

APROXIMACIÓN METODOLÓGICA AL ESTUDIO DE LAS REDES DE CITACIÓN

METHODOLOGICAL APPROACH TO STUDYING CITATION NETWORK

BUENO SÁNCHEZ, Ángel

Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE)

BUENO RODRÍGUEZ, Ángel

E.T.S. de Ingenierías Informática y de Telecomunicación

RAMÍREZ DE LA HIGUERA, Guadalupe

Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE)

Dirección postal del autor: Facultad de Ciencias de la Educación.

Campus Universitario de Cartuja, 18071 Granada

Fecha de finalización del trabajo: abril de 2008

Fecha de recepción por la revista: mayo de 2008

Fecha de aceptación: junio de 2008

C.D.U.: 37.014.25

BIBLID [ISSN: 0214-0484. *Rev. Educ. Univ. Gr.* (2008). Vol. 21:2, págs. 121-135].

RESUMEN

Este trabajo trata de una serie de aplicaciones que se utilizan en la representación gráfica de redes de investigación y producción científica. Específicamente, uno de los proyectos que vamos a utilizar es *Pajek V1.2*, que se enmarca dentro de la línea de investigación sobre representación de redes y sus aplicaciones a la distribución de la educación multicultural a nivel internacional, junto con otras herramientas asociadas. El encadenamiento de éste y otras aplicaciones nos permite llegar a una aproximación de unas redes de citación mediante su representación gráfica.

Palabras clave:

Estudio cuantitativo, redes de citación, metodología, investigación.

ABSTRACT

This work addresses a number of applications that are used in the graphic representation of networks and scientific research production. Specifically, one of the projects that we used was *Pajek V1.2*, which is part of the research on behalf of networks and their applications to the distribution of multicultural education at the international level, together with other associated tools. The chaining of this and other applications allows us to overview approximation of networks citation through its graphic representation.

*Key words:**Scientimetric study, citation networks, methodology, research.*

INTRODUCCIÓN

La cuestión metodológica subyacente en la evaluación de los informes de investigación, considerados éstos como un todo, no ha recibido tanto tratamiento como el diseño de la misma. Desde sus primeras consideraciones, la evaluación de la investigación se ha acometido a través de una aproximación exclusivamente cuantimétrica, centrada en el trabajo con indicadores de productividad, prioritariamente, y/o de citación, en menor medida; o sea, operando con indicadores bibliométricos. Recordemos que la Cuantimetría o Cientimetría, también denominada Metodología Bibliométrica (esas tres acepciones son sinónimas), es ante todo un corpus metodológico inicialmente centrado en el análisis estructural y prognosis de la información, contenida principalmente en publicaciones de investigación científica. López Piñero (1972) ya insistía en denominarla “ciencia de la ciencia”, considerándola como un auténtico programa de investigación de carácter interdisciplinar que interesaba a disciplinas complementarias: la información científica propia de los documentalistas, la historia de la ciencia y la sociología de la ciencia; a estas tres dimensiones podríamos añadir la evaluación científica de las producciones, propia de la Metodología de la Investigación.

López Piñero y Terrada (1992) asocian la Cuantimetría al estudio de una serie de indicadores de actividad científica relativos a la publicación de libros y folletos de temas científicos, a revistas científicas, a trabajos científicos circulantes a nivel nacional e internacional y a autores de las publicaciones. Existe una diversidad de indicadores bibliométricos disponibles, aunque los más utilizados en el contexto español son los de productividad.

Sin embargo, no deberíamos sacralizar los indicadores bibliométricos pues tienen también abundantes limitaciones como señalan Gómez y Bordons (1996). Tales indicadores son susceptibles de complementar, e incluso de corregir, mediante otras aproximaciones suplementarias a este peculiar fenómeno. Kragh (1989) enfatiza los tradicionales métodos historiográficos como medio para denotar el proceso dialéctico de transición de una evidencia cuantitativa de hechos nuevos, de nuevos métodos y experiencias, a unos cambios cualitativos en el propio contenido de la ciencia. También Kragh () señalaba que las técnicas cuantitativas daban una visión demasiado formalista de unos procesos que, en realidad, se hallan marcados por unas influencias nada formales y no racionales” (p. 256).

Es por ello por lo que los trabajos bibliométricos, por su naturaleza metodológica eminentemente descriptiva limitada a un análisis elemental del contenido inserto, pueden considerarse estudios paraevaluativos de la investigación indicada. Será el propio lector el que, tras reflexión crítica pertinente, realice su

valoración específica a la luz de criterios personales, hasta que no dispongamos de estudios comparativos que introduzcan un matiz referencial o de valor.

En concreto, el objetivo de este proyecto, que lleva a cabo Vladimir Batagelj and Andrej Mrvar (2007), de la Universidad de Ljubljana (Eslovenia), es desarrollar una herramienta para el análisis y visualización de grandes redes de datos que puedan tener algunos miles o millones de vértices, aspectos éstos últimos que ya han conseguido, continuando sus investigaciones en el refinamiento, potenciación y compatibilidad de dicha herramienta con otras posibilidades.

