

Diez Años del Congreso Argentino de Agroinformática: Un Análisis Histórico del Alcance Geográfico y Redes de Colaboración

Sandro da Silva Camargo¹, Leonardo Bidese de Pinho¹,
Marcelo Horacio Bosch², Claudio Machado³, and Yanina Bellini Saibene⁴

¹ Programa de Pós-Graduação em Computação Aplicada (PPGCAP),
Universidade Federal do Pampa (UNIPAMPA) & EMBRAPA Pecuária Sul
Bagé, Rio Grande do Sul, Brasil

² Coordinación Nacional de Investigación y Desarrollo,
Instituto Nacional de Tecnología Agropecuaria (INTA)
Buenos Aires, Argentina

³ CIVETAN (Facultad de Ciencias Veterinarias, UNCPBA Argentina, CICPBA,
CONICET)

Tandil, Buenos Aires, Argentina

⁴ Estación Experimental Agropecuaria (EEA) Anguil,
Instituto Nacional de Tecnología Agropecuaria (INTA)
Anguil, La Pampa, Argentina

{sandro.camargo, leonardo.pinho}@unipampa.edu.br,
{bosch.marcelo, bellini.yanina}@inta.gov.ar,
cmachado@vet.unicen.edu.ar

Resumen. Para permitir el intercambio de competencias y recursos, la colaboración en investigación se ha tornado fundamental para el progreso científico. Este trabajo relata un proceso 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 diez ediciones del Congreso Argentino de Agroinformática, así como reconocer el alcance geográfico del evento. La metodología se basó en la aplicación de técnicas de Análisis de Redes Sociales, que permitieron identificar las instituciones más participativas en el contexto de la Agroinformática en Argentina. Fueron analizados 270 trabajos con 1112 autorías de 702 autores distintos. Los resultados enfatizan 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 aporten elementos para ayudar a los organismos gubernamentales a establecer políticas para fortalecimiento de la investigación y desarrollo de la Agroinformática Argentina.

Palabras Clave: Colaboración Científica, Asociación Científica, Colaboración Tecnológica, Asociación Tecnológica.

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 y diseminación de nuevos conocimientos científicos [5]. La forma más evidente de colaboración es la coautoría en publicaciones [13], que son un poderoso instrumento para el análisis de colaboraciones y asociaciones científicas y tecnológicas, haciendo posible obtener la comprensión de los patrones de cooperación entre individuos y organizaciones [12].

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 CAI⁵ participan investigadores, tecnólogos, desarrolladores y empresas relacionadas al sector agroindustrial, de la Argentina y más recientemente de países fronterizos, presentando trabajos relativos a las Tecnologías de la Información y la Comunicación (TIC) aplicadas a problemáticas agropecuaria, agroindustrial y medio ambiental, abarcando desde instancias experimentales a comerciales. El CAI es un evento anual que tuvo su décima edición en 2018, siendo promovido por la Sociedad Argentina de Informática y Investigación Operativa (SADIO). En este trabajo, se analizaron los anales de todas las ediciones del CAI, ocurridas en las siguientes ciudades: Santa Fé (2008), Mar del Plata (2009), Buenos Aires (2010, 2014, 2016 y 2018), Córdoba (2011, 2013 y 2017) y La Plata (2012). No hay anales del CAI de 2012 y 2015, porque durante los años de 2011 y 2014 se decidió probar la organización del evento en forma bianual con el objetivo de contar con una mayor cantidad de contribuciones en cada edición, por lo que estas ediciones del evento no fueron utilizadas en este estudio. Cabe aclarar que en el año 2012 se realizó el congreso pero sin llamado a presentación de trabajos; el evento consistió sólo en charlas técnicas.

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, Pinho y Bellini Saibene [5] construyeron y analizaron la red de colaboración entre instituciones en Agroinformática en Argentina, a partir del análisis de los 86 trabajos publicados en los Anales del 6º, 8º y 9º CAI, llevados a cabo en los años de 2014, 2016 y 2017. 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, así como 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

⁵ <http://47jaiio.sadio.org.ar/index.php?q=cai>

II, III y IV Simposio de la Ciencia del Agronegocio, realizados en los años 2014, 2015 y 2016 respectivamente.

