

Título: Redes de coautoría en la ciencia colombiana: 1999-2004

Primer autor: Andrés Bernal

Dirección: Observatorio colombiano de ciencia y tecnología/Carrera 15 No. 37-59/Colombia/abernal@ocyt.org.co

Segundo autor: Eugenio Llanos

Dirección: Observatorio colombiano de ciencia y tecnología/Carrera 15 No. 37-59/Colombia/ellanos@ocyt.org.co

Resumen:

A partir de los datos referentes a la producción bibliográfica (artículos completos, artículos de revisión, comunicaciones cortas, etc.) de investigadores vinculados a instituciones colombianas registrados en las bases documentales de cuatro SIREs (Index Medicus, Science Citation Index, PsycINFO y CAB Abstracts) construimos una red de coautorías cuyos nodos son investigadores, existiendo un vínculo entre dos nodos si son coautores de un mismo producto bibliográfico, y hemos emprendido su caracterización en términos topológicos.

La red está constituida por 19352 nodos, y abarca el periodo 1961-2004. Su distribución de valencias se ajusta a una regresión potencial, señalando que en su crecimiento opera un mecanismo de vinculación preferente, comportamiento que ya se ha observado en este tipo de redes (Leydesdorff et al, 2005). La baja sensibilidad de la conectividad y distancia media entre nodos frente a la eliminación de nodos de alta valencia revelan una alta solidez. Adicionalmente realizamos un estudio dinámico, analizando las redes formadas en los periodos 1961-1999, 1961-2000.... y 1961-2004, encontrando que durante el periodo 1999-2004 la red ha crecido, incorporando nuevos investigadores, pero al mismo tiempo la distancia media entre nodos ha disminuido, comportamiento que fue observado anteriormente por Barabási (Barabási et al, 2002). Además, la conectividad y distancia media entre nodos no están determinadas por los nodos más centrales. Este resultado es un indicador del desarrollo de una fuerte comunidad científica, y muestra que un mecanismo de vinculación preferente en la investigación colectiva puede conducir a la formación de una red con alta solidez.

Eje temático: Sociedad y producción de conocimiento

Código: 2COL030

País: Colombia

Palabras clave: Redes sociales, redes de coautoría, cohesión, vinculación preferente

Proyectos de investigación: Valoración de los resultados globales de la política del Sistema

Nacional de Ciencia y Tecnología: actores, redes y coordinación

Período de investigación: Iniciación: 2005 ; Finalización:

Presentación

El fortalecimiento de la comunidad científica ha sido uno de los lineamientos estratégicos de la política nacional en ciencia y tecnología en Colombia (Colciencias, 1991). En este sentido, se ha privilegiado la investigación colectiva como modelo paradigmático de actividad investigadora. Sin embargo, hasta el momento en el país no se han adelantado estudios tendientes a la caracterización de las actividades de investigación colectivas, por lo que no se conocen sus dinámicas ni las formas que han adoptado. Para hacer un seguimiento adecuado de las actividades de investigación en el país resulta crucial indagar acerca de las diversas maneras en las que se organiza la investigación colectiva, conocer las formas en que se incorporan nuevos sujetos a la comunidad y en la que se posicionan los diferentes agentes.

Los enfoques de la teoría de redes sociales enfrentan estas preguntas acerca de la colaboración empleando como fuentes de información las coautorías de productos bibliográficos: ya que éstos constituyen el principal resultado de la investigación, las coautorías permiten evidenciar los patrones que toma la colaboración (Barabási, Jeong, Neda, Ravasz, Schubert y Vicsek, 2002;

Moody, 2004; Newman, 2001; Wagner y Leydesdorff, 2005). La presente contribución busca acercarse a la caracterización de patrones de colaboración científica en el país desde el enfoque de los análisis de redes sociales, empleando los datos relativos a los productos bibliográficos producidos por autores con afiliación institucional colombiana presentes en cuatro Servicios de Indexación y Resumen (SIREs).

