

Tipo de artículo: Artículo de revisión
Temática: Inteligencia Artificial
Recibido: 11/12/2017 | Aceptado: 22/01/2018

Análisis estático de la red de colaboración científica en la Revista Cubana de Ciencias Informáticas

Static analysis of the scientific collaboration network in the Cuban Journal of Computer Science

Rafael Bello Lara *

Trabajador no estatal, Unión de Informáticos de Cuba.

*Autor para correspondencia: r3ello@gmail.com

Resumen

La manera en la que se hace ciencia actualmente va más allá de las fronteras entre los países, lo que ha hecho necesario realizar análisis sobre las diferentes publicaciones (personas, instituciones, países), cuánto producen, qué tan usado es lo producido (citas), cómo colaboran los investigadores o instituciones (redes). Conocer esta información les permite a los financiadores ubicar a los investigadores o instituciones que tienen el número más alto de publicaciones o mejor calidad de las mismas, pues son candidatos a recibir mayor financiamiento, así como saber cuáles son las redes o comunidades de colaboración para estimularlas o generar nuevas áreas de colaboración. La Revista Cubana de Ciencias Informáticas contiene artículos de investigadores del país y en muchos casos investigadores extranjeros. El presente trabajo muestra un análisis estático de unos de los componentes de la red de colaboración científica formada por autores de artículos en dicha revista, se propone un operador de agregación que combina las ventajas del operador D-OWA y WA en el cálculo de medida de centralidad compuesta. El resultado de este análisis muestra el nodo (autor) más importante en la red a partir de la medida de centralidad compuesta aplicada y se visualiza la red en escalas de colores según la importancia de los nodos.

Palabras clave: análisis estático, operador de agregación, medidas de centralidad, red de colaboración científica.

Abstract

The way in which science is currently done goes beyond the borders between the countries, which has made it necessary to analyze the different publications (people, institutions, countries), how much they produce, how used they are produced (citations), how researchers or institutions (networks) collaborate. Knowing this information

allows the funders to locate the researchers or institutions that have the highest number of publications or better quality of them, since they are candidates to receive more funding, as well as to know which are the networks or collaboration communities to stimulate them. or generate new areas of collaboration. The Cuban Journal of Computer Science contains articles by researchers from the country and in many cases foreign researchers. This paper shows a static analysis of some of the components of the scientific collaboration network formed by authors of articles in said journal, an aggregation operator is proposed that combines the advantages of the operator D-OWA and WA in the calculation of measurement of compound centrality. The result of this analysis shows the most important node (author) in the network from the measure of applied centrality and the network is visualized in scales of colors according to the importance of the nodes.

Keywords: *static analysis, aggregation operator, centrality measures, scientific collaboration network.*

Introducción

El análisis de redes permite reconocer, evaluar, definir y representar las estructuras sociales subyacentes en base a las relaciones establecidas entre personas. Este análisis no es algo nuevo, pero tiene un gran auge con el mayor uso de Internet al establecerse nuevas redes que conectan a personas de forma mucho menos lineal, con estructuras que son a la vez más líquidas y mutables y, en consecuencia, más difíciles de descifrar.

Históricamente, el análisis de redes aparece como una de las primeras disciplinas en usar la teoría de grafos para hacer ciencia fuera de las matemáticas (SCOTT, 1991). El análisis de redes sociales comenzó creciendo lentamente, a veces a grandes saltos (MILGRAM, 1967; TRAVERS, et al., 1969). Actualmente ha invadido la investigación sobre la Web y de cómo se hace ciencia (BOCCALETTI, 2006), un ejemplo son las redes de colaboración científica, teniendo en cuenta que la manera en la que se hace ciencia actualmente va más allá de las fronteras entre los países, lo que ha hecho necesario realizar análisis sobre las diferentes publicaciones (personas, instituciones, países), cuánto producen, qué tan usado es lo producido (citas), cómo colaboran los investigadores o instituciones (redes), entre otros aspectos; los cuales son estudiados como parte del comportamiento social de la ciencia a través de indicadores bibliométricos (BORDONS, et al., 1999; PRAT, 2001). Conocer esta información les permite a los financiadores ubicar a los investigadores o instituciones que tienen el número más alto de publicaciones o mejor calidad de las mismas, pues son candidatos a recibir mayor financiamiento, así como saber cuáles son las redes o comunidades de colaboración para

estimularlas o generar nuevas áreas de colaboración; además, permite tener indicadores basales para generar políticas que posibiliten luego conocer su impacto (GEROLIN, 2010).