¿Qué implicaciones tiene todo esto para la investigación educativa? Inimaginables. La educación lleva tiempo necesitando programas que ayuden a presentar la información de forma significativa pues, en la actualidad, las investigaciones en educación, de corte fuertemente sociométrico, se quedan cortas a la hora de representar lo datos ya que se encuentran muy mediatizadas por las representaciones basadas en datos de tipo atributivo. Disponer de programas que presente volúmenes extensos de información que, a veces, evite que nos perdamos cuando intentamos tener una visión coherente de los datos, hallar explicaciones significativa de las relaciones entre los términos es de gran importancia.

La producción científica internacional en el área de la integración educativa está experimentado un importante crecimiento en los últimos años. Así, en el período 2000-2006 se identificaron 3.356 documentos realizados por autores a nivel internacional en las bases de datos del *Institute for Scientific Information* versión *ON-LINE*, con un incremento del 47,9%. Esto es, casi el 50% de las investigaciones sobre integración educativa se han realizado en lo últimos 7 años.

Por otra parte, la investigación educativa en general, está señalando un preocupante aumento progresivo de los trabajos realizados por un solo autor. La colaboración es necesaria para el avance del conocimiento porque los problemas requieren abordajes multidisciplinares, y es conveniente para poder complementar insuficiencias. En la investigación educativa se espera que colaboren fundamentalmente investigadores de las grandes áreas de la misma: didáctica, teoría e historia y métodos de investigación, pero también psicólogos, antropólogos, sociólogos, matemáticos, economistas... y de otras áreas del conocimiento. Sin embargo, esta importancia no viene acompañada de explicaciones y descripciones de los volúmenes que todas las disciplinas que precipitan en el campo de la educación aportan con investigaciones sobre los tópicos que llevan a cabo.

PAJEK, programa informático para el análisis de grandes redes, brinda la facilidad de representar gráficamente la situación de la producción científica, ayuda a conocer grupos y redes de trabajo, pudiendo ser analizada y evaluada mediante el empleo de técnicas bibliométricas y representada mediante lo que algunos autores denominan «mapas bibliométricos». El objetivo del presente trabajo es identificar y representar gráficamente la productividad acerca de la integración educativa.

MÉTODO

Para la recuperación de documentos, se han realizado sendas búsquedas en las internacionales *Social Sciences Citation Index (SSCI)* y *Arts & Humanities Citation Index (A&HCI)*, dos Bases de datos de *ISI Web of Science (WOS)*, producido por *Thomson Scientific* (anteriormente *Thomson-ISI*), durante el período: 1956-2007 (actualizada al 8 de julio de 2007).

Para la realización del estudio se han identificado los trabajos de investigación sobre *integración educativa* publicados durante el período 1956-2007 en dichas bases de datos. Para cada uno de los trabajos seleccionados se han identificado todos los datos disponibles en la redefinición de resultados: nombre y los apellidos de los autores, así como el centro de procedencia de éstos (institución, ciudad y país), etc. para posibles extracciones paracualitativas de la búsqueda.

Para la normalización de las autorías se han comprobado las firmas en las que un mismo autor firmaba de dos o más formas diferentes, utilizándose como criterio básico de unificación la existencia de coincidencia en los lugares de trabajo de las firmas. En el caso de las instituciones, se han unificado las diferentes variantes de facultades de las universidades y/o organismos.

Con la información extraída se ha utilizado el programa *Wos2Pajek V0.1*, que gestiona directamente los datos obtenidos de las búsquedas y, automáticamente, confecciona cuatro tipos de ficheros, listos para poder analizarse en *PAJEK*: *año de publicación de los trabajos* (.clu), *descripción de citas emitidas* (.clu), *red de citas* (.net) y *trabajos por autores* (.net), sin necesidad de crear una base de datos relacional, utilizando el programa Microsoft Access o parecido.

Para convertir los ficheros *WoS* en ficheros con formato *PAJEK*, se ha desarrollado el programa *Wos2Pajek* (en lenguaje *PYTHON*). Dicho programa produce los siguientes ficheros:

- Red de citación: trabajo × trabajo;
- Red de autorías (modo 2): trabajo × autor;
- Partición de los trabajos por el año de publicación;
- Partición de los trabajos–descripción completa (1) / solo nombres ISI (0).

En futuras versiones, los autores prometen que se irán desarrollando más opciones para el análisis de las redes: trabajos × discipline, trabajos × países, trabajos × descriptores, trabajos × revistas,...

