
		8	International Astronomical Union Circular

Cuadro 5.8: Referencias contenidas en los trabajos publicados por el IAR (Continuación)

	Primeros 15 años		Ultimos 10 años
		8	Memoirs of the Royal Astronomical Society
		8	New Astronomy
		9	Astronomy and Astrophysics Review
		9	VizieR Online Data Catalog
		10	Bulletin of the Astronomical Institutes of the Netherlands
		10	Nuclear Physics B Proceedings Supplements
		11	Astronomy Letters
		11	Classical and Quantum Gravity
		13	Astronomische Nachrichten
		13	Soviet Astronomy
		16	Australian Journal of Physics
		16	Physical Review
		17	Australian Journal of Physics Astrophysical Supplement
		17	Journal of Physics G Nuclear Physics
		18	Reviews of Modern Physics
		20	Revista Mexicana de Astronomia y Astrofisica
		21	Physica A Statistical Mechanics and its Applications
		22	Modern Physics Letters A
		22	Physics Letters A
		24	Physics Letters B
		25	Physics Reports
		28	Icarus

Cuadro 5.8: Referencias contenidas en los trabajos publicados por el IAR (Continuación)

	Primeros 15 años		Ultimos 10 años
		29	GRB Coordinates Network
		32	Space Science Reviews
		33	The Astronomer's Telegram
		35	New Astronomy Review
		46	Nuclear Instruments and Methods in Physics Research A
		47	Publications of the Astronomical Society of Japan
		58	Astroparticle Physics
		68	Physical Review Letters
		74	Publications of the Astronomical Society of the Pacific
		80	Annual Review of Astronomy and Astrophysics
		80	Astrophysics and Space Science
		111	Science
		151	Physical Review D
		204	Nature
		310	Astrophysical Journal Supplement Series
		321	Astronomy and Astrophysics Supplement Series
		458	Astronomical Journal
		617	**Monthly Notices of the Royal Astronomical Society**
		1886	**Astronomy and Astrophysics**
		2114	**Astrophysical Journal**

Se presentan algunos interrogantes susceptibles de ser desarrollados en futuros trabajos: ¿se repite este rasgo en otras instituciones del país?, ¿pasa

lo mismo en otras instituciones en el exterior?

Como punto de partida puede tomarse el estudio realizado por M. Gómez (2003) en el Instituto de Astrofísica de Canarias, donde se determina que los trabajos publicados y sus referencias se concentran en más de un 60 % en las revistas nucleares como *Astronomy and Astrophysics, Astrophysical Journal y MNRAS*, tal como sucede con los datos recogidos para el IAR en este trabajo. Obsérvense las Tablas 5.7 y 5.8.

5.2. El envejecimiento de la literatura referenciada por el IAR

En el estudio de obsolescencia se trabajó sobre la misma base de datos que se usó para determinar el grupo nuclear de publicaciones tomando como dato fundamental la fecha de las referencias.

En la tabla 5.9 se presentan las edades de las referencias, sus frecuencias absolutas, las frecuencias acumuladas y la utilidad. En la columna 5 puede verse la utilidad teórica y compararla con la utilidad que arrojan los datos observados en la columna 4.

Cuadro 5.9: Edades, referencias y utilidades

Edad	Ref/año	Ref acum	Utilidad	Ut. teórica
0	63	12458	1.0000	1.0000
1	188	12395	0.9949	0.9600
2	352	12207	0.9799	0.9216
3	336	11855	0.9516	0.8847
4	321	11519	0.9246	0.8493
5	358	11198	0.8989	0.8154
6	353	10840	0.8701	0.7828
7	545	10487	0.8418	0.7514
8	429	9942	0.7980	0.7214
9	495	9513	0.7636	0.6925
10	401	9018	0.7239	0.6648
11	394	8617	0.6917	0.6382
12	450	8223	0.6601	0.6127
13	397	7773	0.6239	0.5882
14	268	7376	0.5921	0.5647
15	276	7108	0.5706	0.5421
16	361	6832	0.5484	0.5204
17	245	6471	**0.5194**	**0.4996**
18	272	6226	**0.4998**	**0.4796**
19	259	5954	0.4779	0.4604
20	300	5695	0.4571	0.4420
21	266	5395	0.4331	0.4243
22	304	5129	0.4117	0.4073
23	255	4825	0.3873	0.3911
24	241	4570	0.3668	0.3754
25	245	4329	0.3475	0.3604
26	344	4084	0.3278	0.3460
27	339	3740	0.3002	0.3321
28	317	3401	0.2730	0.3189
29	289	3084	0.2476	0.3061
30	227	2795	0.2244	0.2939
31	236	2568	0.2061	0.2821
32	289	2332	0.1872	0.2708
33	207	2043	0.1640	0.2600
34	268	1836	0.1474	0.2496

Cuadro 5.9: Edades, referencias y utilidades (Continuación)

Edad	Ref/año	Ref acum	Utilidad	Ut. teórica
35	246	1568	0.1259	0.2396
36	184	1322	0.1061	0.2300
37	92	1138	0.0913	0.2208
38	192	1046	0.0840	0.2120
39	99	854	0.0686	0.2035
40	106	755	0.0606	0.1954
41	94	649	0.0521	0.1876
42	70	555	0.0445	0.1800
43	51	485	0.0389	0.1728
44	66	434	0.0348	0.1659
45	59	368	0.0295	0.1593
46	25	309	0.0248	0.1529
47	16	284	0.0228	0.1468
48	55	268	0.0215	0.1409
49	32	213	0.0171	0.1353
50	17	181	0.0145	0.1299
51	19	164	0.0132	0.1247
52	24	145	0.0116	0.1197
53	21	121	0.0097	0.1149
54	13	100	0.0080	0.1103
55	8	87	0.0070	0.1059
56	9	79	0.0063	0.1017
57	10	70	0.0056	0.0976
58	11	60	0.0048	0.0937
59	6	49	0.0039	0.0900
60	8	43	0.0035	0.0864
61	2	35	0.0028	0.0829
62	2	33	0.0026	0.0796
63	2	31	0.0025	0.0764
64	2	29	0.0023	0.0733
65	3	27	0.0022	0.0704
66	2	24	0.0019	0.0676
67	4	22	0.0018	0.0649
68	1	18	0.0014	0.0623

Cuadro 5.9: Edades, referencias y utilidades (Continuación)

Edad	Ref/año	Ref acum	Utilidad	Ut. teórica
69	1	17	0.0014	0.0598
70	3	16	0.0013	0.0574
71	4	13	0.0010	0.0551
72	2	9	0.0007	0.0529
73	2	7	0.0006	0.0508
74	1	5	0.0004	0.0488
75	1	4	0.0003	0.0468
76	1	3	0.0002	0.0449
77	1	2	0.0002	0.0431
78	1	1	0.0001	0.0414

Al utilizar el método del cálculo del máximo de referencias anuales, a partir de las referencias absolutas, se obtuvo un factor de envejecimiento anual de 0,96.

$a = 1 - C/R$

$a = 1 - 545/12458$

$a = 0,96$

Con el método de la vida media se obtuvo el mismo factor de envejecimiento de 0,96.