El presente trabajo expone el análisis estático a la red de colaboración científica formada por los autores de la Revista Cubana de Ciencias Informáticas, proponiendo el operador de agregación D-OWAWA para obtener una medida de centralidad compuesta. El trabajo está organizado de la siguiente manera la sección 2 contiene bases conceptuales del análisis de redes, análisis estáticos, medidas de centralidad y el operador D-OWAWA, la sección 3 está dedicada a mostrar la aplicación del análisis y la discusión de los resultados y en la sección 4 se muestran las conclusiones y futuros trabajos.

Desarrollo

1. ANALISIS DE REDES.

El Análisis de Redes Sociales (ARS) no solo cuenta con un conjunto importante de métodos y algoritmos analíticos, sino que también posee un sistema teórico conceptual propio de naturaleza relacional, que busca generar explicaciones acerca de los fenómenos sociales, centrándose en los patrones y sistemas de relaciones que los constituyen (y no en individuos, grupos, atributos, o categorías). Sin embargo, no existe un acuerdo generalizado sobre muchos de sus conceptos teóricos, métodos analíticos y posiciones epistemológicas (EMIRBAYER J, 1994).

El ARS se focaliza en la estructura reticular de las relaciones sociales, buscando dar cuenta del efecto de las relaciones en el comportamiento de los individuos, los distintos grupos sociales y la sociedad en su conjunto. Abordar los procesos sociales desde una dimensión relacional implica dar cuenta de los patrones y estructuras de las relaciones sociales, y el ARS busca estudiar estos desde su particular configuración en redes (AGUIRRE, 2011). El objeto de estudio del ARS son las redes sociales, su morfología, sus patrones de comportamiento general y su relación dialógica con los individuos que las componen. (BORGATTI, 2009)

Algunas palabras claves básicas de la terminología de ARS son: actores (nodos o vértices) y relaciones (vínculos o aristas). Típicamente, un grafo se representa mediante una serie de puntos que son los vértices, conectados por líneas llamadas aristas.

2. ANÁLISIS ESTÁTICO.

El análisis estático permite selección de los conceptos o las conexiones en los cuales intervenir a partir de la estructura de la red, posibilitando determinar en qué aspectos del sistema incidir o reducir la cantidad de criterios que se analizan. Las propuestas existentes se basan en el análisis a partir de una sola medida de centralidad (centralidad de

grado) (ALTAY, 2011), o la prescripción de un número reducido de estas (OBIEDAT, et al., 2011) a pesar de existir un grupo amplio de medidas (ARCO-GARCÍA, 2008). Este hecho afecta la fiabilidad en el análisis debido a que la determinación de los nodos más importantes de una red puede ser un problema multicriterio (H. JUN, 2010), por lo que limitando el número de criterios se puede llevar a una decisión errónea (M. LEYVA-VÁZQUEZ, 2012).

MEDIDAS DE CENTRALIDAD

La Tabla 1 muestra diferentes medidas de centralidad usadas en el análisis estático.

Tabla 1: Medidas de centralidad más utilizadas para el análisis estático de grafos. (BELLO-LARA, 2010)

Formulación matemática	Descripción
$C(v) = id(v) + od(v)$	Centralidad grado nodal
$C_c(v) = \frac{1}{\sum_{t \in V} d_G(v, t)}$	Centralidad por cercanía
$C_B(v) = \sum_{s \neq v, t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$	Centralidad por intermediación
$C_{VP}(v) = \lambda \sum_j a_{i,j} C_j$	Centralidad de vector Propio
$C_{Bon}(v) = \frac{1}{\beta} \sum_{k=1}^{\infty} \sum_{j=1}^n \beta^k (A^k)_{i,j}$	Centralidad de Bonacich
$C_{Katz}(v) = \sum_{k=1}^{\infty} \sum_{j=1}^n \alpha^k (A^k)_{i,j}$	Centralidad de Katz
$C_{Cca}(v) = \sum_{j=1}^n 2^{-(S)u}$	Centralidad por cercanía de camino aleatorio
$C_{Bn}(v) = \sum_{i \neq j \neq k} R_{j,k}^{(i)}$	Centralidad por intermediación de Newman

MEDIDAS SUGERIDAS

A continuación, se describen en detalles las medidas de centralidad recomendadas:

- Centralidad de grado.

