

CAI, Congreso Argentino de Agroinformática

Congreso Argentino de Agroinformática: Un análisis bibliométrico

Sandro da Silva Camargo¹, Leonardo Bidese de Pinho¹, and
Yanina Bellini Saibene²

¹ Programa de Pós-Graduação em Computação Aplicada,
Universidade Federal do Pampa & Embrapa Pecuária Sul
Bagé, Rio Grande do Sul, Brasil

² Estación Experimental Agropecuaria Anguil,
Instituto Nacional de Tecnología Agropecuaria (INTA)
Anguil, La Pampa, Argentina
{sandro.camargo, leonardo.pinho}@unipampa.edu.br,
yanina.bellini@inta.gov.ar

Resumen. Para viabilizar el intercambio de competencias y recursos, la colaboración en investigación se ha tornado fundamental para el progreso científico. Este trabajo presenta un abordaje para identificar y analizar la red de colaboración entre instituciones en el ámbito de la investigación en agroinformática en Argentina, a partir de los trabajos publicados en los anales de las tres últimas ediciones del Congreso Argentino de Agroinformática. La metodología se basó en la aplicación de técnicas Análisis de Redes Sociales, que permitieron identificar las instituciones más participativas en el contexto de la Agroinformática en Argentina, enfatizando la relevancia del Instituto Nacional de Tecnología Agropecuaria (INTA) y del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) como catalizadores de la investigación en esta área. Se espera que los resultados aquí presentados puedan aportar elementos para ayudar a los organismos gubernamentales a establecer políticas para fortalecimiento de la investigación y desarrollo de la Agroinformática Argentina.

Palabras Clave: Redes, Ciencia, Instituto Nacional de Tecnología Agropecuaria, Consejo Nacional de Investigaciones Científicas y Técnicas

1 Introducción

La colaboración en investigación se ha convertido en una actividad fundamental para el progreso de la ciencia por mejorar la comunicación entre grupos de investigadores, por permitir el intercambio de experiencias y competencias y por facilitar la producción e disseminación de nuevos conocimientos científicos. Medir el nivel de colaboración entre miembros de una comunidad es una tarea compleja, que se puede hacer de distintas maneras. La forma más evidente de colaboración es la coautoría en publicaciones, que se utiliza como un parámetro confiable para la medición de los niveles de colaboración [10].

A fin de evaluar la colaboración en el área de informática aplicada a la agricultura en Argentina, se delimitó el contexto del Congreso Argentino de Agroinformática (CAI), que es el principal foro científico Argentino sobre este tema. En CAI³ participan investigadores, tecnólogos, desarrolladores y empresas relacionadas al sector agroindustrial presentando trabajos relativos a las TICs aplicadas a problemáticas agropecuaria, agroindustrial y medio ambiental, abarcando desde instancias experimentales a comerciales. El CAI es un evento anual que está en su décima edición, siendo promovido por la Sociedad Argentina de Informática e Investigación Operativa (SADIO). En este trabajo, se analizaron los anales de los tres últimos años del CAI, ocurridos en los años 2014, 2016 y 2017, en Buenos Aires en las dos primeras ediciones mencionadas, y Córdoba en su última edición. No hay anales del CAI de 2015, porque durante el CAI 2014 se decidió probar organizar el evento en forma bianual con el objetivo de contar con una mayor cantidad de contribuciones en cada edición, por lo que esta edición del evento no fue utilizada en este estudio.

En la literatura, se pueden encontrar diversos trabajos de aplicación del abordaje de Análisis de Redes Sociales, o *Social Network Analysis* (SNA), para la identificación de las redes de colaboración en investigación a partir de datos de publicaciones. Camargo y colaboradores [4] construyeron y analizaron la red de colaboración en Informática Agropecuaria en Brasil, a partir del análisis de los 139 trabajos publicados en los Anales del X Congreso Brasileño de Agroinformática, realizado en 2015. Como resultado, se identificaron las principales instituciones de investigación en el área y sus interrelaciones. Camargo y colaboradores [3] analizaron la evolución de la red de colaboración entre instituciones en el ámbito de la investigación en la ciencia del agronegocio, a partir del análisis de los trabajos publicados en los anales del II, III y IV Simposio de la Ciencia del Agronegocio, en los años 2014, 2015 y 2016 respectivamente.