En otras áreas de conocimiento, Yu y Wang [16] 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 [15] 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 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 [6] 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, y también de los países fronterizos, en el área de la informática agropecuaria, tomando como base todos los trabajos publicados en los anales de las diez ediciones del CAI. El estudio propone, a partir de evidencias de coautorías de trabajos, qué 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 más efectivas 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 quienes son los miembros clave, buscar colaboraciones para ingresar 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 discute los resultados obtenidos a partir del análisis de los datos y las redes de

colaboración identificadas. La Sección 4 expone un resumen de las conclusiones, las restricciones del enfoque utilizado y las perspectivas de trabajos futuros.

2 Materiales y Métodos

2.1 Fuente de Datos

El presente trabajo utilizó como base los trabajos publicados en los Anales de todas ediciones del CAI y también los datos informados por los autores en el sistema de envío de trabajos. La base de Datos contenía los siguientes campos: Nombre de la Persona, Rol (típicamente autor), ID de la Persona en el Paper, Tipo de Presentación (Exposición Oral, Full Paper, Extended Abstract, Poster o Comunicación Oral), Año de la Edición, Id del Paper, Género de la Persona (Hombre o Mujer), Institución, Dependencia de la Institución, Ciudad, Provincia y País. Además, los anales de la 3^a hasta la 10^a ediciones están disponibles en los sitios de los eventos⁶. Los anales de las demás ediciones fueron obtenidos a partir de los respectivos CD.

2.2 Visualización Georreferenciada

Para la visualización interactiva de datos georreferenciados se utilizó la API (*Application Programming Interface*) Google Charts⁷, que es un servicio gratuito ofrecido por Google [17]. Para esta visualización, fueron utilizados los campos Ciudad, Provincia y País.

2.3 Redes de Colaboración

Para la representación de las colaboraciones en forma de una red social, se utilizó la herramienta Gephi 0.9.1⁸, 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 las redes son una metáfora ampliamente utilizada para representar los miembros de una comunidad y sus ligaciones. 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

⁶ 3° CAI: <http://39jaiio.sadio.org.ar/node/86>,

4° CAI: <http://40jaiio.sadio.org.ar/node/86.htm>,

5° CAI: <http://42jaiio.sadio.org.ar/proceedings/simposios/CAIContribuciones.htm>,

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

7° CAI: no disponibles,

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

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

10° CAI: <http://47jaiio.sadio.org.ar/cai>

⁷ <https://developers.google.com/chart/interactive/docs/>

⁸ <https://gephi.org/>

relaciones y el contexto social de los miembros, en detrimento de las propiedades de los propios miembros [11].

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 [3–5, 9]. 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, utilizando una combinación simple de dos elementos, totalizando $C_{n,2}$ colaboraciones.

Desde el punto de vista del 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 [10]. El tamaño del nodo es directamente proporcional al número de trabajos en que algún afiliado 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. Foulds [8] presenta algunos ejemplos de aplicaciones de la teoría de los grafos.

2.4 Métricas de Redes

El análisis de redes sociales utiliza un conjunto de métricas para describir las características de los grafos [7]. Para este trabajo, se analizaron las siguientes métricas: cantidad de nodos, cantidad de aristas, grado promedio, grado promedio ponderado, modularidad, diámetro, 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 ecuación:

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

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 en 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 define por la siguiente ecuación:

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

La métrica de modularidad es una medida de la estructura de la red, midiendo la fuerza de división de la red en comunidades. Redes con alta modularidad tienen colaboraciones densas entre las instituciones de diferentes comunidades. Por otro lado, las redes con baja modularidad tienen pocas colaboraciones entre instituciones de comunidades diferentes. El valor de la modularidad debe estar en el rango continuo entre 0, para redes donde no hay ninguna colaboración, y 1, para redes donde todas las instituciones tienen colaboración con todas las demás.

La métrica del diámetro indica el camino más corto entre dos nodos más distantes en una red. La distancia entre dos nodos es representada por la cantidad de nodos existentes entre ellos. Así, el diámetro representa el tamaño lineal de una red. La densidad de una red es definido como la proporción de la cantidad de aristas existentes por la cantidad de aristas posibles. Proporciones cercanas a cero indican que hay pocas colaboraciones. Proporciones cercanas a uno indican que las instituciones colaboran con todas las demás.