La actual versión utilizada de *Wos2Pajek V01*, requiere que el usuario introduzca 6 parámetros:

El directorio de *Wos*: *path* al directorio en el cual el subdirectorío con el proyecto in se localiza;

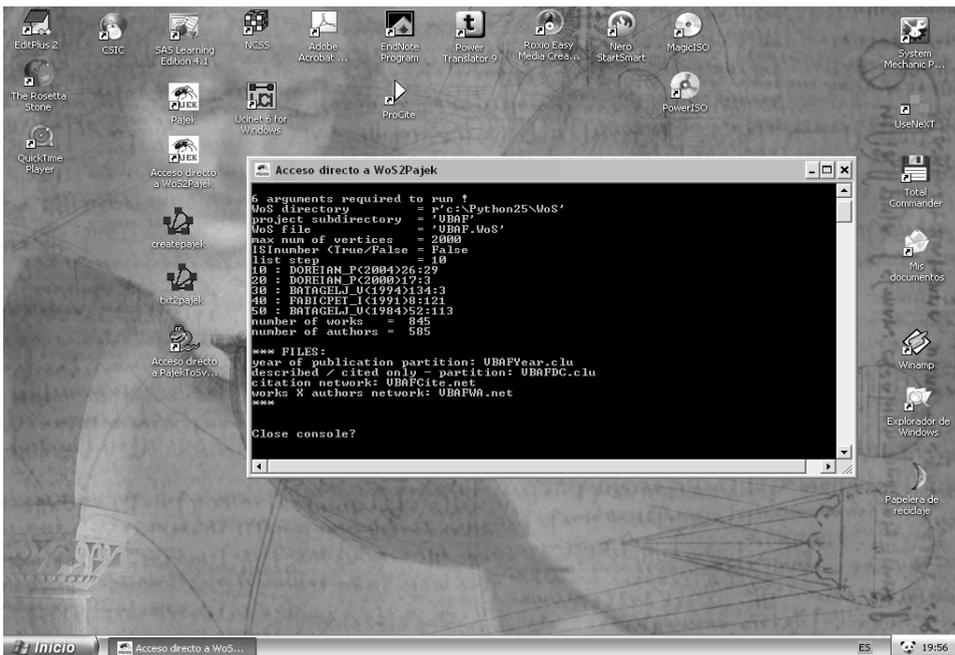
Nombre del subdirectorío del proyecto; se usa también como parte del nombre de los ficheros de salida;

- Nombre de fichero *WoS*;
- maxnum* – estimación del número de todos los vértices (número de trabajos + Número de trabajos citados);
- Uso de nombres *ISI* / nombres abreviados;
- Step* – información acerca de los $k^* \text{ step}$ (pasos) por un trazo; *step* = 0 – no hay trazo.

El programa fuente puede ser ejecutado de diferentes maneras, usando el poco amigable interprete de *PYTHON*. En este sentido, señalar que es un lenguaje muy potente a la hora de tratar las características de las bases de datos y de los formatos *ISI*, pues soporta tanto mayúscula como en minúscula, así como ajusta los prefijos de los nombres. En este sentido, *PYTHON* cuando trabaja con *Wos2Pajek* tiene en cuenta que los nombre pueden tener diferentes características por provenir de *ISI*. Para mejorar la precisión, *Wos2Pajek* soporta también nombres cortos (similar a los nombres usados en las salidas de *HISCITE*).

Paralos cálculo de los indicadores y la construcción de las redes se ha utilizando otro programa denominado *CreatePajek*, que traslada desde una hoja *MS Excel* los datos contenidos a formato *Pajek*.

GRAFICO 1: PANTALLA DE LA GENERACIÓN DE LOS FICHEROS POR EL PROGRAMA PYTHON



Los archivos generados por *Wos2Pajek* son los siguientes: *VBAFYear.clu*, *VBAFDC.clu*, *VBFAcite.net* y *AVBFAWA.net*, que son ficheros generados a partir

del introducido de nuestra búsqueda en *ISI*, una vez depurado de errores. Hemos mantenido los nombres de los ficheros según vienen definidos, en prevención de posibles dificultades dada la poca amigable facilidad que ofrece el intérprete de *PYTHON*.

Una vez obtenidos los ficheros, desde *PAGET* se realizó los siguientes “settings” o configuraciones al fichero cargado:

- Net/Partition/Degree/Input
- Partition/Binarize [1-(k-1)]
- Net/Partition/Degree/Output
- Partition/Binarize [0]
- [select partition 1]
- [select partition 2]
- Partitions/Min(V1,V2)
- Operations/Extract from Network/Partition [0]

RESULTADOS

a. Las bases de datos

En primer lugar, hemos realizamos una búsqueda en la base de datos *ERIC*, a través del interfaz *WebSPIRS Versión 5.12*. *WebSPIRS* (el sistema de recuperación de información en *Internet de SilverPlatter*) permite realizar búsquedas en las bases de datos de *SilverPlatter* a través de un navegador de Internet.