La centralidad de grado ($C(v)$) se calcula a partir de la suma de su grado de entrada ($id(v)$) y grado de salida ($od(v)$), tal como se expresa en la siguiente ecuación:

$$C(v) = id(v) + od(v) \quad (1)$$

La centralidad en un grafo indica cuán fuertemente está relacionado un nodo con otros a partir de sus conexiones directas.

- Intermediación.

La intermediación se calcula mediante la siguiente expresión:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (2)$$

Donde σ_{st} representa el número de caminos de longitud menor desde el nodo s hasta el nodo t y $\sigma_{st}(v)$ es el número de caminos de menor longitud que van de s a t a través de v . En un grafo indica la importancia de un nodo en el flujo de la información (OBIEDAT, y otros, 2011).

- Cercanía.

La cercanía se define como:

$$C_c(v) = \frac{1}{\sum_{t \in V} d_G(v, t)} \quad (3)$$

donde $t \neq v$ y $d_G(v, t)$ es la distancia del camino más corto entre v y t . En el caso de un grafo brinda información sobre cuán rápido se difunde la información de un nodo por la red (OBIEDAT, y otros, 2011). Se podrán tener en cuenta además otro conjunto de medidas como la centralidad por vector propio (BORGATTI, 2005) y la centralidad de Bonacich (CRIADO, y otros, 2012). El decisor podrá tomar un conjunto de estas medidas de centralidad en dependencia de los factores que desea tener en cuenta.

Operadores de agregación.

Los operadores de agregación son un tipo de función matemática empleada para la fusión de la información.

Combinan n valores en un dominio D y devuelven un valor en ese mismo dominio (TORRA, et al., 2007).

Denominando esas funciones \mathbb{C} , los operadores de agregación son funciones de forma:

$$\mathbb{C}: N^n \rightarrow N \quad (4)$$

Los operadores de agregación presentan múltiples aplicaciones en diversos dominios (CALVO, et al., 2010). En la toma de decisiones su papel fundamental está en la evaluación y en la construcción de alternativas (TORRA, et al., 2007). Su empleo se enmarca fundamentalmente ESTUDIO DE CASOS

en la toma de decisiones multicriterio, a continuación, se muestran dos de los operadores de agregación más usados y que sirven de base para el operador empleado en la investigación.

- **Media ponderada:** la media ponderada (WA, por sus siglas en inglés) es uno de los operadores de agregación más empleados en la literatura. Se define de la siguiente forma:

Definición 1 Un operador WA tiene asociado un vector de pesos V , con $v_i \in [0,1]$ y $\sum_{i=1}^n v_i = 1$, teniendo la siguiente forma:

$$WA(a_1, \dots, a_n) = \sum_{i=1}^n v_i a_i \quad (5)$$

donde v_i representa la importancia/relevancia de la fuente de datos a_i .

Los operadores cuyo vector de peso dependen de los de los argumentos forman parte de la familia de operadores neat-OWA (FERNÁNDEZ, 2008).

- **El operador OWA dependiente (D-OWA)** (Z. XU, 2006) penaliza con una menor fiabilidad los elementos que se alejen de la media aritmética (μ), disminuyendo la influencia de estos valores en la agregación según la siguiente expresión matemática:

$$D - OWA(a_1, \dots, a_n) = \frac{\sum_{j=1}^n s(a_j, \mu) a_j}{\sum_{j=1}^n s(a_j, \mu)} \quad (6)$$

donde $s(a_j, \mu)$ es una función de similitud entre el elemento a_j y μ (BOONGOEN, y otros, 2008).

Operador propuesto.