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 de intermediación se calcula por la siguiente ecuación:

$$i(n) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(n)}{\sigma_{st}}$$

donde σ_{st} es el número total de caminos más cortos del nodo s hasta el nodo t , y $\sigma_{st}(n)$ es el número de estos caminos que pasan por n .

La métrica *PageRank* es una abordaje 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 las instituciones que tienen un rol de liderazgo en la red [14].

3 Resultados y Discusión

3.1 Visualización Georreferenciada

Se realizó un análisis sobre los países de las instituciones de todos los autores de cada trabajo, a fin de visualizar el alcance geográfico del evento. La Figura 1 muestra los países de las instituciones que ya publicaron trabajos en el CAI. La cantidad de autorías de cada país se muestra en la Tabla 1. Se puede ver que instituciones de diversos países de América del Sur ya publicaron en el CAI y, eventualmente, algunas instituciones de otros continentes. En la última edición, hay un porcentaje importante de trabajos de Brasil y Estados Unidos.



Fig. 1: Países de las instituciones de los autores de los trabajos publicados.

Como cerca de 87% de las autorías procede de la Argentina, se realizó un análisis específico sobre las provincias y ciudades de las instituciones de los autores de este país. La Tabla 2 presenta las cantidades de autorías por provincia. Así se identifica una gran concentración en el desarrollo de trabajos publicados con autores de las provincias de Buenos Aires, Córdoba, Santa Fé y La Pampa, que tuvieron trabajos publicados en todas las ediciones del evento. La Figura 3a presenta un *boxplot* de la distribución, por provincia, de los autores de los trabajos. Bajo un análisis estadístico, la cantidad de autores de estas provincias son consideradas como valores atípicos superiores, o sea, numéricamente distante del resto de los datos.

Otras 14 provincias tuvieron trabajos publicados en algunas ediciones del CAI, como puede ser analizado en la Tabla 2. De los 971 autores de instituciones argentinas, 21 no declararon su provincia en el momento del envío del trabajo. Las provincias de Formosa, La Rioja, Salta, Santa Cruz y Tierra del Fuego no tuvieron trabajos publicados en la historia del CAI.

Bajo una perspectiva más detallada, se analizaron las ciudades de los autores, según el resultado presentado en la Figura 2. Se encontraron autores en 57 ciudades argentinas distintas. La Figura 3b presenta un *boxplot* de la distribución de las ciudades de las instituciones de los autores de los trabajos. Las ciudades cuya frecuencia de autorías son consideradas *outliers* son: Córdoba (145 autorías), Santa Fé (118), Anguil (95), Buenos Aires (71), Castelar(59) y San Juan (49).

Tabla 1: Cantidad de autorías por país para todos los trabajos publicados en el CAI.

País	2008	2009	2010	2011	2013	2014	2016	2017	2018	Total
	1°	2°	3°	4°	5°	6°	8°	9°	10°	País
Argentina	135	105	117	116	57	53	141	112	135	971
Brasil			1						47	48
Estados Unidos	1						3		22	26
Chile								8	13	21
Uruguay			5			6		1	1	13
España			2	5				1		8
Venezuela				1				3		4
Paraguay								3		3
Colombia						2				2
Costa Rica				2						2
Alemania								1		1
<i>No declarado</i>				1			9	3		13
Total Edición	136	105	125	125	57	61	153	132	218	1112