A fin de determinar la base de datos que mas se ajustara a nuestros objetivos, buscamos en el tesoro que nos proporcionara los mejores descriptores asociados a nuestro problema de búsqueda. Asociados a nuestro descriptor de búsqueda “integración educativa”, obtuvimos 18 bases de datos, con las cuales procesamos el gráfico 2.

Como nos muestra el gráfico, *ERIC*, *MEDLINE* y *GlobalHealth* son las bases de datos más aptas para la búsqueda de nuestros descriptores, tanto por la cercanía de los vértices y el valor de su “crossing” como por la cercanía de sus vértices a sus líneas. Aunque, por su cobertura, la mejor es *ERIC*. Se omiten los demás gráficos generados desde el “info” del programa, que resalta los vértices con diferentes colores cuando se cumple algunas de las características señaladas.

GRAFICO 2: RED DE BASES SILVERPLATTER

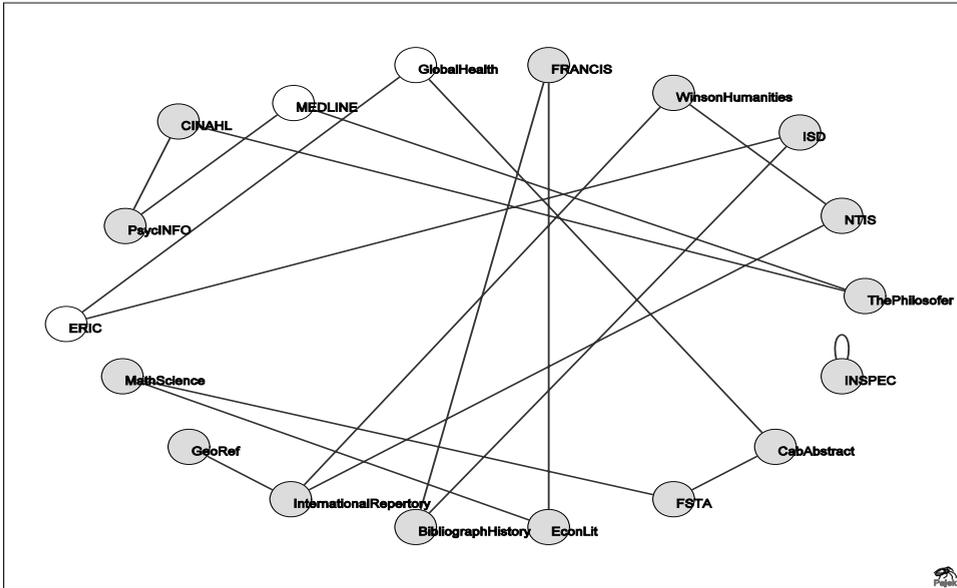
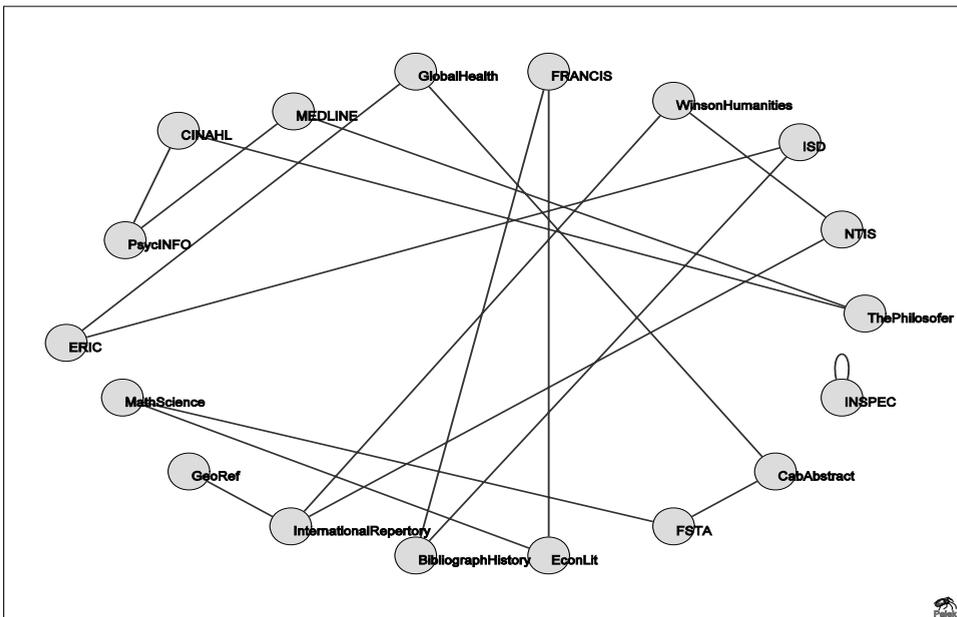


GRAFICO 3: NUMERO DE “CROSSING”