Para conocer con exactitiud el valor de h e ilustrar los cálculos, interpolamos en el intervalo de 17 a 18 años, según los datos de la Tabla 5.9:

$18 - (18 - 17) * (0,5 - 0,5)/(0,5 - 0,52)$

$h = 17,99$

Aplicando la ecuación (2):

$a^h = 0{,}5$

Al despejar a, resulta:

$a = exp(1n(0,5)/h)$

Sustituyendo h por su valor, se obtiene:

$a = exp(1n(0,5)/17{,}99) = 0,96$

Concluyendo, la pérdida de utilidad por año para la literatura consultada por el IAR sería del 4 %. De modo que cada año que pasa la utilidad queda reducida a un 96 % respecto de la del año anterior (Figura 5.5).

Por su parte, la vida media para la literatura consultada por el IAR es de 18 años.

Para calcular el índice de Price se tomaron las referencias comprendidas entre el año 0 y el año 5.

$IP = 1260/12458 * 100$

$k = 0,5$

$IP = 10,1$

De acuerdo a este bajo valor obtenido para el índice de Price podría suponerse que hay un alto componente de literatura de interés histórico consultada por el IAR en cuatro décadas. Las implicancias de tal afirmación serán analizadas más adelante en este trabajo.

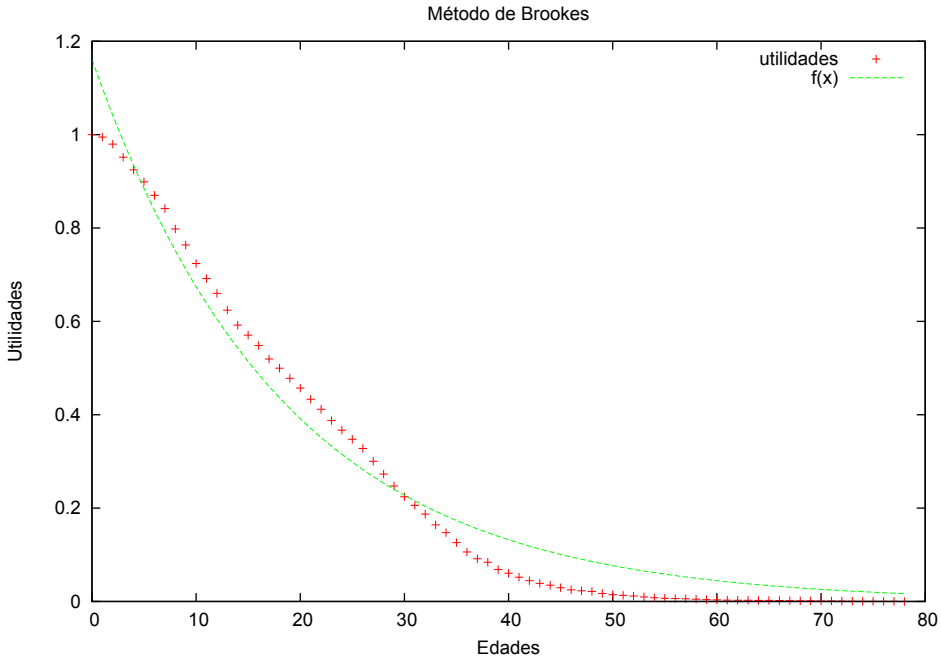


Figura 5.5: Envejecimiento de la literatura referenciada por el IAR

5.3. Determinación de los frentes de investigación del IAR

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

Como se dijo anteriormente el estudio se realizó sobre las palabras clave y autores presentes en la base de datos de contribuciones del IAR entre el año 1999 y 2008. Pudo determinarse que las palabras clave con mayor peso y mayor poder de intermediación y cercanía son: *Interstellar medium, Stars, Radio continuum y Gamma ray-astronomy* (Tabla 5.10).

Cuadro 5.10: Palabras clave

Keywords	Grado Nodal	Intermediación	Cercanía
Interstellar medium	28.000000	0.050385	0.826389
Stars	27.000000	0.049414	0.806233
Radio continuum	27.000000	0.044079	0.806233
Gamma-ray astronomy	26.000000	0.040499	0.787037
Early-type stars	21.000000	0.027813	0.703310
Jets	22.000000	0.022673	0.674603
Galaxies	21.000000	0.021368	0.648148
Stellar winds	23.000000	0.021287	0.718599
Outflows	23.000000	0.019807	0.718599
Gamma-ray observations	22.000000	0.019536	0.688657
Black holes	19.000000	0.019126	0.648148
Interstellar bubbles	19.000000	0.018144	0.674603
X-ray emission	21.000000	0.017925	0.703310
X-ray binaries	22.000000	0.014777	0.718599
Radio emission lines	13.000000	0.013093	0.601010
Binary stars	20.000000	0.011098	0.688657
Observations	20.000000	0.010106	0.688657
Gamma-ray theory	19.000000	0.009111	0.661111
Radiation mechanisms	20.000000	0.009058	0.648148
Milky Way Galaxy	12.000000	0.008261	0.590278
Wolf-Rayet stars	16.000000	0.007651	0.635684
Non-thermal mechanisms	18.000000	0.006930	0.648148
Supernova remnants	15.000000	0.006596	0.612140
Dynamics	15.000000	0.005932	0.601010
Kinematics	15.000000	0.005932	0.601010
HII regions	13.000000	0.005406	0.601010
Surveys	13.000000	0.004413	0.601010
BL Lacertae objects	16.000000	0.003269	0.579922
Quasars	9.000000	0.002901	0.560264
Microquasars	14.000000	0.002665	0.601010
Cosmic rays	9.000000	0.001793	0.560264
Active galaxies	12.000000	0.001538	0.590278
Gamma-ray sources	9.000000	0.001168	0.550926
HI line emission	12.000000	0.001106	0.550926
Mass-loss	11.000000	0.001021	0.579922

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.

Podemos 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**. Para astrofísica de altas energías se asocian principalmente las palabras clave: *X-ray binaries, binary stars, microquasars, gamma-ray theory, X-ray emission y observations*; para medio interestellar: *interstellar bubbles, Wolf-Rayet stars, HII regions y stellar winds*. Hay muchos términos compartidos como *supernova remnants, outflows y radio continuum* que podrían dar cuenta del contacto existente entre ambas líneas. Con respecto a términos como *stars*, que comparten fuertemente la relación con ambos grupos de palabras clave podría atribuirse a la asignación de un término general que obviamente se aplica a ambas líneas (ver Figura 5.6).

La indización es un aspecto a considerar pues, a pesar de haberse normalizado la lista de palabras clave, pueden verse en la red algunas palabras que bien podrían formar parte de una misma categoría como *gamma-ray theory, gamma-ray observations y gamma-ray astronomy*, de hecho las dos primeras son términos específicos de la tercera. Esto supone una nueva perspectiva a la hora de realizar el análisis de palabras clave pues el nivel de especificidad de cada palabra puede implicar agrupaciones diferentes y por tanto modificaciones en los resultados del estudio.

Podría atribuirse la presencia de nodos en la periferia de la red a algunos términos que tienen que ver con temáticas de trabajos que el IAR ha publicado en colaboración con otras instituciones.

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 líderes en esos grupos, como el caso de Romero y Arnal. En la Tabla 5.11 se presentan los indicadores de centralidad correspondientes. Puede observarse que Arnal posee un grado nodal menor que otros autores como Benaglia, Torres o Combi, no obstante se mantiene como líder en su propio grupo (figuras 5.7 y 5.8).