El autor propone la creación de medidas de centralidad compuestas a partir del operador D-OWAWA que resuman un grupo de medidas seleccionadas por el decisor. La utilización de este operador permite combinar las ventajas del operador D-OWA y WA en el cálculo de medidas compuestas (MUNDA, y otros, 2003).

Definición 2 Un operador D-OWAWA es una función $\mathbb{R}^n \rightarrow \mathbb{R}$ de dimensión n si tiene un vector de ponderaciones W asociado, con $W = \sum_{j=1}^n w_j = 1$ y $w_j \in [0,1]$ tal que:

$$D - OWAWA(a_1, a_2, a_3 \dots a_n) = \sum_{i=0}^n \hat{v}_i a_i \quad (7)$$

Donde a_i tiene asociado una ponderación \hat{v}_i

$$\hat{v}_i = \beta w_i + (1 - \beta) \omega_i \text{ con } \beta \in [0,1] \quad (8)$$

ω_i corresponde al i-ésimo elemento del vector de pesos calculado a partir del operador D-OWA.

3. ESTUDIO DE CASOS

En la Revista Cubana de Ciencias Informáticas (RCCI), colaboran un total de 582 investigadores de todo el país (los datos se obtuvieron hasta el Volumen 8 Numero 3, 2014 Julio-Septiembre) principalmente de la Universidad de las Ciencias Informáticas (UCI), Universidad Central Martha Abreu de Las Villas (UCLV), Instituto Superior Politécnico José Antonio Echeverría (CUJAE) e instituciones dedicadas a la investigación. La Figura 1 muestra la red donde se identificaron 123 grupos, resaltándose con distintos colores los 4 de mayor tamaño.

Es muy común al realizar análisis en grafos no conexos estudiar el componente principal o subgrafo conexo de mayor cantidad de nodos. Teniendo en cuenta que el componente principal está compuesto por 65 nodos, mostrar gráficos interactivos de éste dificulta la visualización en un pequeño espacio. Partiendo de lo anteriormente descrito se decide utilizar en la investigación el segundo componente que tiene 21 nodos, como muestra la Figura 2, permitiendo mostrar su visualización en el documento al aplicar el análisis.

En esta red existe mucha colaboración entre autores, sin embargo, no es común encontrar la colaboración de dos o más investigadores en más de un artículo, por lo que la relación entre los investigadores está dada por la cantidad de artículos que han publicado en conjunto.

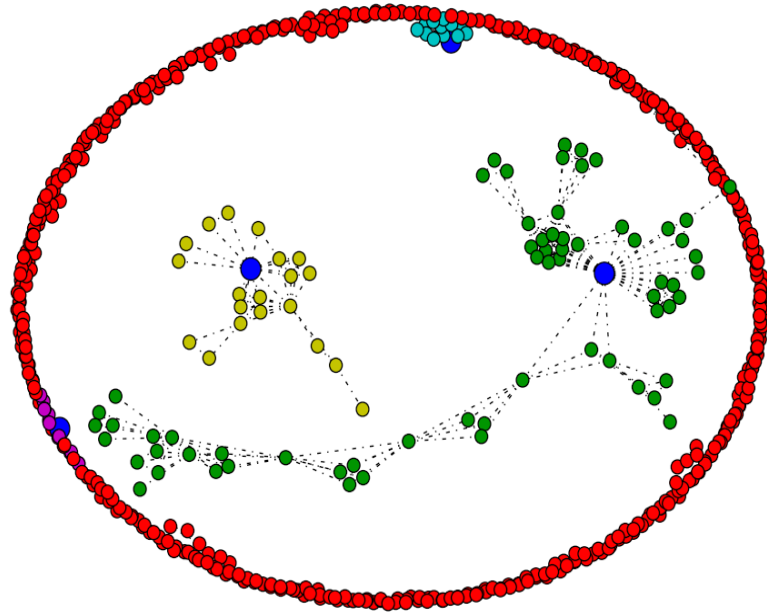


Figura 1: Red de colaboración científica RCCI.