Metodología y datos

Extrajimos todos los documentos de autores vinculados a instituciones colombianas de las bases de datos de Science Citation Index, Index Medicus, PsycINFO y CAB Abstracts. Estos cuatro SIREs abarcan aproximadamente el 40% de la producción bibliográfica de la ciencia nacional publicada en revistas indexadas internacionalmente en los últimos cuatro años (OcyT, 2005). Las bases de datos consultadas comprenden artículos científicos completos, artículos de revisión, comunicaciones cortas, cartas al editor, y abarcan el periodo 1961-2004.

A partir de estos datos construimos una *red de coautorías* cuyos vértices son los autores de tales documentos, existiendo una arista entre dos vértices si los investigadores respectivos son autores de un mismo documento. Adicionalmente, consideramos las redes generadas por los documentos publicados hasta cada uno de los años comprendidos en el periodo 1999-2004, con el propósito de analizar la evolución de la colaboración científica durante este periodo.

Para minimizar los errores debidos a las diferentes formas en que los autores escriben sus nombres, sometimos dichos nombres a un proceso de normalización previo a la construcción de la red:

1. Asociamos cada nombre con una expresión regular:

$$R_i = \text{Apellidos}.*\text{-primer nombre}.*\text{-segundo nombre}.*$$

2. Construimos el conjunto P de subconjuntos C_m del conjunto R de estas expresiones regulares definido como:

$P = \{C_m: \forall R_i, R_j \in C_m \text{ } s(R_i) \sim R_j \text{ o } s(R_j) \sim R_i \text{ y } \forall C_m, C_n \in P \text{ no se satisface } C_m C_n \text{ o } C_n C_m\}$,
donde $s(R_i)$ es la cadena de caracteres que representa a la expresión regular R_i .

3. Los miembros de cada pareja de subconjuntos C_m, C_n con $n \neq m$ tales que $C_m \cap C_n \neq \emptyset$ fueron distribuidos en los siguientes subconjuntos de R:

a) $C_m - C_n$,

b) $C_n - C_m$,

c) N conjuntos unitarios, cada uno conteniendo uno de los miembros de $C_m \cap C_n$.

4. El paso anterior se repitió hasta que P fuera una partición de R, esto es, $C_m \cap C_n = \emptyset \forall C_n, C_m \in P, n \neq m$.

Los nombres asociados con cada uno de los miembros de una misma clase C_m fueron considerados equivalentes, y los identificamos con el nombre más largo de la clase.

Evaluamos tres propiedades de la red de coautorías: conexidad, centralidad y distancia media entre nodos. La conexidad y la distancia media están relacionadas con la intensidad de la colaboración y la facilidad que tiene la información de circular a través de los nodos de la red (Wasserman y Faust, 1999). La centralidad de un nodo está relacionada con el prestigio de un actor, en la medida que un nodo conectado con una gran cantidad de nodos en la red representa un actor visible y reconocido dentro de su comunidad. La centralidad como medida global permite determinar si la red está construida alrededor de un pequeño número de actores de elevado prestigio.

Como indicadores de centralidad consideramos tres magnitudes:

- Valencia: Un nodo puede considerarse central en la medida que esté conectado directamente a un gran número de nodos (Wasserman y Faust, 1999). La forma de la curva de distribución de valencias permite diferenciar una red altamente centralizada, en la que una minoría de nodos abarca un gran número de conexiones (i.e. como sucede cuando opera un

mecanismo de vinculación preferencial), de una con baja centralidad, en la que todos los nodos se encuentran conectados aproximadamente al mismo número de nodos.