En otras áreas de conocimiento, Yu y Wang [12] construyeron y analizaron una red de colaboración, a partir de más de 10.000 publicaciones anteriores a 2015. El foco de este trabajo fue restringido a dispositivos médicos regenerativos implantables, derivados de animales, basados en tecnología de ingeniería de tejidos. Se analizó el proceso de evolución de la red y el estado actual de la colaboración. Como resultado, el trabajo evidenció la escasa colaboración entre instituciones, el potencial de fortalecimiento de la interacción entre diferentes países, y que las instituciones en los extremos de la red podrían perfeccionar su colaboración y compartir sus descubrimientos con otras instituciones más centrales. Wu y Duan [11] midieron las actividades de colaboración científica en investigación psiquiátrica a nivel de autores, instituciones y países. En este trabajo se consultó más de 36.000 trabajos. Como resultado, además de la descripción del comportamiento colaborativo en el área de psiquiatría, también se sugirió que la colaboración puede ofrecer evidencias científicas y sugerencias razonables para basar la creación de políticas para guiar el financiamiento de la investigación psiquiátrica en el futuro. Andrade y colaboradores [1] discutieron cómo los procesos de colaboración científica pueden ser identificados y caracterizados

³ <http://47jaiio.sadio.org.ar/index.php?q=cai>

a través del abordaje de Análisis de Redes Sociales y redes complejas. También se discute que la colaboración en comunidades científicas ha sido estudiada como uno de los elementos básicos para explicar la difusión del conocimiento.

Por fin, Cechinel [5] hizo un análisis de la evolución de la colaboración científica entre países durante los primeros siete años de la Conferencia Latinoamericana de Objetos y Tecnologías de Aprendizaje (LACLO). El autor sugirió que los resultados podrían ayudar a la comunidad a establecer políticas enfocadas en la integración de países que aún no estaban bien conectados dentro de la comunidad.

Inspirado por el análisis de los resultados de trabajos relacionados, el presente estudio busca identificar y presentar la red de colaboración entre las instituciones de investigación argentinas en la área de la informática agropecuaria, tomando como base los trabajos publicados en los anales del Congreso Argentino de Agroinformática (CAI). El estudio se propone presentar, a partir de evidencias de coautorías de trabajos, que instituciones desempeñan un papel destacado en el área. Como impactos esperados, se pretende que la red de colaboración aquí identificada contribuya con la creación de políticas gubernamentales de fomento a la investigación y desarrollo en informática agropecuaria, que la propia comunidad pueda conocerse mejor a partir de la identificación de sus miembros y colaboraciones, y que otras instituciones interesadas en comenzar a desarrollar investigaciones en esta área puedan, a partir del conocimiento de quién son los miembros clave, buscar colaboraciones para ingreso en la red de investigación.

El resto de este trabajo está organizado de la siguiente manera: La Sección 2 presenta una breve descripción de las características de la base de datos utilizada para este estudio y el enfoque de Análisis de Redes Sociales. La Sección 3 presenta los resultados obtenidos a partir del análisis de los datos y las redes de colaboración identificadas. La Sección 4 presenta un resumen de los descubrimientos, las restricciones del enfoque utilizado y las perspectivas de trabajos futuros.

2 Materiales y Métodos

El presente trabajo utilizó como base los trabajos publicados en los Anales del 6º, 8º e 9º Congreso Argentino de Agroinformática, que están disponibles en los sitios de los eventos⁴. Un trabajo con autores de dos o más instituciones diferentes fue considerado como producto de una colaboración, de acuerdo con el enfoque utilizado en otros trabajos [3, 4, 7]. Si se tiene un trabajo cualquiera con n instituciones, donde $n \geq 2$, para el enfoque de Redes Sociales fueron consideradas como colaboración todas las combinaciones de las n instituciones, combinadas 2 a 2, totalizando $C_{n,2}$ colaboraciones.