En la figura 5.7 pueden observarse todos los autores de la base de datos mientras que en la figura 5.8 se discriminan los autores de acuerdo a su

pertenencia o no al IAR.[1]

Cuadro 5.11: Autores

Autores	Grado Nodal	Intermediación	Cercanía
Romero, G. E.	19.00000	0.35301	0.80645
Paredes, J. M.	10.00000	0.01893	0.55556
Torres, Diego F.	13.00000	0.03674	0.59524
Combi, J. A.	12.00000	0.01527	0.58140
Arnal, E. M.	11.00000	0.10949	0.64103
Benaglia, P.	15.00000	0.17390	0.71429
Cappa, C. E.	9.00000	0.05793	0.60976
Bosch Ramon, Valenti.	8.00000	0.00333	0.53191
Martí, J.	10.00000	0.00638	0.55556
Ribó, M.	10.00000	0.00638	0.55556
Orellana, M.	5.00000	0.01128	0.51020
Cellone, S.	5.00000	0.00000	0.49020
Albert, J.	2.00000	0.00000	0.46296
Morras, R.	8.00000	0.04421	0.59524
Bajaja, E.	5.00000	0.00000	0.41667
Lemarchand, Guillermo A.	2.00000	0.00000	0.46296
Pöppel, W. G. L.	5.00000	0.00000	0.41667
Butt, Yousaf	10.00000	0.00592	0.55556
Niemela, V. S.	6.00000	0.02267	0.53191
Mirabel, I. F.	8.00000	0.00393	0.53191
Testori, J. C.	10.00000	0.09883	0.62500
Cichowolski, S.	6.00000	0.00000	0.42373
Kaufman Bernadó, M. Marina	7.00000	0.00114	0.52083
Pineault, S.	6.00000	0.00000	0.42373
Dame, Thomas	6.00000	0.00000	0.51020
Goss, W. M.	6.00000	0.00400	0.43860

Podemos decir en términos generales que hemos encontrado ciertas correspondencias en los agrupamientos de palabras clave y de autores, lo cual

[1]Cabe aclarar que hay autores que se presentan en la red como pertenecientes al IAR pues poseen trabajos dentro de la institución durante los años 2000 a 2001 pero que hoy no pertenecen al IAR, como es el caso de Torres. Si se tiene en cuenta este hecho puede afirmarse que existe una fuerte colaboración entre Romero y autores externos.

nos ayuda a corroborar la validez de la metodología empleada.