- *Closeness centrality*: Un nodo puede considerarse central en la medida en que se encuentre a una distancia muy corta de un gran número de autores (Wasserman y Faust, 1999). Por tanto, puede emplearse la distancia media de un nodo a todos los demás nodos de la red como un indicador de su centralidad. La suma de las desviaciones en el valor de este indicador calculado sobre todos los nodos de la red será a su vez un indicador de su centralidad global.
- *Betweenness centrality*: Un nodo puede considerarse central si forma parte de un gran número de caminos de mínima distancia entre parejas de nodos en la red, pues esto sugiere que es crucial para la comunicación entre los distintos actores que la conforman (Wasserman y Faust, 1999). El número de caminos geodésicos de las cuales forme parte un nodo será entonces un indicador de su centralidad. La suma de las desviaciones del valor de este indicador sobre todos los nodos de la red será a su vez un indicador de su centralidad global.

Por último, evaluamos la dependencia de la conexidad de la red y de la distancia entre sus nodos frente a la eliminación de los nodos más centrales. Para ello eliminamos progresivamente el 30% de los nodos de la red, empezando por los nodos más centrales, calculando en cada etapa el tamaño de la máxima componente de la red resultante y la distancia media entre sus nodos. Este cálculo nos permite evaluar si la estructura de la red está determinada por los autores más centrales.

Resultados

1. La productividad de la red es muy baja

En la figura 1 se trazó la distribución de número de documentos publicados por investigador. Se observa que el 64% de los investigadores han publicado un solo artículo durante el periodo 1961-2004. El promedio es de 2.39 artículos por investigador en 43 años, valor que está muy por debajo de los promedios internacionales calculados por Newman (Newman, 2001).

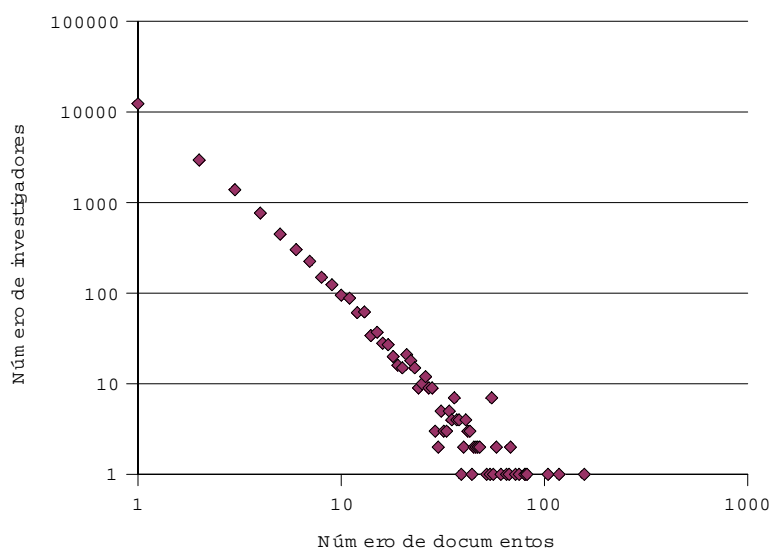


Figura 1: Número de documentos por investigador 1961-2004

2. Número de autores por documento

En la Figura 2 se trazó la distribución de autores por documento. El valor medio de autores por producto es igual a 3.77, valor que coincide con el promedio internacional calculado por Newman (Newman, 2001).