También se detectan dos “peatones” o bases de datos que no se encuentran conectadas o relacionadas con otras en lo referente a nuestro tópico de investigación: *GeoRef* e *INSPECT*

El total de registros recuperados en *ERIC* fue de 14.440, algo excesivo para poder gestionarlo con los medios disponibles. Además, el tratamiento de tantos datos no proporcionaría una gran diferencia informativa que si se trabaja con una base de datos que contenga el núcleo de la producción sobre nuestro problema; nos estamos refiriendo a *Thompson-ISI*.

b. Herramientas de representación de la información relacional

En segundo lugar, en nuestro intento de trabajar con todo el fichero recuperado de *ISIS*, hemos utilizado la primera versión del programa *Wos2Pajek*. Tras ciertas dificultades, se generó desde *PAJEK* el gráfico de citas desde nuestra base de datos, sin embargo, el programa apenas genera la red de citas (Ver gráfico 3).

Las causas pueden ser diversas, dado que esta es una versión experimental de *Wos2Pajek* (que además no es muy pródiga en explicaciones ni indicaciones), las deficiencias siempre acompañan a estos prototipos de reciente aparición.

GRAFICO 4: PANTALLA DE WOSTOPAGEK

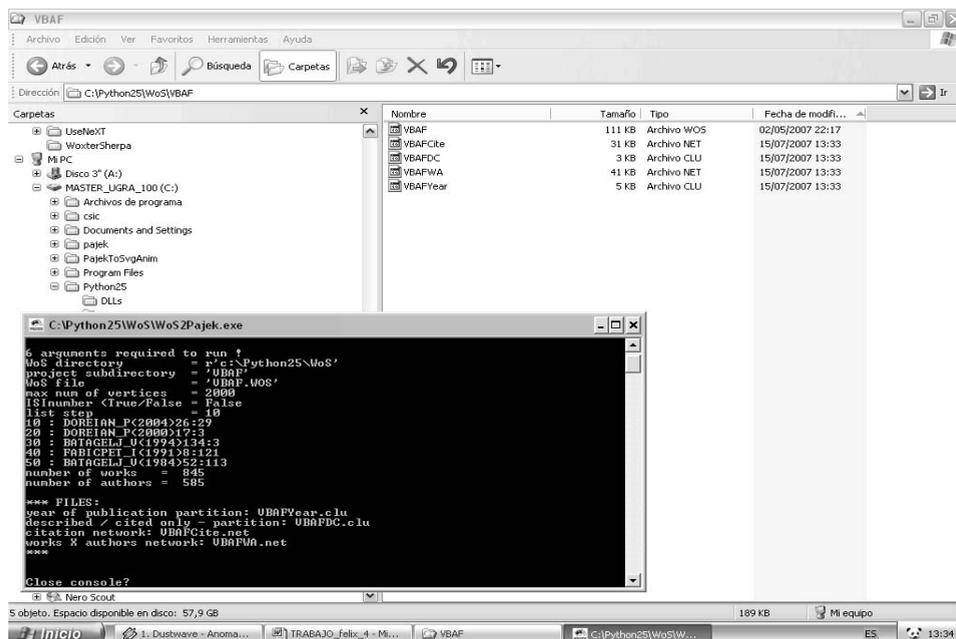
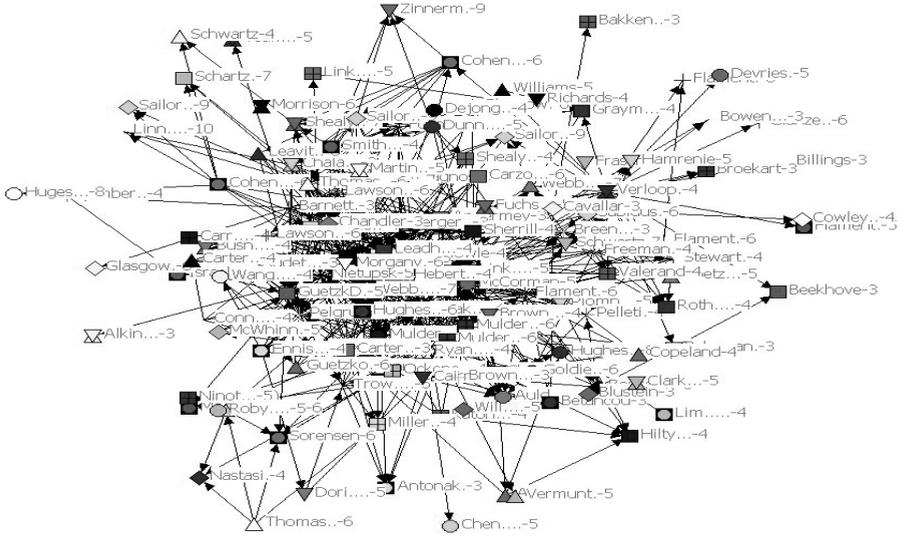
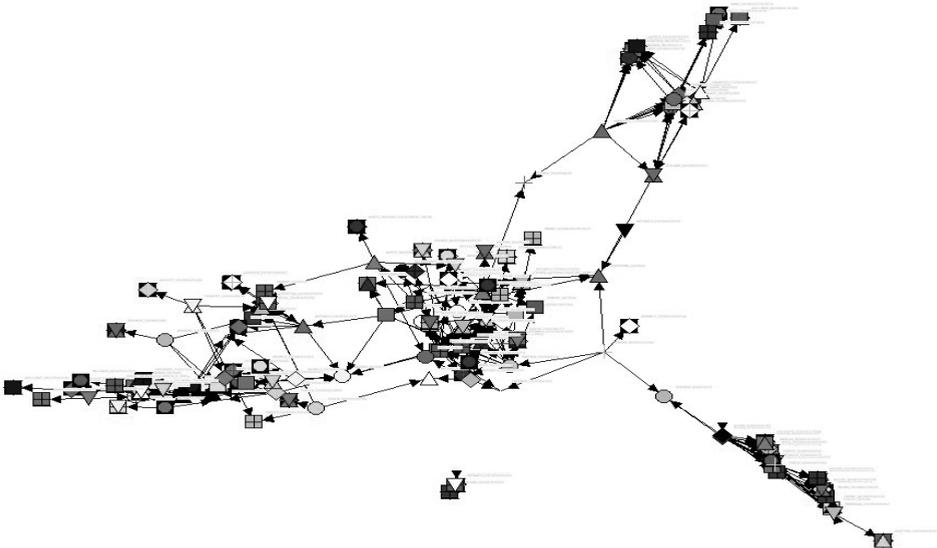