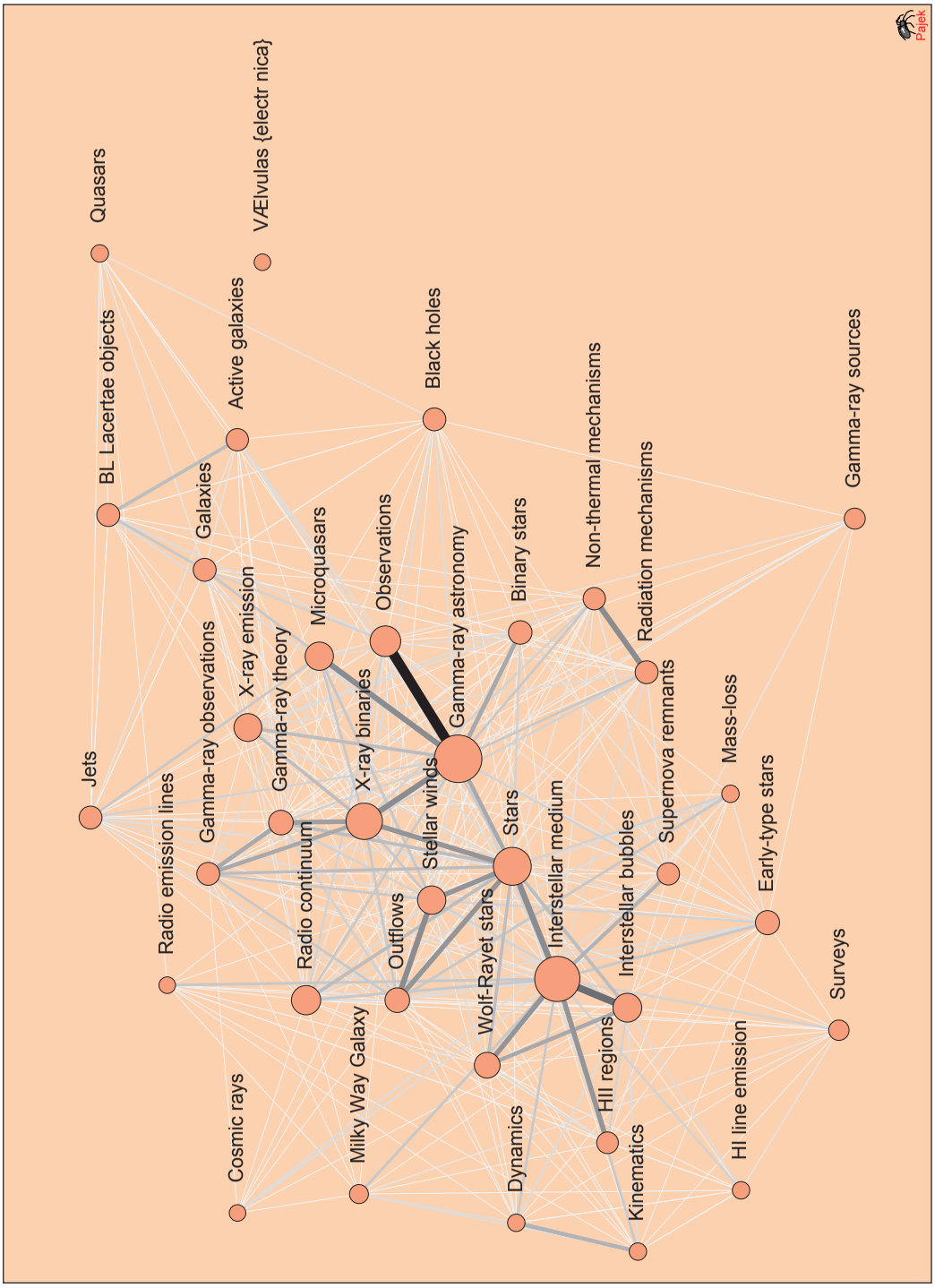


Figura 5.6: Palabras clave

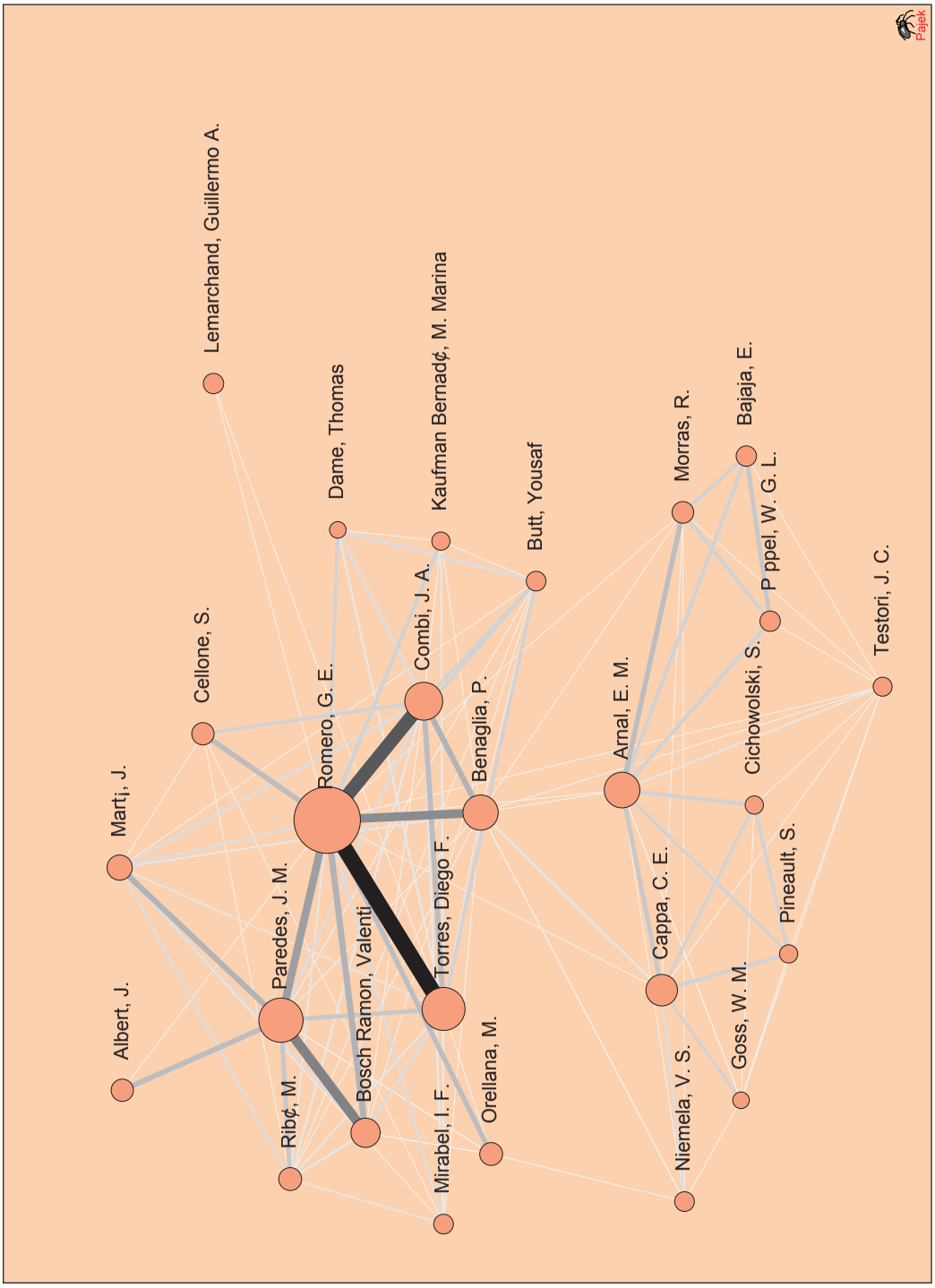


Figura 5.7: Total de autores de la base de datos

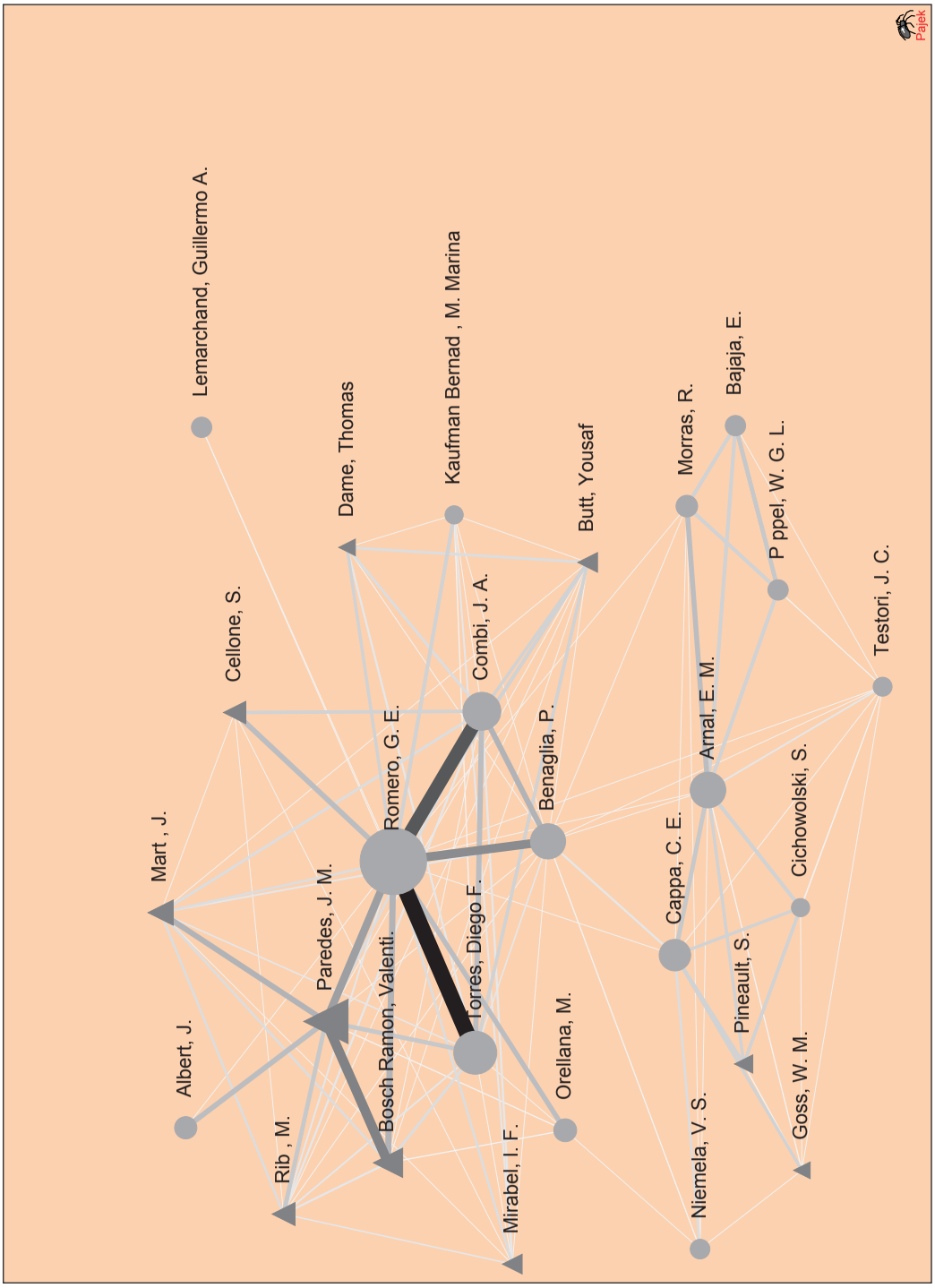


Figura 5.8: Autores del IAR y de otras instituciones

Capítulo 6

Discusión y conclusiones

A partir del análisis de los trabajos publicados por los investigadores del IAR y de las referencias contenidas en los mismos, se pudieron determinar ciertos patrones de consumo y producción de información de los científicos de la institución que pueden ser el sustento para futuras decisiones en el desarrollo de la colección de la Biblioteca. El análisis e interpretación de los resultados obtenidos confirman ciertas características en los hábitos de consumo y producción de información de sus usuarios, que se venían observando informalmente en los últimos diez años.