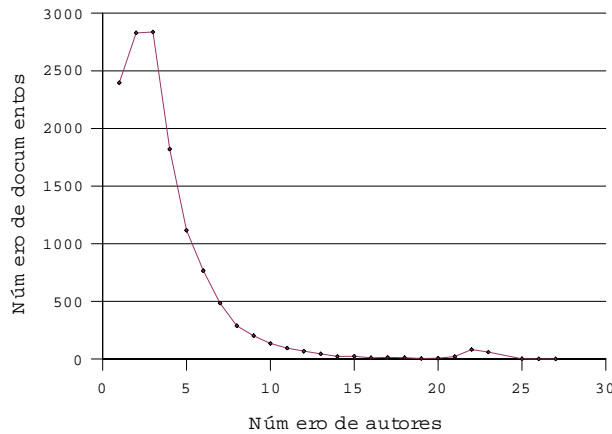


Figura 2: Número de autores por documento

3. La distribución de valencias es característica de un mecanismo de vinculación preferente.

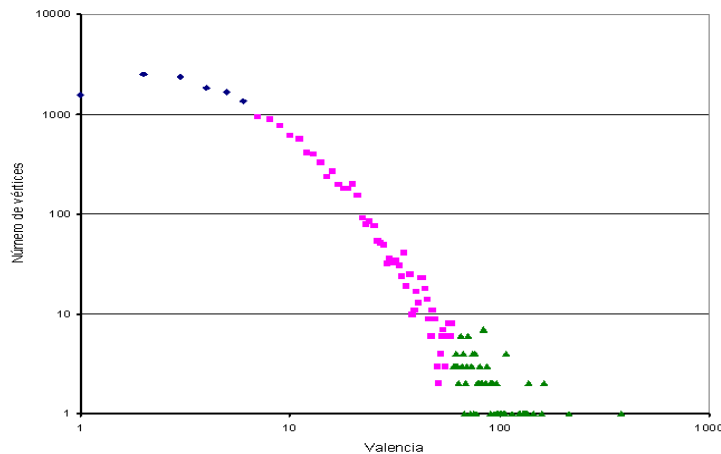


Figura 3: Distribución de valencias. Escala logarítmica

En estudios anteriores sobre la estratificación en las ciencias se han identificado grandes desigualdades (Merton, 1968; Allison, Long y Karuze, 1982). “Aunque la mayoría de los científicos trabajan en la oscuridad, un pequeño número de científicos recibe un reconocimiento desproporcionado” (Moody, 2004). Esta desigualdad se extiende a la colaboración, donde la mayoría de los investigadores tiene un grupo reducido de colaboradores, mientras un pequeño número de investigadores está rodeado por una gran cantidad de colaboradores (Newman, 2000; Leydesdorff y Wagner, 2005).

Este comportamiento ha sido explicado por Barabási y Albert (Barabási y Albert, 1999) mediante un modelo de *vinculación preferente*, donde la probabilidad de que un autor publique un artículo en colaboración con otro autor es proporcional al número de investigadores con quienes ha publicado un artículo anteriormente. La distribución de valencias de una red de coautorías que exhibe este comportamiento se ajusta a una ley de potencias $f(x)=Ax^{-\tau}$.

En la Figura 3 se encuentra la gráfica, en escala logarítmica, de la distribución de valencias de la red de coautorías 1961-2004. Se observa que la región media (color violeta) de la gráfica se ajusta razonablemente a una ley de potencias, presentándose desviaciones en las regiones de bajas (color

azul) y de altas valencias (color verde). Estas desviaciones también pueden ser explicadas mediante el modelo de vinculación preferente (Barabási, Jeong, Nédá, Ravasz, Schubert y Viczek, 2002; Leydesdorff y Wagner, 2005). Este comportamiento ya habla de una red con un alto grado de centralidad.

4. Alrededor del 70% de los nodos están en una sola componente y su tamaño aumenta en el tiempo

En la figura 4 se trazaron los gráficos de distribución de vértices en componentes según su tamaño, para cada uno de los años en el periodo analizado. Hay tres fenómenos que cabe destacar: Primero, en todos los años aparece una componente que abarca alrededor del 70% de los nodos y es tres órdenes de magnitud más grande que la segunda mayor componente. Segundo, la fracción de nodos comprendidos en esta componente aumenta a través de los años. Tercero, la fracción de nodos comprendidos en las componentes de menor tamaño (entre 1 y 10 nodos) disminuye en el tiempo. Esto indica que a medida que aumenta el tamaño de la red, incorporando nuevos investigadores, también aumenta su conexidad, es decir, la colaboración entre investigadores.