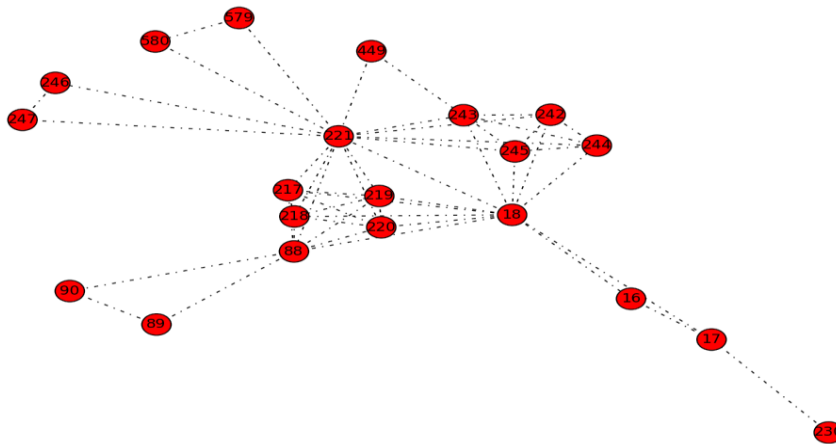


Figura 2: Componente de la red elegido para analizar.

Una vez aplicada el análisis propuesto los resultados se muestran en la Tabla 2, en este caso el análisis estático a partir de las medidas propuestas y la agregación de las mismas con el operador D-OWAWA.

Tabla 2: Centralidad de los investigadores a partir de su grado de cooperación entre ellos.

Investigador	C	C_B	C_C	D-OWAWA
--------------	---	-------	-------	---------

16	0.05	0.0	0.444444444444444	0.13097866077998
17	0.075000000000000	0.099999999999999	0.454545454545454	0.17958063608218
18	0.350000000000000	0.34210526315789	0.71428571428571	0.43814263837234
217	0.150000000000000	0.0	0.57142857142857	0.20760534429599
90	0.05	0.0	0.39215686274509	0.11806591795596
449	0.05	0.0	0.454545454545454	0.13347763347763
579	0.05	0.0	0.454545454545454	0.13347763347763
580	0.05	0.0	0.454545454545454	0.13347763347763
88	0.2	0.189473684210526	0.606060606060606	0.29761829226738
89	0.05	0.0	0.39215686274509	0.11806591795596
218	0.150000000000000	0.0	0.57142857142857	0.20760534429599
219	0.150000000000000	0.0	0.57142857142857	0.20760534429599
220	0.150000000000000	0.0	0.57142857142857	0.20760534429599
221	0.375	0.486842105263157	0.76923076923076	0.52243110292848
230	0.025	0.0	0.3174603174603	0.0891243506728
242	0.125	0.0	0.52631578947368	0.18477043673012
243	0.150000000000000	0.01842105263157	0.54054054054054	0.20540672179348
244	0.125	0.0	0.52631578947368	0.18477043673012
245	0.125	0.0	0.52631578947368	0.18477043673012
246	0.05	0.0	0.454545454545454	0.13347763347763
247	0.05	0.0	0.454545454545454	0.13347763347763

En este caso se evaluaron solo las tres medidas de centralidad propuestas en la investigación:

C : centralidad de grado.

C_B : centralidad por intermediación.

C_C : centralidad por cercanía.

Además, para calcular la medida compuesta D-OWAWA se ajustó el parámetro β con valor 0, por lo que en este caso todas las medidas tienen igual importancia para el decisor.

Como se puede observar en la Tabla 2 el investigador con id 221 es el más importante de la red pues posee los valores más altos de las 3 medidas de centralidad, por ende, la medida compuesta propuesta resulta mayor que la de los demás investigadores. A continuación, se muestra en la Figura 3 la visualización de los nodos teniendo en cuenta su importancia en una escala de colores y tamaños:

221>18>88>217~218~219~220~243>242~244~245>17>449~579~580~246~247>16>90~89>230

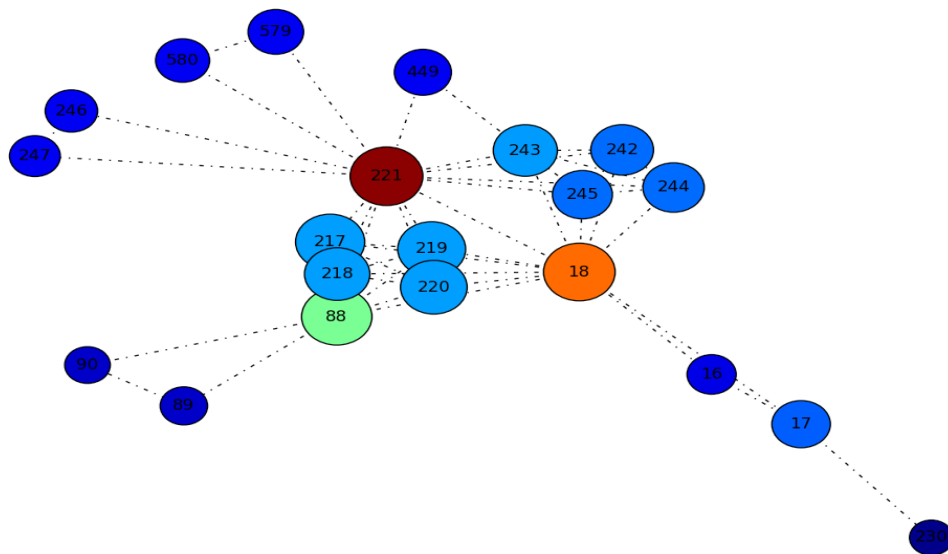


Figura 3: Visualización del componente analizado a partir de la medida compuesta.

Comparación con otro método para el ordenamiento multicriterio:

Dominancia de Pareto en un contexto de maximización.

Dado un vector $u = \{u_1, \dots, u_k\}$, se dice que domina a otro vector $v = \{v_1, \dots, v_k\}$ si y sólo si:

$$\forall i \in \{1, \dots, k\}, u_i \geq v_i \text{ y } \exists i_0 \in \{1, \dots, k\}, u_{i_0} > v_{i_0} \quad (9)$$

Si ordenamos los nodos a partir del teorema de dominancia de vectores obtenemos los siguientes resultados:

221>18>88>17~243~217~218~219~220>222~242~244~245>449~579~580~246~247>16>90~89>230

Sin embargo, ordenando a partir de la medida compuesta D-OWAWA se obtiene como resultado:

221>18>88>243>217~218~219~220~222>242~244~245>17>449~579~580~246~247>16>90~89>230

Si contamos la cantidad de empates obtenidos en el orden por cada uno de los métodos, se muestra como el uso de la medida de centralidad compuesta a partir del operador D-OWAWA permite realizar mejor ordenamiento ya que se obtienen menos empates entre los nodos. (BELLO-LARA, 2010)

Conclusiones

Es posible la utilización de un grafo para representar y modelar la relación entre los autores de artículos científicos. El análisis estático permite la selección de los nodos más importantes de la red. Sin embargo, generalmente se emplea una sola de las medidas de centralidad, dejando a un lado otro número importante de estas medidas.

La utilización del operador D-OWAWA permitió la agregación de distintas medidas de centralidad de un modo flexible. Posibilitando el ordenamiento de los nodos de acuerdo a la medida de centralidad.

Recomendaciones

Aplicar este análisis en otros casos de redes de colaboración, analizar los datos de la revista con los nuevos volúmenes publicados y cotejar los resultados con los expuestos en el trabajo para medir el comportamiento de la red.

Referencias

- AGUIRRE, JULIO LEONIDAS. 2011. Introducción al análisis de redes sociales. Buenos Aires: s.n., Diciembre de 2011. 82.
- ALTAY, Ayca and KAYAKUTLU, Gülgün. 2011. Fuzzy cognitive mapping in factor elimination: A case study for innovative power and risks. s.l.: Procedia Computer Science, 2011. Vol. 3, 1111-1119.
- ARCO-GARCÍA, L. 2008. Agrupamiento basado en la intermediación diferencial y su valoración utilizando la teoría de los conjuntos aproximados " Tesis presentada en opción al grado científico de Doctor en Ciencias Técnicas, Universidad Central de las Villas. 2008.
- BELLO-LARA, RAFAEL. 2010. Técnica para el análisis estático de redes basado en lógica difusa y operadores de agregación "Tesis presentada en opción al grado científico Master en Ciencias, Universidad de las Ciencias Informáticas, La Habana 2010.
- BOONGOEN, Tossapon y SHEN, Qiang. 2008. Clus-DOWA: A new dependent OWA operator. En Fuzzy Systems. [(IEEE World Congress on Computational Intelligence). IEEE International Conference on. IEEE]. s.l.: FUZZ-IEEE, 2008. págs. 1057-1063.