GRAFICO 5: RED DE CITACIONES EN INTEGRACIÓN EDUCATIVA CON UCINET



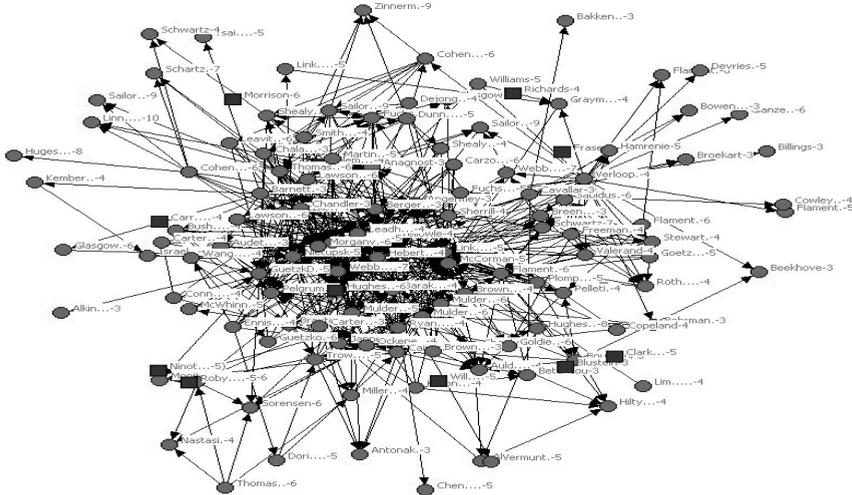
En esta segunda representación de la red, podemos observar ese núcleo compacto, que parece denotar una cohesión en el número de relaciones de las citas de los autores. Dado que este programa dispone de varias utilidades para explorar las redes, nos ha parecido interesante analizar algunas de ellas. En este sentido, vamos a analizar ciertas posibilidades que nos ofrece el programa *NETDRAW*, incluido en *UCINET*:

GRAFICO 6: RED G



Este gráfico de la red nos muestra una de las más importantes medidas de centralidad y la agrupación de los nodos según comparten sus relaciones. Este caso, se ve claramente el punto de corte que puede desconectar la red.

GRAFICO 7: ANÁLISIS DE ISLOTES



Este análisis denota los casos aislados en la red, muy útil para ver las redes centrales de la red y los agujeros estructurales. En este caso, la red denota poder, información, libertad...