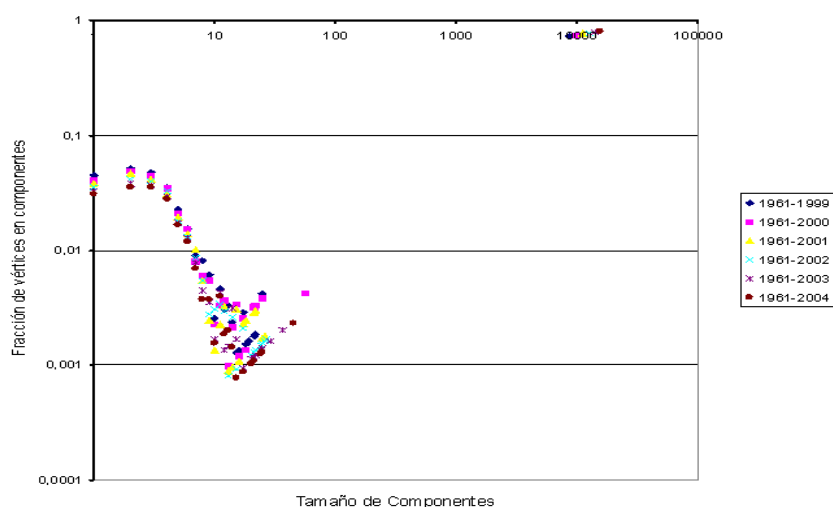


Figura 4: Variación en la distribución de componentes 1999-2004. Escala logarítmica

5. La centralidad de la red aumenta en el tiempo.

En la figura 5 se trazaron los gráficos de variación de los valores de *closeness centrality* y *betweenness centrality* a lo largo del periodo 1999-2004. Aunque se observa una caída en el último año, la tendencia global es de un aumento en la centralidad de la red, bien sea entendida como la existencia de un conjunto reducido de nodos cercanos a los demás nodos de la red, bien sea entendida como la facultad de estos nodos de vincular a los demás miembros de la red a través de caminos geodésicos. Este comportamiento es el predicho por el modelo clásico de vinculación preferente, pues en la medida que unos pocos investigadores acumulen una gran cantidad de colaboradores, su carácter central dentro de la red irá en aumento.

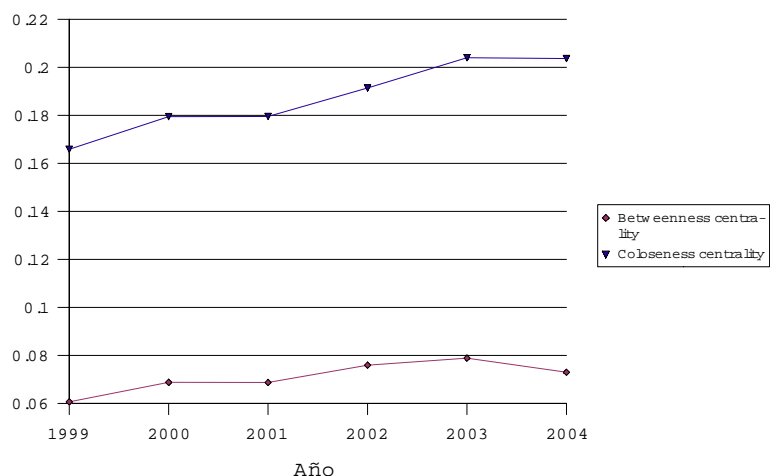


Figura 5: Variación de la centralidad 1999-2004.

6. La valencia media aumenta y la distancia media entre nodos disminuye en el tiempo.

Este resultado, atípico dentro del modelo de vinculación preferente (Barabási, Jeong, Néda, Ravasz, Schubert y Viczek, 2002), es señal de un fuerte incremento en la colaboración: No sólo aumenta el número de investigadores conectados, sino que además se encuentran más cercanos y su número de colaboradores es cada vez más elevado. Puesto que además el mecanismo de vinculación preferente conduce a un aumento en la centralidad, se esperaría que estas características estuvieran determinadas por los nodos más centrales de la red, quienes actuarían como puentes que vinculan investigadores que de otra forma no estarían conectados, o que disminuyen la distancia que los separa. La eliminación de estos investigadores centrales tendría entonces un efecto dramático sobre la estructura de la red.