Para la visualización interactiva de datos georeferenciados se utilizó la API (*Application Programming Interface*) Google Charts⁵, que es un servicio gra-

⁴ 6ºCAI: <http://43jaiio.sadio.org.ar/proceedings/CAI/search.html>,

8ºCAI: <http://45jaiio.sadio.org.ar/node/86>,

9ºCAI: <http://www.clei2017-46jaiio.sadio.org.ar/node/86>

⁵ <https://developers.google.com/chart/interactive/docs/>

tuito ofrecido por Google [13]. Para la representación de las colaboraciones en forma de una red social, se utilizó la herramienta Gephi 0.9.1 ⁶, que es una herramienta gratuita y de código abierto para la creación, análisis y explotación de redes complejas [2].

La aplicabilidad de las Redes Sociales para análisis de colaboración en una red de investigación se debe al hecho de que son una metáfora ampliamente utilizada para representar las relaciones entre miembros de una comunidad. El enfoque de análisis de redes sociales es un conjunto de técnicas enfocadas en el estudio de una estructura social, considerando primariamente los datos de las relaciones y el contexto social de los miembros, en detrimento de las propiedades de los propios miembros [9].

Desde el punto de vista de la análisis de redes sociales, una red de colaboración puede representarse en la forma de un grafo $G(N, A)$, donde N es el conjunto de nodos que aquí representan las instituciones de enseñanza o investigación, y A es el conjunto de aristas, teniendo que cada arista a_z es un subconjunto de N con sólo dos elementos [8]. El tamaño del nodo es directamente proporcional al número de trabajos en que algún filiado de la institución fue coautor. Las aristas del grafo se definieron por $A(n_x, n_y)$ donde n_x y n_y son instituciones colaboradoras en un trabajo publicado. Para cada vez que un par de instituciones colaboradores se repite en otros trabajos, el grosor de esta arista se incrementa en el grafo.

El análisis de redes sociales utiliza un conjunto de métricas para describir las características de los grafos [6]. Para este trabajo, se analizaron las siguientes métricas: Cantidad de nodos, Cantidad de aristas, Grado promedio, Grado promedio ponderado, densidad y componentes conectados. En relación a las métricas de los nodos, se utilizaron tamaño, grado, intermediación y *pagerank*.

El grado de un nodo indica la cantidad de aristas que inciden sobre este nodo, o sea, en cuantas colaboraciones esta institución participó. El grado promedio es el promedio de los grados de todos los nodos del grafo. Esta métrica se calcula por la siguiente fórmula:

$$GP = \frac{2|A|}{|N|}$$

donde GP es el Grado Promedio, $|A|$ es la cantidad de Aristas, o colaboraciones, y $|N|$ es la cantidad de Nodos, o instituciones. Cuando dos instituciones colaboran en varios trabajos, el peso de su arista es la cantidad de colaboraciones. Mientras el grado promedio considera a todas las aristas con peso 1, el grado promedio ponderado considera el peso de cada arista. La cantidad de componentes conectados representa cuántos grafos disjuntos hay en la red. Cuanto menor sea la cantidad de componentes, más conectados son los nodos.

La densidad de un grafo indica cuánto su número de aristas está cerca del número máximo de aristas posibles para este grafo. La densidad mínima es 0 para grafos desconexos y 1 para grafos completos. La métrica de densidad se calcula por la siguiente fórmula:

⁶ <https://gephi.org/>

$$D = \frac{2|A|}{N(N-1)}$$

La métrica de intermediación de los nodos, o *betweenness centrality*, es una métrica que cuantifica la frecuencia o el número de veces que un nodo actúa como un puente a lo largo del camino más corto entre otros dos nodos. Así, nodos con mayor intermediación, reconocidamente, tienen un papel fundamental en la estructura de la red, principalmente en el proceso de difusión de información.

La métrica *PageRank* es una abstracción desarrollado por Google para inferir la reputación de una página web y aumentar su prioridad de posición en el resultado de las búsquedas. La aplicación del algoritmo *pagerank* en la red de colaboración indicaría con cuantas más instituciones una institución participa, mayor sería su importancia. Además, cuanto mayor es la reputación de una institución, proporcionalmente mayor será la importancia de sus colaboraciones, así como son premiadas las colaboraciones desde diferentes comunidades poco relacionadas entre sí.