- BORDONS, María y ÁNGELES ZULUETA, M. ^a. 1999. Evaluación de la actividad científica a través de indicadores bibliométricos. *Revista española de cardiología*. 1999. Vol. 52, 10, págs. 790-800.
- BORGATTI, Stephen P. 2005. Centrality and network flow. *Social networks*. 2005. Vol. 27, 1, págs. 55-71.
- BORGATTI, Stephen P., et al. 2009. Network analysis in the social sciences. 2009. Vol. 323, 5916, págs. 892-895.
- CALVO, Tomasa y BELIAKOV, Gleb. 2010. Aggregation functions based on penalties. *Fuzzy sets and Systems*. 2010. Vol. 161, 10, págs. 1420-1436.
- CRIADO, Regino, ROMANCE, Miguel y SÁNCHEZ, Ángel. 2012. Interest point detection in images using complex network analysis. *Journal of Computational and Applied Mathematics*. 2012. Vol. 236, 12, págs. 2975-2980.
- EMIRBAYER J, M. GOODWIN. 1994. Network Analysis, Culture and the Problem of Agency. *The American Journal of Sociology*. Mayo de 1994. Vol. 99, 6, págs. 1411-1454.
- FERNÁNDEZ, J. M. Doña. 2008. Modelado de los procesos de toma de decisión en entornos sociales mediante operadores de agregación OWA. " Tesis en opción al grado de doctor en ciencias, Universidad de Málaga: s.n., 2008.
- GEROLIN, Jerônimo, et al. 2010. Ten-year growth in the scientific production of Brazilian Psychiatry: the impact of the new evaluation policies. *Revista Brasileira de Psiquiatria*. 2010. Vol. 32, 1, págs. 6-10.
- H. JUN, et al. 2010. Evaluating Node Importance with Multi-Criteria, in *Green Computing and Communications (GreenCom), 2010 IEEE/ACM Int'l Conference on & Int'l Conference on Cyber, Physical and Social Computing (CPSCom)*. 2010. págs. 792-797.
- M. LEYVA-VÁZQUEZ, et al., 2012. Modelo para el análisis estático en mapas cognitivos difusos, in *I Congreso Internacional de Ingeniería Informática y Sistemas de Información*. La Habana: s.n., 2012.
- MILGRAM, Stanley. 1967. The small world problem. *s.l.: Psychology today*, 1967. Vol. 2, 1, págs. 60-67.
- MUNDA, Giuseppe y NARDO, Michela. 2003. On the methodological foundations of composite indicators used for ranking countries. *Ispra, Italy: Joint Research Centre of the European Communities*. 2003.
- OBIEDAT, Mamoon, SAMARASINGHE, Sandhya y STRICKERT, Graham. 2011. A new method for identifying the central nodes in fuzzy cognitive maps using consensus centrality measure. 2011.

- PRAT, Anna Maria. 2001. Evaluación de la producción científica como instrumento para el desarrollo de la ciencia y la tecnología. Acimed. 2001. Vol. 9, págs. 111-114.
- SCOTT, John. 1991. Social network analysis: a handbook, 2000. s.l.: 梃, 1991. Vol. 20, pág. 6.
- TORRA, Vicenç y NARUKAWA, Yasuo. 2007. Modeling decisions. Information Fusion and Aggregation Operators. s.l.: Springer, 2007.
- TRAVERS, Jeffrey y MILGRAM, Stanley. 1969. An experimental study of the small world problem. s.l.: Sociometry, 1969. págs. 425-443.
- . 1993. Families of OWA operators. s.l.: Fuzzy sets and systems, 1993. Vol. 59, 2, págs. 125-148.
- Z. XU. 2006. Dependent OWA operators. s.l.: Modeling Decisions for Artificial Intelligence, 2006. págs. 172-178.