La aplicación del modelo de Bradford como forma de determinar el grupo nuclear de publicaciones permitió revelar que existe una marcada tendencia de los investigadores a consumir y a publicar artículos en dos o tres títulos de la disciplina. La alta concentración de referencias en *Astrophysical journal , Astronomy & astrophysics y Monthly notices of the R. A. S.* permite afirmar sin dudas que el núcleo de publicaciones usadas por el IAR está representado por esos tres títulos.

Estos resultados introdujeron algunos interrogantes relacionados con hábitos de comportamiento del grupo de investigadores estudiado. Estudios previos como los de Abt (1995), Gómez (2003) e Isaksson (2007) dan cuenta de resultados similares lo cual podría conducir a pensar en una regularidad de comportamiento de los investigadores dentro de la astronomía.

Cuando se aplicó el modelo al conjunto de referencias incluidas en los trabajos publicados se observó que la dispersión de la distribución de estas referencias era diferente a la dispersión comprobada por Bradford, cosa que

no ocurría al aplicar el mismo modelo al conjunto de trabajos publicados. Este comportamiento particular demuestra que los investigadores del IAR prefieren citar un reducido número de publicaciones aunque publiquen sus artículos en un rango más amplio que sí responde al modelo de Bradford aplicado.

Es importante mencionar que en este estudio el modelo de Bradford no ha sido reproducido tal como fue llevado a cabo por su autor. Bradford analiza los artículos de las revistas de una especialidad durante tres años, o sea la distribución de la literatura sobre un tema específico en las distintas fuentes; mientras que en este estudio se analiza el consumo y la producción de diez años en todas las fuentes en las que los artículos fueron publicados independientemente de la especificidad del área temática. Esta variación en la aplicación del modelo podría tener ciertos efectos en los resultados originadas por diferencias en el periodo de tiempo estudiado o bien en las fuentes de datos utilizadas. No obstante, la aplicación del método puede considerarse válida en tanto sirve para analizar y comparar patrones de comportamiento relativos al consumo y la producción de los investigadores de la institución que sin dudas pueden utilizarse como elementos para evaluar la colección.

Puede decirse, analizando las referencias de los primeros quince años y los últimos diez, que no han surgido nuevas publicaciones relacionadas con temáticas específicas. Sin embargo se nota la aparición de un grupo de referencias a revistas de física, que sin llegar a compararse con las revistas líderes, comienzan en los últimos años a ocupar un lugar que en los primeros quince no ocupaban, y que podrían dar cuenta del desarrollo de una nueva línea de investigación que se condice con el desarrollo de nuevos instrumentos como los telescopios que trabajan en la frecuencia de rayos X y rayos gamma y que poseen un fuerte componente de estudios teóricos en al área de la física.

Estos resultados dan lugar a pensar en futuros estudios aplicables a otros institutos de la Argentina o del exterior que permitirían confirmar o descartar este comportamiento para la disciplina en forma general, o también un estudio basado en Bradford para las publicaciones dentro del área.

Con respecto al envejecimiento de la literatura referenciada por los investigadores del IAR se pudo comprobar que existe un alto componente de literatura de interés histórico reflejado por el bajo índice de Price obtenido y por el nivel elevado de la vida media el cual se ubicó en los 18 años. Estos resultados se condicen con estudios previos realizados por Abt (1996) donde se

observa una vida media también elevada. Los estudios de utilidad realizados demuestran que la colección de la biblioteca del IAR pierde una utilidad del 4,4 % por cada año que transcurre. Estos datos podrían servir para justificar por ejemplo, inversiones en restauración y preservación de la colección, y ser una herramienta para desarrollar políticas de expurgo.

El análisis de co-ocurrencia de palabras y de coautoría asociado al análisis de redes sociales permitió comprobar una correspondencia entre las dos temáticas predominantes en el IAR y los dos grupos de investigación existentes, mediante la detección de las palabras clave más representativas de su produccción.

Se ha observado la predominancia de términos relacionados con astrofísica de altas energías y medio interestelar que se corresponden con las dos líneas de investigación y se encuentran a su vez conectadas por palabras clave comunes como *Stars o Supernova remnants*. Es importante mencionar que la historia del IAR se inicia en la década del sesenta, por lo cual los términos de indización se corresponden con temáticas recientes en el desarrollo de la disciplina astronómica, por lo cual las temáticas incluidas en la base de datos se corresponderían con los resultados obtenidos por Abt (1996) en la segunda parte de su estudio.

Un elemento que ha influido en la detección de estos términos fue la indización de los documentos. Como ejemplo puede mencionarse la asignación de términos genéricos como *Gamma-ray astronomy* que engloban a otros más específicos como *Gamma-ray theory* o *Gamma-ray observations* y que, si fueran absorbidos por el término genérico modificarían la visualización de la red.

En futuros estudios se podría trabajar con las palabras clave presentes en los títulos, que probablemente representan en forma más específica el contenido de los documentos aunque no dejen de presentar problemas de sinonimia susceptibles de normalizar.

De todo lo expuesto puede concluirse que la aplicación de métodos bibliométricos indirectos en la evaluación de la colección de una biblioteca puede ser una herramienta eficaz no solo dentro del área de la astronomía sino en otras áreas en diferentes bibliotecas especializadas.

Bibliografía

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

[2] 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.

[3] Beaver, D. d., & Rosen, R. (1978). Studies in scientific collaboration. Part I: the profesional origins of scientific coautorship. *Scientometrics*, 1(1978), 65–84.

[4] Bordons, M. (2001). Aspectos metodológicos en la obtención de indicadores bibliométricos. *Cuadernos de Indicios*, (Jun), 17–26.

[5] Börner, K., Chen, C., & Boyack, K. W. (2005). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, 37(1), 179–255.

[6] Bradford, S. C. (1934). Sources of information on specific subjects. Originally Published in *Engineering. A Illustrated Weekly Journal*, for 26 January 1934 (Vol. 137 Nº 3550) (London) 137, 85–86.

[7] Brookes, B. C. (1969). Bradford's law and the bibliography os science. *Nature*, 224(Dec. 6), 953–956.

[8] Brookes, B. C. (1970). Obsolescence of special library periodicals : sampling errors and utility contours. *Journal of American Society of Information Science*, (Sep.-Oct.), 320–329.

[9] Burton, R. E., & Kebler, R. W. (1960). The half-life of some scientific and technical papers and technical literature. *American Documentation*, 11(1).

[10] Busha, C. H. , & Harter, S. P. (1990). *Métodos de investigación en Bibliotecología: técnicas e interpretación*. México: : UNAM.

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

[12] Garfield, E. (1970). Citation indexing for studying science. *Nature*, 227(Aug), 669–771.

[13] Glänzel, W. 2. A concise introduction to Bibliometrics & its history. [Página web]. URL http://www.ecoom.be/index.php?id=31. [Recuperado el 15 de julio de 2010]

[14] Gómez, M. (2003). A bibliometric study to manage a journal collection in an Astronomical Library: some Results. *Library and Information Services in Astronomy IV*, 214–222.

[15] Hanneman, R. A., & Riddle, M. (2005) Introduction to social network methods [Página web]. URL http://www.faculty.ucr.edu/ hanneman/nettext/ [Recuperado el 15 de julio de 2010].