3 Resultados y Discusión

Se realizó un análisis sobre las ciudades de los autores principales de los trabajos, a fin de buscar evidencias georeferenciadas sobre la investigación en el área de Agrolinformática. En este sentido, la Figura 1 permite identificar una gran concentración en el desarrollo de trabajos publicados en las provincias de La Pampa, Buenos Aires y Córdoba de forma consistente durante los tres años analizados. El resto de la región Pampeña, Cuyo y Noreste Argentino (NAE) aparecen con más fuerza en el 8° y 9° CAI. No hubo trabajos producidos por autores principales de la región Patagonia Sur de Argentina. De acuerdo con la Tabla 1, las ciudades con mayor contribución en la producción de trabajos para las tres últimas ediciones del CAI fueron: Córdoba (14 trabajos), Buenos Aires (10) y Anguil (7). Además de los datos presentados en esta tabla, también hubo 9 ciudades con dos trabajos publicados y 19 ciudades con un solo trabajo. Se verifica también que en el 8° y 9° CAI, la ciudad con más artículos publicados es la ciudad anfitriona del evento.

Las Figuras 2a, 2b y 3 muestran las redes de colaboración relativas a los anales del 6°, 8° y 9° CAI, enfatizando la última edición. Se puede evidenciar que el nivel de colaboración entre las instituciones tiene una tendencia creciente. Cuando el peso de las aristas era uno, el valor fue omitido. Se mostraron los pesos de aristas con un valor superior a uno. En relación a los tamaños de los nodos, los menores representan instituciones con un único trabajo publicado. Los nodos mayores representan instituciones con más trabajos publicados. Los tamaños máximos se presentan en la figura 4a. Los colores de los nodos son influenciados por la métrica de intermediación, presentada en la figura 4c. Cuanto menor el valor de la métrica de intermediación, más cian es el color del nodo. Cuanto más rojo es el color, mayor es la intermediación. Los valores medios se representan por el color blanco para los nodos.

En la Tabla 2 se presentan las métricas de las redes de colaboración generadas a partir de las publicaciones en las tres últimas ediciones del CAI. La cantidad