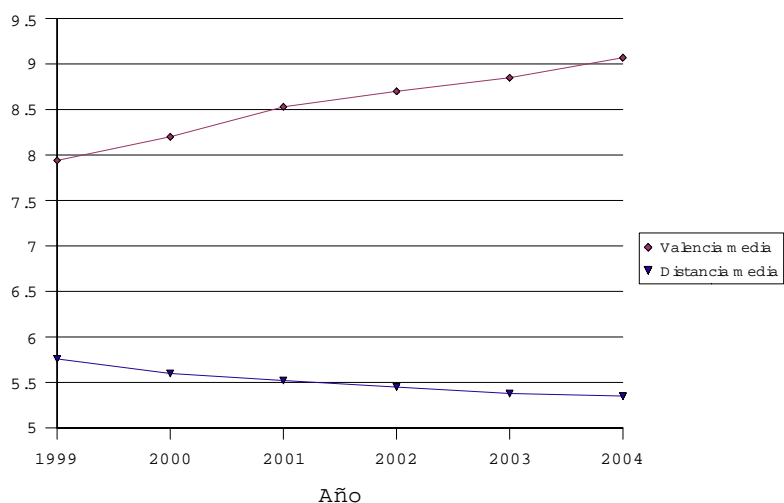


Figura 6: Variación de la valencia media y de la distancia media 1999-2004

7. La conexidad y distancia media entre nodos no están determinadas por los actores más centrales.

Para confrontar la hipótesis arriba planteada, eliminamos sucesivamente los nodos de más alta valencia de la red, calculando en cada etapa el tamaño de la máxima componente y la distancia media entre nodos en las redes resultantes. Realizamos este proceso para cada uno de los años en el periodo analizado. Las curvas obtenidas se encuentran en las figuras 7 y 8.

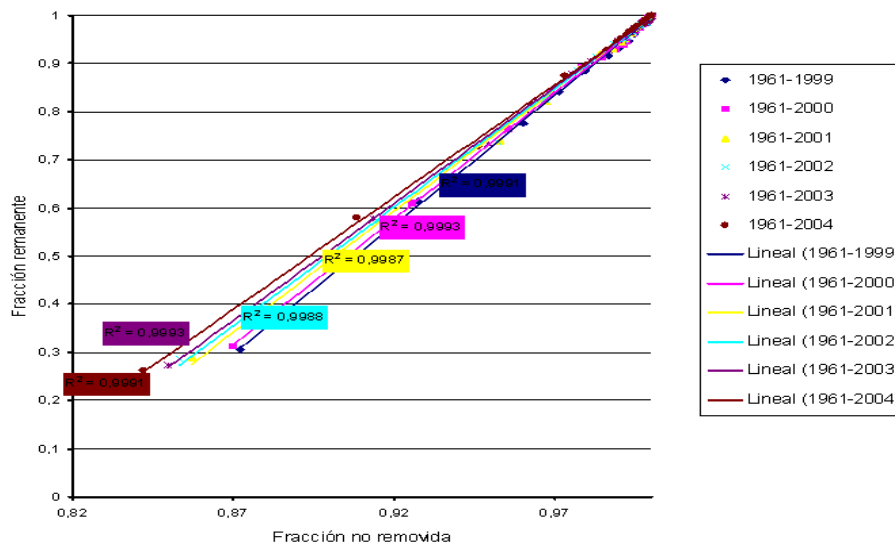


Figura 7: Efecto de la eliminación de los vértices centrales en la conexidad de la red, 1999-2004. La fracción no removida es igual a la fracción de nodos que no fueron eliminados explícitamente de la red original. La fracción remanente se define como la relación entre el tamaño de la máxima componente de la red resultante y el tamaño de la máxima componente antes de eliminar los nodos.