GRAFICO 8: ANÁLISIS DE K CORES

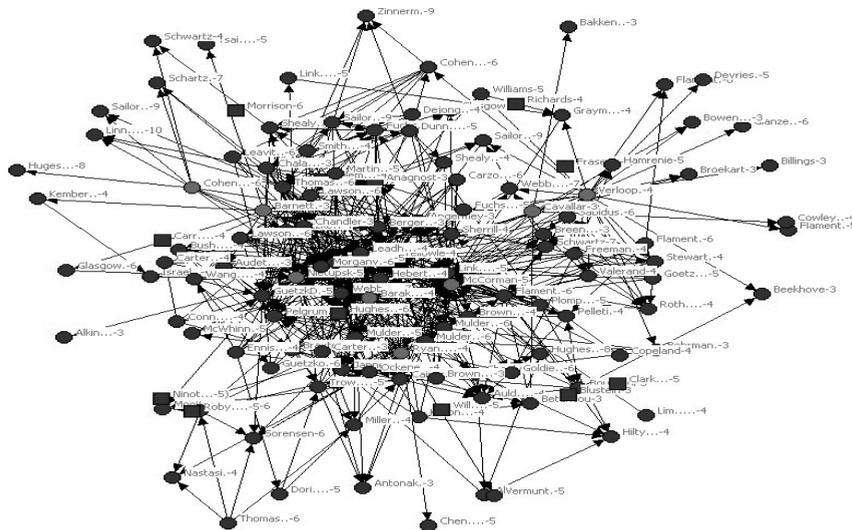
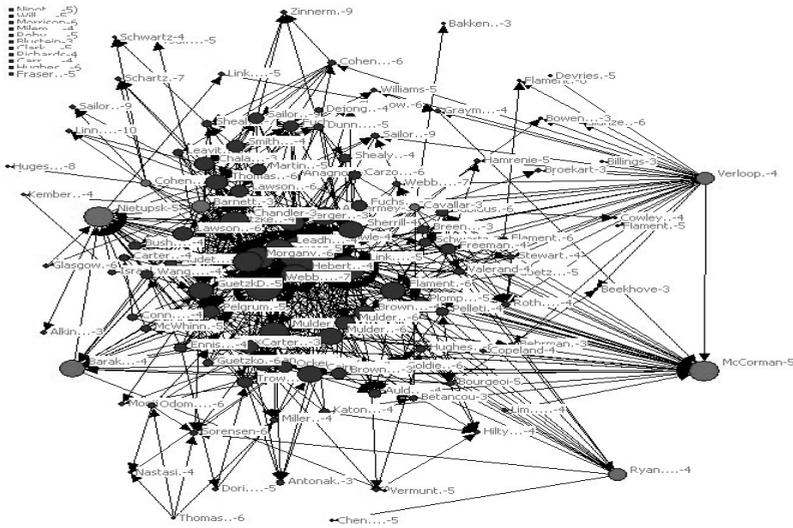
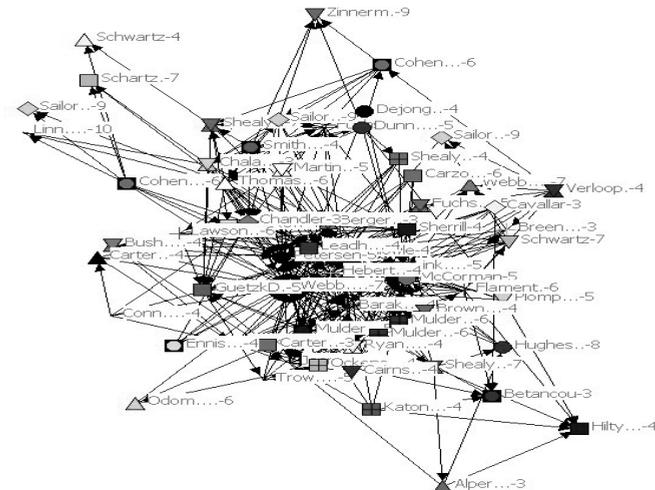


GRAFICO 11: ANÁLISIS COMPONENTES PRINCIPALES



El gráfico denota el número de componentes principales de la red, en color azul los emisores de citas y en rojo los principales receptores.