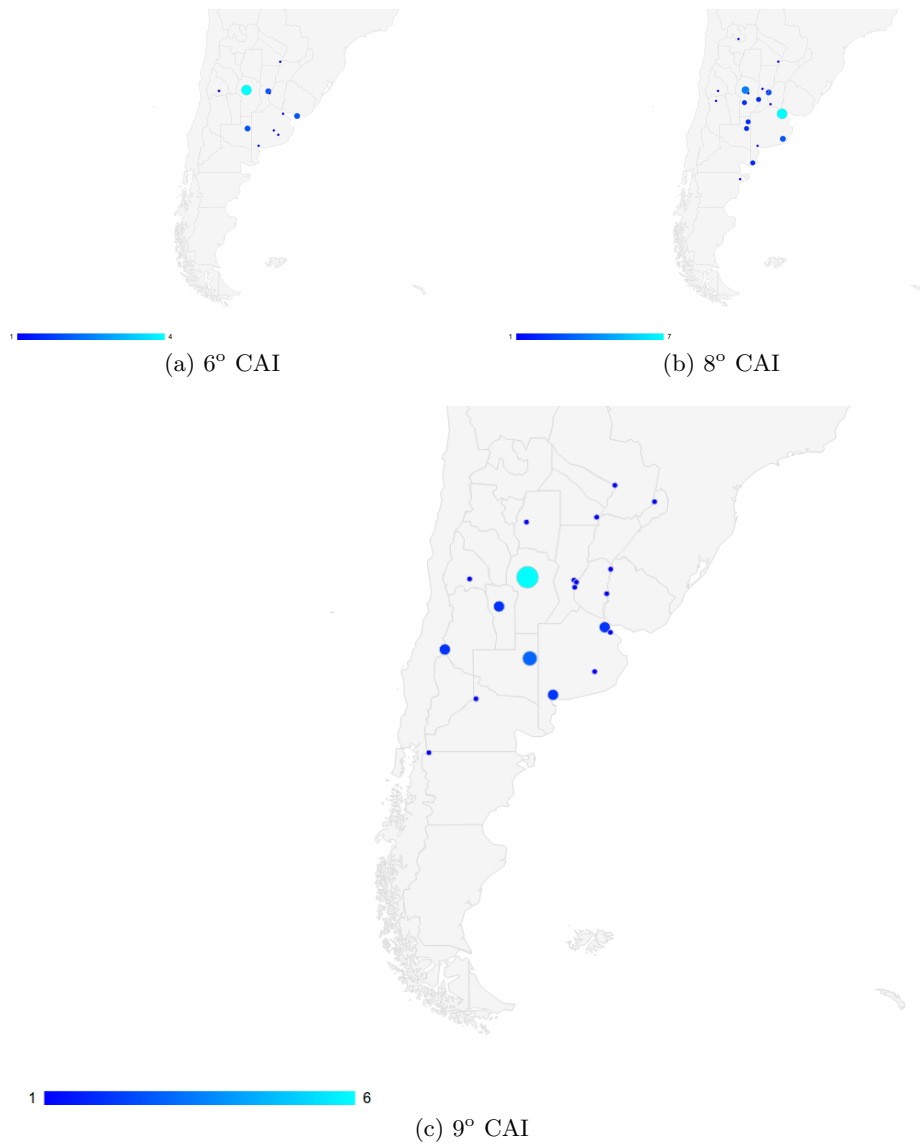


Fig. 1: Ciudades de las instituciones de los autores principales de los trabajos publicados

de nodos muestra que la octava edición fue la de mayor cantidad de instituciones participantes, consideradas las filiaciones de los autores. También queda claro el incremento del tamaño del evento de la sexta para la octava edición, lo cual puede obedecer a la decisión de los organizadores de no realizar un llamado a

Tabla 1: Ciudades con mayor cantidad de autores principales de los trabajos publicados en el CAI

Ciudad	6° CAI	8° CAI	9° CAI	Total
Córdoba	4	4	6*	14
Buenos Aires	1*	7*	2	10
Anguil	2	2	3	7
Bahía Blanca	1	1	2	4
Paraná	0	3	1	4
Santa Fé	2	1	1	4
San Juan	1	1	1	3
Balcarce	0	3	0	3

* Ciudad anfitriona del evento

presentación de trabajos durante el 2015. Además, la cantidad de aristas representa la cantidad de colaboraciones entre las instituciones. De la misma forma, la octava edición fue la que presentó mayor cantidad de colaboraciones entre instituciones. Los grados de los nodos muestran una menor cantidad de colaboraciones, o aristas, en el 6° CAI, y un aumento de las colaboraciones en las ediciones siguientes, con énfasis especial en el 8° CAI.

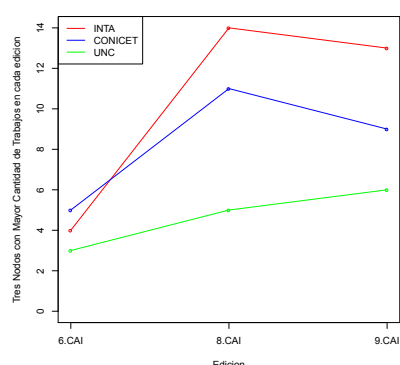
Tabla 2: Metricas de las redes de colaboración de cada año

Metrica	6° CAI	8° CAI	9° CAI
Cantidad de Nodos	19	33	28
Cantidad de Aristas	16	50	38
Grado Promedio	1,684	3,03	2,714
Grado Promedio Ponderado	1,789	3,758	3,143
Diametro da Red	4	2	3
Densidad del Grafo	0,094	0,095	0,101
Modularidad	0,448	0,298	0,361
Componentes Conectados	7	10	4
Tamaño Promedio del Camino	2,373	1,688	2,161
Trabajos publicados	18	37	31

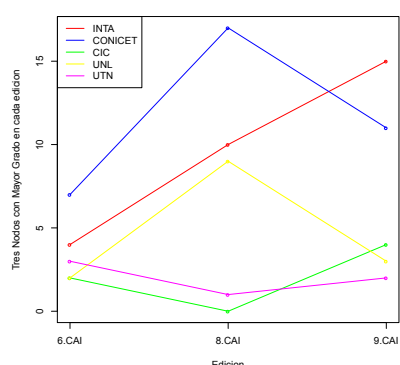
La densidad de los grafos de las tres ediciones muestra una característica muy positiva de un leve incremento de las colaboraciones. Sin embargo, el valor cercano a cero, indica que todavía hay un potencial para aumentar estas colaboraciones.