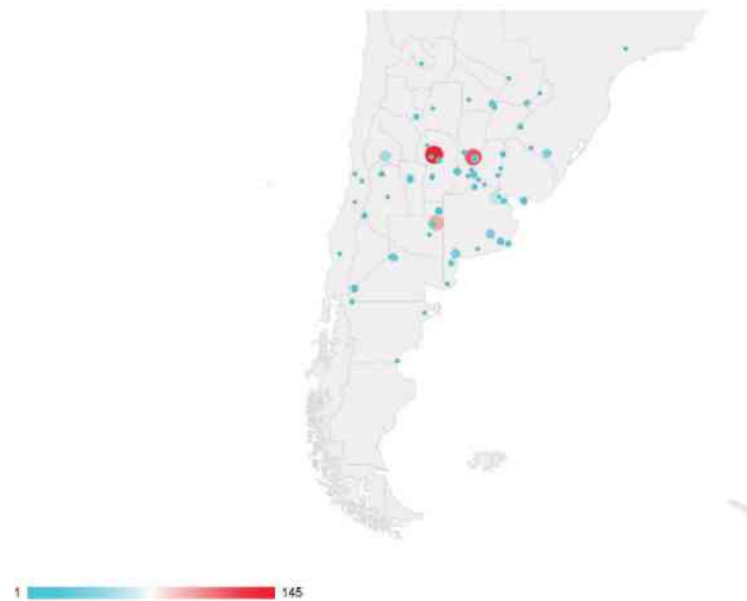


Fig. 2: Ciudades de las instituciones de los autores de los trabajos publicados con énfasis en el mapa de Argentina.

3.2 Redes de Colaboración

Las Figuras 4, 5 y 6 muestran las redes de colaboración relativas a los anales de un primer y segundo período de cuatro años cada uno y del 10° CAI, enfatizando

Tabla 2: Cantidad de autorías por provincia argentina para todos los trabajos publicados en cada edición del CAI.

Año	2008	2009	2010	2011	2013	2014	2016	2017	2018	Total
Provincia / Edición	1°	2°	3°	4°	5°	6°	8°	9°	10°	Provincia
Buenos Aires	36	41*	45*	29	6	17*	39*	27	18*	258
Córdoba	2	3	17	45*	19*	15	40	21*	19	181
Santa Fé	9*	32	21	11	11	10	17	11	30	152
La Pampa	38	7	8	15	9	3	13	12	26	131
San Juan	22	7		5		2	4		8	48
Río Negro	16		13	5			8	4		46
Entre Ríos		6	6		1	4	5	5	7	34
Chaco	2	2	7	2	4		3	5		25
San Luis		6						9	10	25
Misiones					2			5	9	16
Mendoza					1		3	4		8
Catamarca	3			1	3					7
Corrientes	1	1				2			1	5
Chubut	1						2	1		4
Jujuy				1			2			3
Neuquén	1							2		3
Santiago del Estero	2							1		3
Tucumán								1		1
<i>No Declarado</i>	2			2	1		5	4	7	21
Total Edición	135	105	117	116	57	53	141	112	135	971

* Provincia anfitriona del evento

la última edición. En los grafos presentados, los nodos representan las instituciones de los autores y las aristas representan las colaboraciones entre autores de diferentes instituciones. Cuando el peso de las aristas era uno, el valor fue omitido. Se mostraron los pesos de las aristas con un valor superior a uno. El tamaño de cada nodo es proporcional a cantidad de trabajos publicados. Los nodos con circunferencias menores representan instituciones con un único trabajo publicado. Los tamaños máximos se presentan en la Figura 7a. Los colores de los nodos son influenciados por la métrica de intermediación, presentada en la Figura 7c. 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 con el color blanco.

En el primer CAI (2008 - Santa Fé), presentado en la Figura 4a, ya es posible verificar la importancia del Instituto Nacional de Tecnología Agropecuaria (INTA), en términos de medida de centralidad. Todos los demás nodos tienen una centralidad muy baja. Cuatro instituciones no tuvieron ninguna colaboración. Acerca del tamaño de los nodos, INTA tuvo participación en 15 trabajos publicados, Universidad Nacional San Juan (UNSJ) en 5, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y Universidad de Buenos