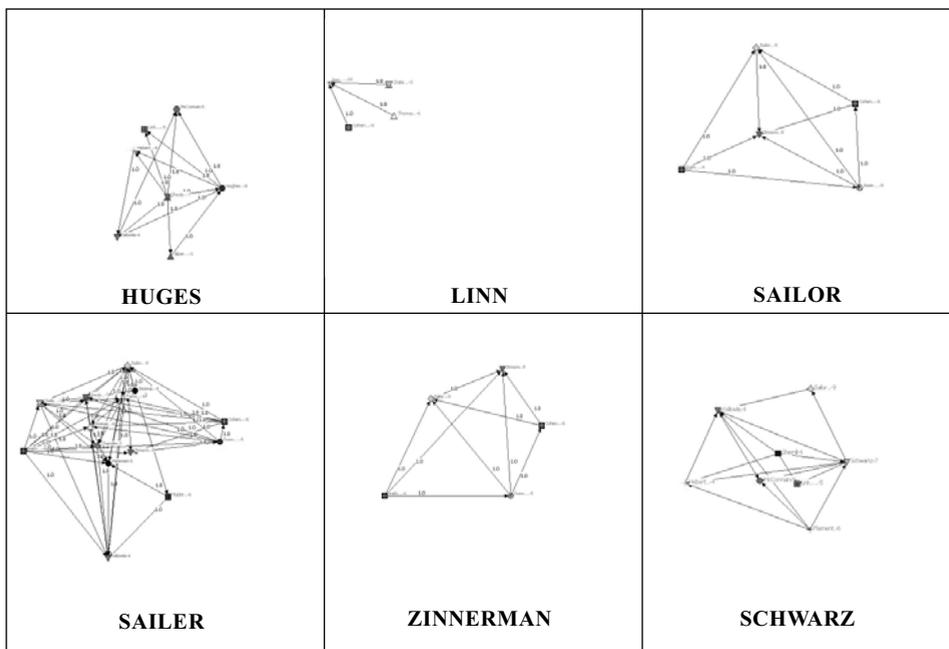
GRAFICO 12: EGOS PRODUCTIVOS



Como podemos ver en el gráfico, los autores más citados (que no productores) se sitúan en la periferia, recibiendo las influencias desde el núcleo de la red.

Finalmente, se ha pretendido comparar las subredes de los autores más citados, como se estructuran, que tipo de relaciones tienen, su densidad, número de componentes, etc.

TABLA 1: SUBREDES DE AUTORES MÁS CITADOS



Parece clara nuestra hipótesis de que cada una de las subredes correspondientes a los autores más citados tiene una estructura diferente. Los gráficos denotan que dichas relaciones son bastantes diferentes para casi todos ellos. En este sentido, Sailer y Zinnerman muestran una red con 4 componentes muy cohesionados, mientras que Hugues y Scharz los son de 7. La red con mayor riqueza en sus relaciones es la de Sailer, también muy cohesionada y sin muestra de puente alguno.

RESUMEN

La metodología aplicada ha permitido identificar los autores más productivos y la composición de 11 clústeres de autores de intensa colaboración en el área de integración educativa localizados en Thompson-ISI durante el período 1957-2007 (11 julio), caracterizando mediante indicadores bibliométricos su actividad científica. No obstante, es importante realizar algunas observaciones en relación con el método aplicado. Pues debido a las intenciones del estudio, solamente hemos identificado la citación, pero no las instituciones de dichos

autores y otros posibles indicadores de productividad y colaboración, que también son posibles.

En este sentido, en este breve análisis de redes sociales se han mostrado algunos de los conceptos que utilizan algunas herramientas para construir hipótesis relacionadas con diversos factores que determinan la estructura de las relaciones sociales (en cualquier grupo preseleccionado), aplicados a una muestra de autores que trabajan en el campo de la integración educativa.

El análisis de redes sociales es un método de análisis científico que puede ser de gran utilidad para conocer los patrones de relaciones que se establecen en el interior de una determinada estructura social. Se ha visto aquí una breve descripción de alguna de ellas aplicadas al campo de la educación, revelándose como un potente método complejo y sofisticado que puede ayudar a conocer ciertas cuestiones del campo educativo.

La metodología del análisis de redes sociales aplicada como una herramienta de diagnóstico, puede ayudar a enfrentar los desafíos asociados con la planificación e implementación de proyectos educativos, demostrando las tendencias institucionales, identificando las redes de los autores para evitar la replicación de esfuerzos y facilitando la construcción de capital social entre los diferentes actores. La visualización de redes sociales a nivel de comunidades y organizaciones permite comprender las redes locales mostrando las interacciones entre actores y las personas con influencia.

Por otro lado, el análisis de las subredes que conforman cada uno de los autores e investigadores tendría un gran valor descriptivo en educación, no solo a nivel para conocer los autores más importantes mediante la citación, como es en nuestro caso, sino también de la red social en su área en general.

REFERENCIAS

- Batagelj, V. and Mrvar, A. (2007). *Pajek - Analysis and Visualization of Large Networks*. In Juenger, M., Mutzel, P. (Eds.). Berlin: Graph Drawing Software. Springer (series Mathematics and Visualization).
- Garfield, E. (2007). *Histcite*. Disponible en: <http://www.histcite.com/>
- Gómez, I. y Bordóns, M. (1996). Limitaciones en el uso de los indicadores bibliométricos para la evaluación científica. *Política Científica*, 46, 21-26.
- ISI Web of Science (2007). Disponible en: <http://www.thomsonreuters.com/content>
- Kragh, H.S. (1990). *An Introduction to the Historiography of Science*. Cambridge, England: Cambridge University Press.
- López Piñero, J. M. (1972). *El análisis estadístico y sociométrico de la literatura científica*. Valencia: Centro de Documentación e Informática Médica/CSIC.
- López Pinero, J. M., y Terrada, M. L. (1992). Los indicadores bibliométricos y la evaluación de la actividad médico-científica. III. Los indicadores de producción, circulación y dispersión, consumo de la información y repercusión, *Medicina Clínica*, 98, 142-8.
- Python Software Foundation (2008). *Python Programming Language*. Disponible en: <http://www.python.org/>
- Thomas, H. (2007). Py2Exe. Disponible en <http://www.py2exe.org/>