El comportamiento lineal de estas gráficas muestra que, en promedio, el efecto de la eliminación de un nodo sobre la estructura de la red es independiente de la centralidad de dicho nodo. Esto sugiere una fuerte solidez, pues la estructura de la colaboración no se pierde debido a la ausencia de unos pocos investigadores centrales¹. Más aún, puede observarse que la pendiente de las rectas en la figura 7 disminuye en el tiempo, lo cual indica que esta solidez va aumentando, esto es, que el efecto medio de la eliminación de un investigador sobre la estructura de la red es cada vez menor.

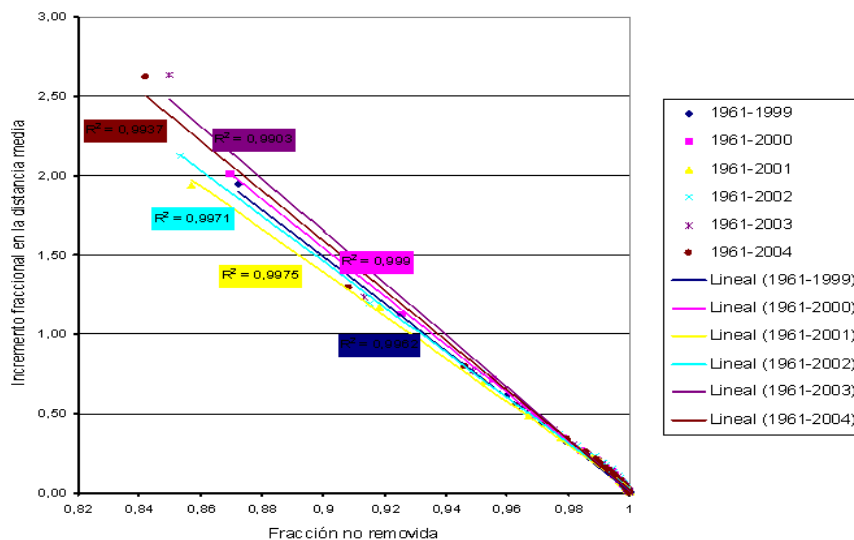


Figura 8: Efecto de la eliminación de los vértices centrales en la distancia media entre nodos, 1999-2004. La fracción no removida es igual a la fracción de nodos que no fueron eliminados explícitamente de la red original. El incremento fraccional en la distancia media es igual al cociente entre la distancia media en la red resultante y la distancia media en la red antes de

¹Este resultado debe interpretarse con cuidado, pues es legítimo preguntar si dos investigadores habrían colaborado para publicar un artículo en ausencia de un tercer investigador central, esto es, de elevado prestigio. Esta pregunta debería ser resuelta desde otros enfoques metodológicos.

eliminar los nodos.

Bibliografía:

Barabási, Albert-Laszlo; Reka, Albert (1999). Emergence of Scaling in Random Networks. *Science*, 286,509

Barabási, A ; Jeong, H ; Neda, Z ; Ravasz, E ; Schubert, A ; Vicsek, T (2002). Evolution of the social network of scientific collaborations. *PHYSICA A* 311, 590

Colciencias (1991). Ciencia y Tecnología para una Sociedad Abierta. Colciencias.

Moody, James (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review*, 69,4,238

Newman, M.E.J (2001). *The structure of scientific collaboration networks*. Proceedings of the National Academy of Sciences USA 98, 404

Observatorio Colombiano de Ciencia y Tecnología (2005). Indicadores de Ciencia y Tecnología 2004. En prensa.

Wagner, Caroline ; Leydesdorff, Loet (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy*, in press

Wasserman, Stanley; Faust, Katherine (1999). *Social Network Analysis: Methods and Applications*. Cambridge University Press, 178-188.