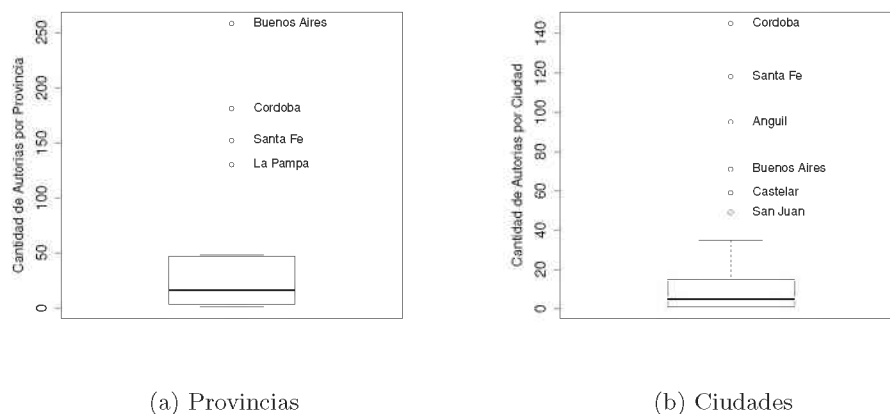


Fig. 3: Provincias y Ciudades Argentinas de las instituciones de los autores de los trabajos publicados.

Aires (UBA) en 3. Fueron identificados dos trabajos en colaboración entre INTA y UBA, así como dos colaboraciones entre INTA e Idea Sys. Las demás colaboraciones fueron de un solo trabajo entre las instituciones involucradas. También puede ser verificada la presencia de una institución extranjera, la *University of California* de Estados Unidos.

En el segundo CAI (2009 - Mar del Plata), presentado en la Figura 4b, es posible reforzar la importancia del INTA, en términos de la centralidad. Todos los demás nodos tuvieron la centralidad cero. Tres instituciones no tuvieron ninguna colaboración. Acerca del tamaño de los nodos, INTA tuvo 12 trabajos publicados, Universidad Nacional del Litoral (UNL) tuvo 7 y Universidad Nacional del Centro de la Provincia de Buenos Aires (UNICEN) tuvo 4. Todas las colaboraciones fueron de un solo trabajo entre las instituciones involucradas. Se verifica que esta edición tuvo su tamaño reducido en relación con la anterior, con menos instituciones y menos colaboraciones. Tampoco hubo participación de instituciones extranjeras.

En el tercer CAI (2010 - Buenos Aires), presentado en la Figura 4c, se mantiene la importancia del INTA y crece la importancia del CONICET, en términos de centralidad. Todos los demás nodos tuvieron centralidad muy baja. Tres instituciones no tuvieron ninguna colaboración. Acerca del tamaño de los nodos, INTA tuvo 15 trabajos publicados, CONICET y Universidad Tecnológica Nacional (UTN) tuvieron 3. Fueron identificadas seis colaboraciones entre instituciones. Puede ser verificada la presencia de cuatro instituciones extranjeras, siendo dos de Uruguay, una de Brasil y una de España. Esta edición también fue la mayor de los primeros cuatro años, en relación a cantidad de instituciones.

En el cuarto CAI (2011 - Córdoba), presentado en la Figura 4d, es posible verificar la importancia de la Universidad Nacional de Córdoba (UNC) y la importancia de CONICET se mantiene, en términos de la centralidad. Todos los demás nodos tuvieron la centralidad muy baja. Cuatro instituciones no tuvieron ninguna colaboración. Acerca del tamaño de los nodos, UNC tuvo 10 trabajos publicados, CONICET tuvo 8 e INTA tuvo 7. Se enfatizan las seis colaboraciones entre CONICET e INTA, así como las cuatro colaboraciones entre UNC e INTA. Se verifica la presencia de cuatro instituciones extranjeras, siendo dos de España, una de Costa Rica y una de Venezuela.