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

[17] Henneken, E. A., Kurtz, M. J., Accomazzi, A., Grant Carolyn S. , Thompson, D., Bohlen, E., & Murray, S. S. (2009). Use of astronomical literature : a report of usage patterns. *Journal of Informatrics*, 3, 1–8.

[18] Izquierdo, L. R., & Hanneman, R. A. (2006). *Introduction to the formal analysis of social networks using mathematica*. Burgos.

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

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

[21] Kessler, M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10–25.

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

[23] Miguel, S., Caprile, L., & Jorquera-Vidal, I. (2008). Análisis de cotérminos y de redes sociales para la generación de mapas temáticos. *El profesional de la información*, 17(6), 637–646.

[24] Pajek : Program for Large Network Analysis [Página web]. URL http://pajek.imfm.si/doku.php [Recuperado el 3 de noviembre de 2009].

[25] Perianes-Rodríguez, A., Olmeda-Gómez, C., & Moya y Anegón, F. d. (2008). Introducción al análisis de redes. *El profesional de la información*, 17(6), p. 664–669.

[26] Perianes Rodríguez, A., Olmeda Gómez, C., & Moya Anegón, F. (2009). Detecting, identifying and visualizing research groups in co-authorship networks. *Scientometrics*, 82, (2), p. 307–319.

[27] Person, O. (2001). All author versus first author citations. *Scientometrics*, 50(2), 339–344.

[28] Persson, O. Bibexcel [Página web]. URL http://www8.umu.se/inforsk/Bibexcel/ [2009, November 3].

[29] 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.

[30] Price, D. J. d. S. (1965). Networks of scientific papers. *Science*, 149, 510–515.

[31] Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation*, 24, 348–349.

[32] Ruiz Baños, R., & Bailon-Moreno, R. (1997). Métodos para medir el envejecimiento de la literatura científica. *Boletín de aa Asociación Andaluza de Bibliotecarios*, (46), 57–75.

[33] Ruiz de Osma, E. (2006). Aplicación del modelo Bradford en la producción científica del área biomédica de la Universidad de Granada (1988-1996). *Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.*, (2do. sem), 23 p.

[34] Sanz Casado, E. (1994). *Manual de estudios de usuarios*. Madrid: Fundación Germán Sánchez Ruipérez.

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

[36] Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2009). Comparative study on methods of detecting research fronts using different types of citation. *Journal of the American Society for Information Science and Technology*, 60(3), 571–580.

[37] Small, H. (1973). Co-citation in the scientific literature : a new measure of the relationship betweeen two documents. *Journal of the American Society for Information Science*, 24(4), 265–269.

[38] Solla Price, D. J. d. (1970). *Citation Measures of hard science, soft science, technology, and nonscience. Communication among scientists and engineers*, Lexington, MA: D.C. Heath and Company.

[39] Sonnenwald, D. (2007). Scientific Collaborations. *Annual Review of Information Science and Technology*, 41, p. 643–681.

[40] Spinak, E. (1998). Indicadores bibliométricos. *Ciencia da Informação*. 27(maio/ago), 141–148.

[41] Trimble, V. (2009). Telescopes in the mirror of scientometrics. *Experimental Astronomy*, 26, 133–147.

[42] Valderrama-Zurián, J. C. et al. (2006). Coautorship networks and institutional collaboration in Revista española de cardiología publications. *Revista Española de Cardiología*, 60(2), 117–130.

Capítulo 7

Anexo: Herramientas y técnicas para la recolección y preparación de los datos

En este anexo se describirán los procedimientos y las herramientas utilizadas para obtener la información de base para el estudio, y para la redacción del texto. Se ha seleccionado un conjunto de datos a extraer de las bases del IAR con el objetivo de encontrar elementos de información que reflejen de algún modo la realidad objeto de este estudio.

7.1. Obtención de referencias

Las referencias de los trabajos publicados fueron descargadas desde el ADS[1] en formato etiquetado por *tags* textuales (formato Procite). Cada archivo contenía las referencias que correspondían a cada uno de los trabajos publicados, también denominados contribuciones. Como primer paso hubo que relacionar cada una de las referencias obtenidas con el número de trabajo en la base de datos. Para ello se aplicó un *script* denominado **ref.sh**[2] que tomó como fuente los archivos descargados. El resultado se almacenó en un archivo de texto.

[1] NASA Astrophysical Data System http://adswww.harvard.edu/

[2] Al final del anexo se incluye el texto del script.

```
./ref.sh *.txt  > con_ref_todos.txt
```

7.2. Aplicación de utilitarios Cisis

El archivo de texto obtenido se transformó en una base de datos Isis mediante la aplicación de los utilitarios Cisis[3] y un *pft* que convierte las etiquetas textuales del ADS en etiquetas numéricas. De los utilitarios Cisis se aplicó el **mx**, una aplicación de propósito general que permite operar con las bases Isis desde la línea de comandos, y el **id2i** un utilitario con el que es posible transformar archivos de texto en bases de datos Isis.

```
mx seq=archivo_texto create=base1 now -all
mx base1 pft=@pct2id-tesis.pft now > base2.id
id2i base2.id create=base
```

7.2.1. Obsolescencia

Para trabajar con datos de obsolescencia fue necesario aislar las fechas de cada referencia:

```
mx base pft="v7.4/" > fechas -all now
```

El listado se procesó con una planilla de cálculos aplicando la función *contar si* para conocer la cantidad de referencias por año y las edades de las referencias.

[3]http://bvsmodelo.bvsalud.org/php/level.php?lang=es&component=28&item=1

7.2.2. Núcleo de revistas

Para aplicar el modelo de Bradford fue necesario aislar los títulos de las publicaciones y los tipos de publicación:

```
mx base pft="v5,'|',v2/" > titulos -all now

v2 (tipo: journal, conference, etc)
v5 (título)
```

El paso siguiente fue recuperar las contribuciones correspondientes a esos años. Se aislaron los quince primeros años y los últimos diez de trabajos del IAR y se obtuvieron sus referencias:

```
mx base pft="v1/" > v1 -all now

mx base pft="v1,'|',v5,'|',v2/" > titulos_nc -all now
```

Se procesaron con una hoja de cálculos separando los trabajos de cada periodo.

Otra de las fuentes de datos utilizada fue la base de datos de trabajos publicados del IAR (**contri**), la cual es Isis, está en formato bibliográfico MARC21 y reune todos los trabajos publicados por los investigadores del IAR desde su fundación.

Los datos a extraer de esta base fueron los números de contribución, los títulos de las revistas donde se publicaron los trabajos, las fechas de publicación y las palabras clave. Los campos a procesar fueron el 953, 773, campo 008 y 653 respectivamente.

```
mx biblio lw=8000 "pft=( v953[1]^a, '|', v8[1]*7.4, '|', \
v653^a, /, )" > keywords -all now
```

Una vez obtenida la lista, se ordenó alfabéticamente por palabra clave y se procedió a normalizar los términos.