La menor cantidad de componentes de la red del 9° CAI, muestra la existencia de sólo cuatro subgrafos. El primer subgrafo es compuesto únicamente por la Universidad Paraguaya-Alemana (UPA) del Paraguay, que no tuvo colabora-

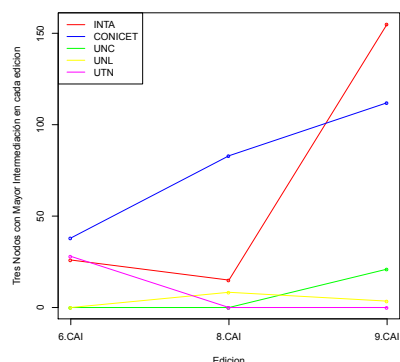
También se analizaron las métricas de los nodos, a fin de identificar las instituciones más relevantes en el contexto de la investigación en agroinformática en Argentina. Con el fin de simplificar la visualización de los resultados, se seleccionaron solamente las tres instituciones con mayor valor para cada métrica, a cada año. La Figura 4 presenta los resultados de las métricas de nodos: cantidad de trabajos, grado, intermediación y *pagerank*.



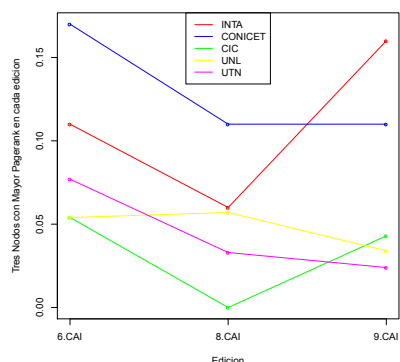
(a) Cantidad de trabajos



(b) Cantidad de colaboraciones



(c) Intermediación de los nodos



(d) Págerank de los nodos

Fig. 4: Métricas de los nodos en las últimas tres ediciones del CAI

Conforme presentado en la figura 4a, las mismas tres instituciones (INTA, CONICET e UNC) fueron las que tuvieron mayor cantidad de publicaciones en las tres ediciones del evento. El INTA sobresale como la mayor cantidad de

publicaciones en las dos últimas ediciones consolidándose como la institución más productiva en el contexto reciente del CAI.

En relación a los grados de los nodos, que representan la cantidad de colaboraciones, INTA y CONICET fueron las dos instituciones con más trabajos en colaboración en las tres últimas ediciones del evento. Además, la cantidad de colaboraciones del INTA presenta una tendencia linealmente creciente. La tercera institución con mayor cantidad de colaboraciones fue diferente en cada edición del evento. En 6° CAI, esta posición fue ocupada por la Universidad Tecnológica Nacional (UTN). En 8° CAI, por la Universidad Nacional del Litoral (UNL). En 9° CAI, por la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC).

En relación al *pagerank*, INTA y CONICET fueron las dos instituciones con mayor valor para esta métrica en las tres ediciones del evento. Sin embargo, el INTA presentó un gran incremento de *pagerank* en la última edición del evento. Por otro lado, el CONICET tuvo una reducción a partir del 6° CAI.

En el 6° CAI, la UTN fue la tercera institución con mayor *pagerank*. Sin embargo, esta métrica de UTN viene disminuyendo en los eventos siguientes. En el 8° CAI, la UNL obtuvo el tercer mayor *pagerank*, aliado a la quinta posición en la edición siguiente, lo que muestra cierta estabilidad en sus contribuciones en los dos últimos años. En el 9° CAI, el CIC fue la tercera institución con mayor *pagerank*, con un valor muy similar al 6° CAI, pero sin un valor relevante en el 8° CAI.

4 Conclusiones y Trabajos Futuros

Este trabajo presentó un panorama de la colaboración científica en agroinformática en Argentina a partir de la identificación de las redes de colaboración científica entre las instituciones que tuvieron artículos publicados en los anales del 6°, 8° y 9° Congreso Argentino de Agroinformática (CAI), ocurridos en los años 2014, 2016 y 2017.