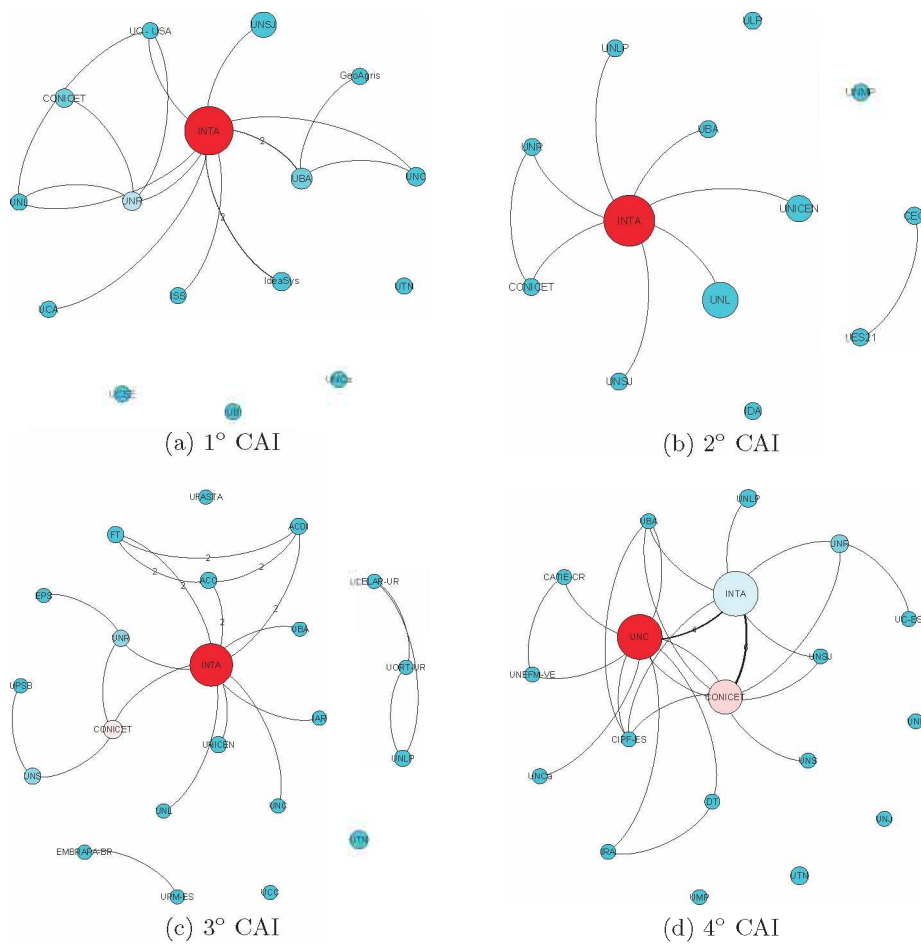


Fig. 4: Redes de colaboración del primer período de cuatro años

En el quinto CAI (2013 - Córdoba), presentado en la Figura 5a, muestra una interrupción casi completa de la red de colaboración en agroinformática, a través de una red poco conectada. Esta fue la edición con menos instituciones y menos colaboraciones en la historia del CAI. Todos los nodos tuvieron la centralidad igual a cero. Una institución no tuvo ninguna colaboración. Por la primera vez, dos de las instituciones más importantes, INTA y CONICET, no tienen ningún trabajo en conjunto. Acerca del tamaño de los nodos, UNC tuvo 6 trabajos publicados, INTA tuvo 5 y UNL tuvo 4. Se enfatiza la existencia de solamente cuatro colaboraciones. Tampoco hubo participación de instituciones extranjeras.

En el sexto CAI (2014 - Buenos Aires), presentado en la Figura 5b, se presenta una red apenas más conectada que el año anterior. CONICET e INTA se destacan por la métrica de intermediación. Todos los nodos tuvieron la centralidad igual a cero. Tres instituciones no tuvieron ninguna colaboración. Todas las colaboraciones fueron de un único trabajo. Acerca del tamaño de los nodos, UNC e INTA tuvieron 4 trabajos publicados y CONICET tuvo 3. Se verifica la presencia de cuatro instituciones extranjeras, siendo dos de Uruguay y dos de Colombia.

En el octavo CAI (2016 - Buenos Aires), presentado en la Figura 5c, muestra el evento con mayor cantidad de instituciones diferentes, así como la mayor cantidad de colaboraciones. Las instituciones con mayor métrica de centralidad fueron CONICET, INTA y UNL. Todos los demás nodos tuvieron la centralidad muy baja. Cuatro instituciones no tuvieron ninguna colaboración. Fueron encontradas seis colaboraciones entre CONICET e INTA, así como tres colaboraciones entre CONICET y UNL, lo que refuerza la importancia del CONICET. Acerca del tamaño de los nodos, INTA tuvo 14 trabajos publicados y CONICET tuvo 11. Se verifica la presencia de cinco instituciones extranjeras, de los Estados Unidos. Una peculiaridad de esta edición es que muchas instituciones no están en la red de colaboración de INTA y CONICET.