Otro dato a extraer de esta base fue la lista de autores cada uno de los cuales hubo que relacionar con los números de trabajo. Para ello se listaron los campos 100 y 700, para después unirlos en la hoja de cálculo. Se agregó el año de publicación para poder seleccionar los periodos estudiados.

```
mx biblio lw=8000 "pft=if p(v100) then ( v953[1]^a, '|', \
v100^a, '|',v8[1]*7.4, /, ),fi" \ > autores -all now

mx biblio lw=8000 "pft=if p(v700) then ( v953[1]^a, '|', \
v700^a, '|', v8[1]*7.4, /, ),fi" \ > autores1 -all now
```

7.3. Selección de datos relacionales

La aplicación de las técnicas de análisis de co-ocurrencia de palabras y de autores requiere el ordenamiento de los datos mediante tablas relacionales[4]. Una vez obtenida la lista de palabras clave mediante los procedimientos explicados más arriba se procedió a procesarlas mediante una hoja de cálculos e importarlas en *Access*.

Se crearon diferentes consultas para contar la cantidad de palabras clave y de autores, y las co-ocurrencias en cada grupo. El objetivo de estos procesos fue la creación de una matriz de datos, fuente desde donde obtiene los datos el Pajek[5], software que diagrama las redes y que será presentado en el siguiente punto de este anexo.

[4]Por cuestiones relacionadas con los tiempos de desarrollo del trabajo no se ha podido incursionar en el lenguaje SQL que permite interactuar con bases relacionales, por tal motivo se optó por trabajar con un software propietario como *Access*. Información de carácter introductorio en http://en.wikipedia.org/wiki/SQL (véase especialmente la lista de referencias)

[5]http://pajek.imfm.si/doku.php

7.4. Análisis de redes: *Pajek*

Para generar y visualizar las redes de co-ocurrencia de palabras clave y de autores se utilizó el software *Pajek*.

Para que la aplicación pueda construir las redes es necesario trabajar con matrices de datos. Estas matrices se deben transformar en archivos de texto de entrada con extensión `.net`, `.vec`, `.clu`, etc.

Para graficar la red de autores por ejemplo, se abrió con el *Pajek* el archivo `autores.net`, donde se encuentra la lista de autores. Luego se abrió también el archivo `autores.vec`, en el que se configuraron valores representativos de los diferentes tamaños de los nodos en función de las frecuencias de ocurrencia.

La red de autores se visualizó a partir de la herramienta *draw y draw-vector* del menú principal del *Pajek*. Luego se aplicó el algoritmo *Kamada Kawai*. Un trabajo similar se llevó a cabo para analizar la red de palabras clave.

Se utilizaron diferentes opciones de configuración para exportar las imágenes a partir de las opciones de exportación de la herramienta *draw*. Finalmente, para calcular las medidas de centralidad se utilizaron las opciones de *transformar la red*.

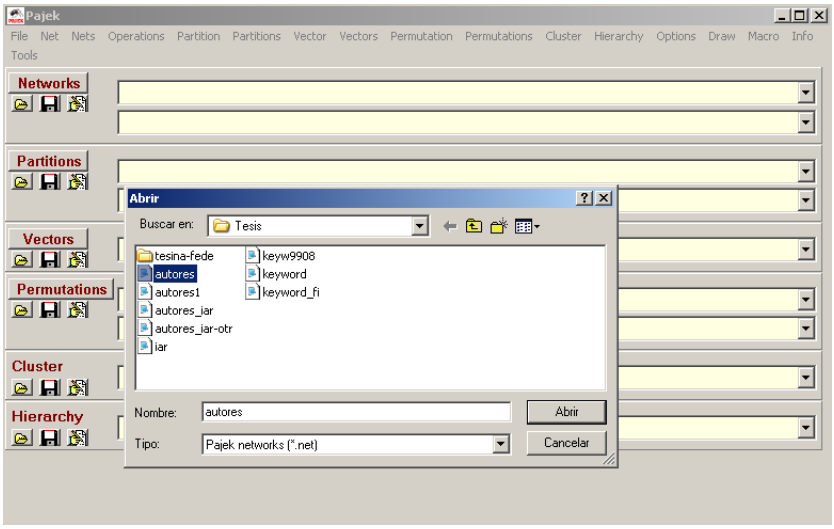
7.5. Formateo del texto

El texto de esta tesina fue generado con LaTeX[6], un poderoso lenguaje de composición tipográfica orientado especialmente a la creación de libros, documentos científicos y técnicos que contienen fórmulas matemáticas. El motivo fundamental de la elección tiene que ver con la necesidad —o tal vez la exigencia— de calidad del producto final obtenido. Otro de los motivos que determinaron la elección es que LaTeX es software libre[7], y por el hecho de trabajar con herramientas que separan el contenido de la apariencia que posee el texto lo cual da una libertad que los procesadores de texto no ofrecen.

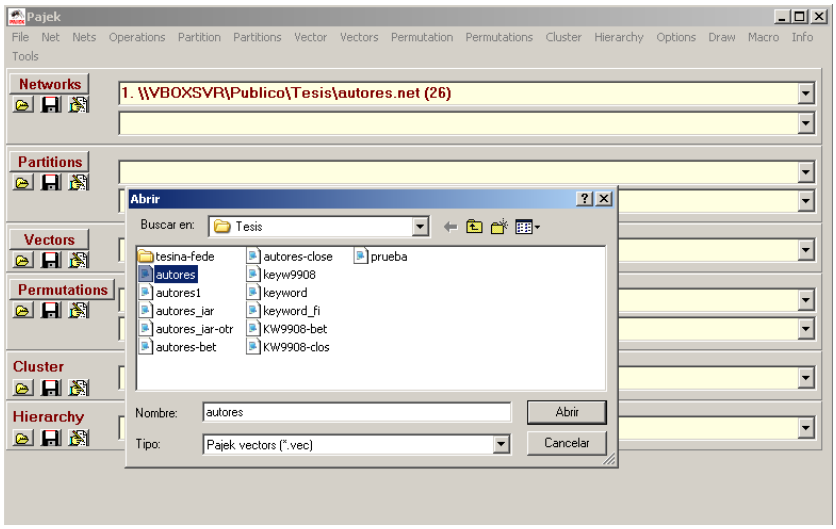
Las imágenes fueron generadas como eps *Encapsulated PostScript Image*, uno de los formatos de gráficos que gestiona LaTeX.

[6] http://www.latex-project.org/

[7] http://www.gnu.org/philosophy/free-sw.html



(a) autores.net



(b) autores.vec

Figura 7.1: autores

Los datos que se encontraban en las hojas de cálculo y que permitieron generar las tablas fueron extraidos mediante sentencias disponibles en el sistema operativo Linux. Las tablas en LaTeX deben poseer una estructura que obliga incluir delimitadores de columna y de fila. La utilización del comando