Las análisis realizadas en este trabajo muestran las evidencias del constante crecimiento de la importancia del Instituto Nacional de Tecnología Agropecuaria (INTA) en el contexto de la agroinformática en Argentina. En este mismo entorno, el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) también viene presentando un papel muy representativo. El análisis geográfico de la red de colaboradores presenta la necesidad de analizar aquellas áreas del país que no están encontrando en el CAI un espacio de publicación de sus trabajos como la Patagonia y parte del Noroeste Argentino (NOA); pudiendo convertirse en una posibilidad de crecimiento del congreso.

Como restricción del abordaje aplicado, se destaca que parte del trabajo de identificación de las instituciones involucradas en los artículos fue realizada a través de un proceso no automatizado. Como trabajo futuro, se pretende utilizar la biblioteca OpenNLP para superar esta restricción.

Cuanto a los países vecinos, hay un claro potencial de internacionalización con la expansión del CAI. Los trabajos de instituciones extranjeras se restringieron a

contribuciones puntuales, en ediciones específicas, provenientes de Chile, Uruguay y Paraguay, además de colaboraciones de instituciones argentinas con Alemania y Estados Unidos. La participación de miembros del comité de otros países puede ayudar a contribuir con esta internacionalización y también podría contribuir para una mejor difusión de las llamadas de trabajo.

References

1. Andrade, M.T.T., Braga, P., Carneiro, T.K.G., Ribeiro, N.M., Moret, M.A., de Barros Pereira, H.B.: Contextualized analysis of social networks: Collaboration in scientific communities. *Social Networking* 3(2) (Febrero 2014)
2. Bastian, M., Heymann, S., Jacomy, M.: Gephi: an open source software for exploring and manipulating networks. In: *International AAAI Conference on Weblogs and Social Media*. vol. 8, pp. 361–362. The AAAI Press, Menlo Park, California (2009)
3. Camargo, S.S., Fiss, R.E., Camargo, F.N.P., Lampert, V.N.: Colaboração científica: uma análise evolutiva de redes sociais no simpósio da ciência do agronegócio. In: *Anais do V Simpósio da Ciência do Agronegócio*. Porto Alegre, Brasil (2017)
4. Camargo, S.S., Fiss, R.E., Camargo, F.N.P., Volk, L.B.S., Trindade, J.P.P.: Um panorama da colaboração científica na pesquisa agroinformática brasileira. In: *Anais do XI Simpósio Brasileiro de Agroinformática (SBIAGro)*. pp. 527–536. Campinas, Brasil (2017)
5. Cechinel, C.: Scientific collaboration between countries in laclo from a social network analysis perspective. In: *Oitava Conferência Latino-Americana de Objetos e Tecnologias de Aprendizagem. LACLO 2013*, Valdivia, Chile (2013)
6. Diestel, R.: *Graph Theory*. No. 173 in *Graduate Texts in Mathematics*, Springer (1997)
7. Newman, M.E.J.: Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National academy of Sciences of the United States of America* 101(suppl 1), 5200–5205 (2004)
8. Newman, M.E.J.: *Networks: An Introduction*. Oxford University Press, New York, NY, USA (2010)
9. Otte, E., Rousseau, R.: Social network analysis: a powerful strategy, also for the information sciences. *Journal of information Science* 28, 441–453 (2002)
10. Su, Y.F.: *Study on the Cooperation Network of Spory Research Papers*. Shanghai University of Sport, Shanghai (2010)
11. Wu, Y., Duan, Z.: Social network analysis of international scientific collaboration on psychiatry research. *Int J Ment Health Syst* 9(2) (Enero 2015)
12. Yu, S.Y., Wang, H.M.: Scientific collaboration: a social network analysis based on literature of animal-derived regenerative implantable medical devices. *Regen Biomater* 3(3), 197–203 (Mayo 2016)
13. Zhu, Y.: Introducing google chart tools and google maps api in data visualization courses. *IEEE Computer Graphics and Applications* 32(6), 6–9 (Noviembre 2012)