En el noveno CAI (2017 - Córdoba), presentado en la Figura 5d, se muestra una sólida red de colaboración con casi todas las instituciones argentinas que participaron del evento. Solamente la Universidad Nacional San Luis (UNSL) y la Universidad Tecnológica Nacional (UTN) no componen la red Argentina, siendo que la última no tuvo ninguna colaboración. Las instituciones con mayor métrica de centralidad fueron INTA y CONICET. Todos los demás nodos tuvieron la centralidad muy baja. Sobre el tamaño de los nodos, INTA tuvo 30 trabajos publicados, CONICET y UNC tuvieron 17. Se verifica la presencia de ocho instituciones extranjeras, siendo dos de Chile, dos de Paraguay, dos de Venezuela, una de España y una de Alemania. Las instituciones de los tres primeros países colaboraron entre sí. Las demás, colaboraron con instituciones de la Argentina. Además, había un autor de Uruguay sin filiación que no tuvo ninguna colaboración.

En el décimo CAI (2018 - Buenos Aires), presentado en la Figura 6, se presenta una sólida red de colaboración con casi todas las instituciones que participaron de esta edición del evento. Tres instituciones no participaron de trabajos en colaboración: UTN, Fundación Universidad Argentina de la Empresa (FUADE)

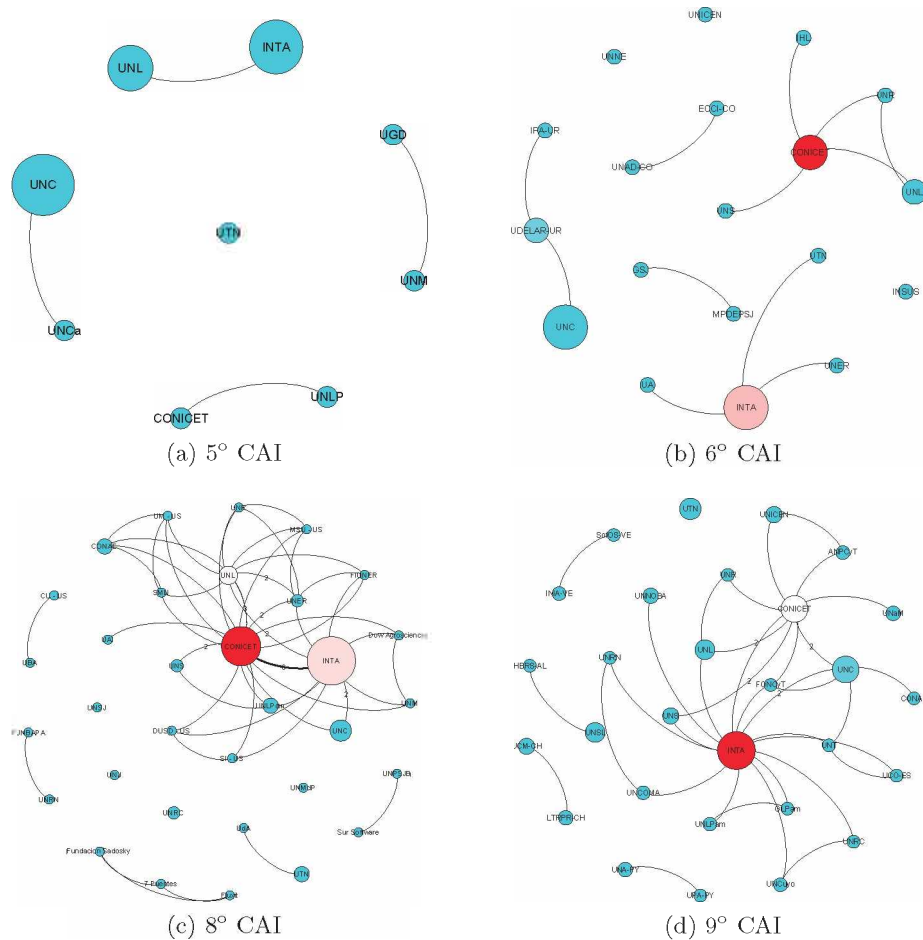


Fig. 5: Redes de colaboración de lo segundo período de cuatro años

y UNSL. Las instituciones con mayor métrica de centralidad fueron CONICET, INTA y Universidad Católica del Maule (UCM). Acerca del tamaño de los nodos, INTA tuvo 19 trabajos publicados, CONICET tuvo 14 y Universidade Federal do Pampa (UNIPAMPA) tuvo 13. Se verifica la presencia de trece instituciones extranjeras, siendo cinco de Chile, cuatro de Estados Unidos, tres de Brasil y una de Sudafrica. La UCM de Chile ejerce un papel muy importante por ser un enlace con la red Argentina. En este sentido, UNIPAMPA puede cumplir el mismo rol en relación a Brasil.