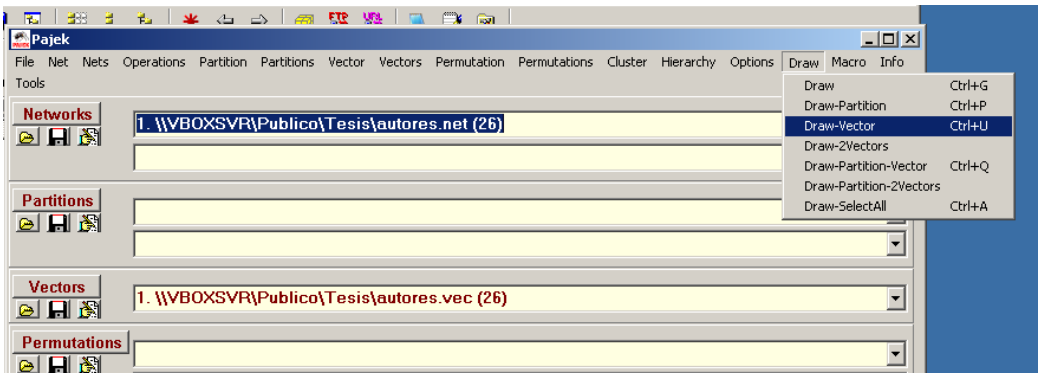


Figura 7.2: Draw y Draw Vector

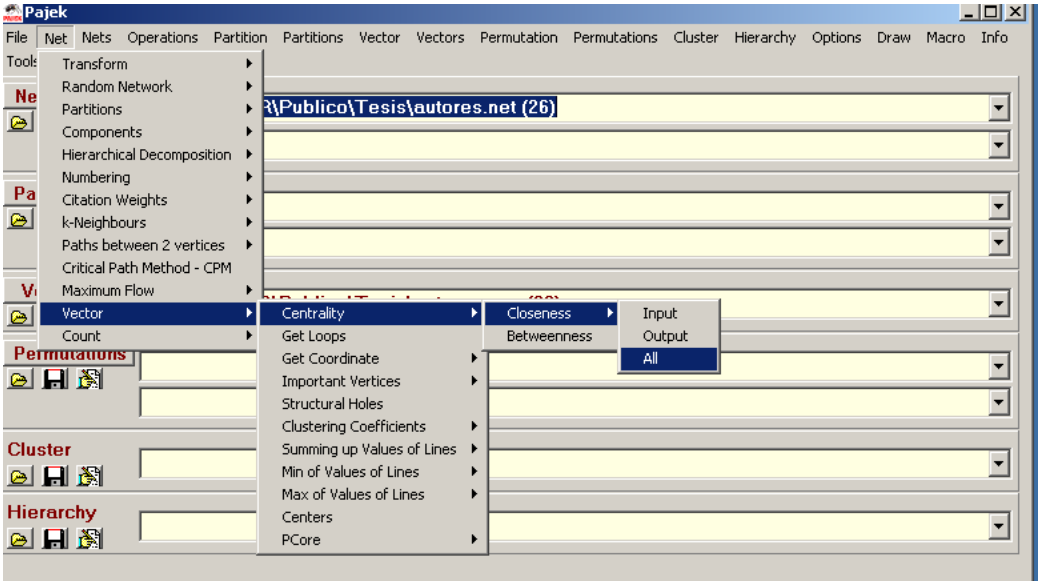


Figura 7.3: Transformar la red

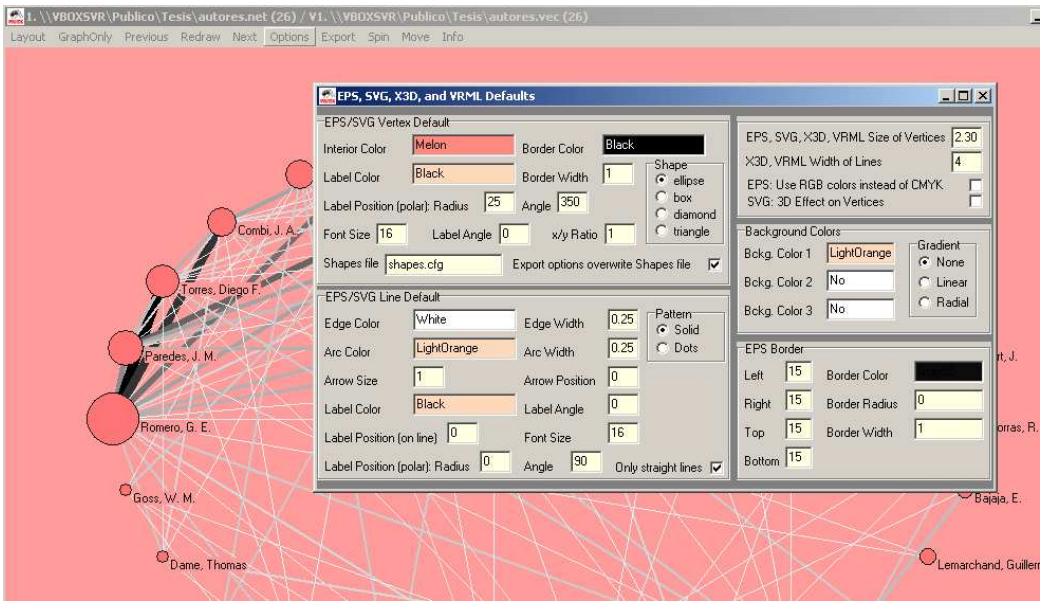


Figura 7.4: Opciones de visualización

awk[8] facilitó la extracción de los datos desde el archivo .csv generado por la hoja de cálculo, y la inclusión de los mencionados delimitadores. Se transcribe la línea que se aplicó para generar los datos de la tabla.

```
awk -F"\&" 'BEGIN{OFS="&"}{print $5,$6,$7,$8,$9"\\\\"}' \
fechas.csv
```

Para generar los gráficos se utilizó el graficador *Gnuplot*[9]. Se incluye a modo de ejemplo la lista de comandos ejecutados para generar el gráfico:

```
gnuplot> set title 'Modelo de Bradford aplicado a los últimos \
diez años de referencias'
gnuplot> set ylabel 'Acumulado de referencias'
```

[8]http://en.wikipedia.org/wiki/AWK
[9]http://www.gnuplot.info/

```
gnuplot> set xlabel 'Log. acumulado de revistas'
gnuplot> f(x)=x*a+b
gnuplot> fit f(x) 'tabla2_b' u 5:4 via a,b
gnuplot> plot 'tabla2_b' u 5:4,f(x) w l
```

7.6. Scripts utilizados

Script ref.sh

```
#!/bin/sh
# Chequeo de Sanidad
if [ $# -lt 1 ]; then
echo "ERROR: Debe especificar uno o varios nombres de archivos"
echo "Use:"
echo "  ref.sh cont-100.txt"
echo "  ref.sh *.txt"
exit 1
fi

# La parte interesante :-D
for file in $*; do
# Nos quedamos con la parte del nombre del archivo que nos interesa
# y le agregamos 4 ceros adelante :-D
filenumber=0000`echo $file|cut -d '-' -f 2|cut -d '.' -f 1`

sed /'TY  - '/i"NR  - $filenumber" $file
done
```

Script pct2id-tesis.pft

```
lw(8000),
select v1.5
```

```
case 'NR  -' : /#,'!ID 0'/, '!v1!',v1*6,
case 'TY  -' : /#,'!v2!',v1*6,
case 'T1  -' : /#,'!v3!',v1*6,
case 'A1  -' : /#,'!v4!'v1*6,
case 'JO  -' : /#,'!v5!',v1*6,
case 'VL  -' : /#,'!v6!',v1*6,
case 'Y1  -' : /#,'!v7!',v1*6,
case 'SP  -' : /#,'!v8!',v1*6,
case 'EP  -' : /#,'!v9!',v1*6,
case 'KW  -' : /#,'!v10!',v1*6,
case 'UR  -' : /#,'!v11!',v1*6,
case 'N2  -' : /#,'!v12!',v1*6,
case 'ER  -' : /#,'!v13!',v1*6,
elsecase       ,v1
endsel
```