ciones posibles, habiendo potencial para incremento de las coautorías. También se puede verificar que hay una tendencia decreciente, con la disminución de la razón de las colaboraciones existentes dividida por las colaboraciones posibles, visto que los mayores valores históricos fueron detectados en las primeras ediciones del evento.

Acerca de la modularidad de la red, es posible percibir una tendencia de disminución, con el menor valor observado en la última edición. Esta métrica indica que la tendencia de colaboración es solamente internas en las comunidades. En la Figura 6 se percibe que la comunidad de Chile y Argentina solamente colaboran por el vínculo entre CONICET y UCM. Por otro lado, la comunidad de Brasil está muy integrada con la comunidad Argentina, por medio del INTA.

Acerca de los componentes conectados, hay una leve tendencia de crecimiento. En contraste, la última edición tuvo el menor valor: cuatro componentes. Se puede verificar que hay tres componentes compuestos por las instituciones que no tuvieron colaboraciones: UNSL, UTN y FUAJ, y un gran componente, donde están INTA, CONICET y las demás instituciones.

En relación a los trabajos publicados, hay también una tendencia de crecimiento. En la última edición, esta cantidad alcanzó su mayor valor en la historia del evento.

Tabla 3: Métricas de las redes de colaboración de cada año

Año	2008	2009	2010	2011	2013	2014	2016	2017	2018
Métrica / Edición	1°	2°	3°	4°	5°	6°	8°	9°	10°
Cantidad de Nodos (Instituciones)	16	13	22	19	9	19	33	29	33
Cantidad de Aristas (Colaboraciones)	15	9	21	25	4	12	50	35	51
Grado Promedio	1,88	1,39	1,91	2,63	0,89	1,26	3,03	2,41	3,09
Grado Promedio Ponderado	2,12	1,39	2,46	3,5	0,89	1,26	3,76	2,69	4,42
Diámetro de la Red	4	2	4	4	1	2	2	3	5
Densidad de la Red	0,13	0,12	0,09	0,14	0,11	0,07	0,10	0,09	0,10
Modularidad	0,26	0,27	0,43	0,21	0,75	0,72	0,27	0,41	0
Componentes Conectados	5	5	6	5	5	8	10	6	4
Tamaño Promedio del Camino	2,00	1,69	2,17	2,12	1,00	1,43	1,69	2,07	2,66
Trabajos Publicados	28	25	28	28	27	18	37	31	58

Así, el análisis de los nueve grafos, sobre un punto de vista histórico, muestra el incremento de la cantidad de trabajos, de instituciones y de colaboraciones.

3.4 Análisis de las Métricas de los Nodos

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. En la historia del CAI, autores de más de 125 instituciones tuvieron

participación en el evento, además de profesionales autónomos y autores que no declararon filiación. Las diez instituciones con mayor cantidad de autorías son presentadas en la Tabla 4 y suman 889 autorías, o sea, más de 70% del total. Con el fin de simplificar la visualización de los resultados, se seleccionaron solamente las cuatro instituciones con mayor cantidad de autorías en la historia del CAI. Así, son presentadas las métricas de las primeras cuatro instituciones: INTA, CONICET, UNC y UNL, las cuales suman 651 autorías, o sea, más de 50% del total de 1265 filiaciones declaradas en todos los trabajos presentados en la historia del congreso. El valor de 1265 es mayor que las 1112 autorías porque algunos autores tienen doble filiación. Otra observación relevante en estos datos, es que la UNIPAMPA tuvo 36 autorías teniendo participado solamente del 10° CAI. Todas las demás instituciones tuvieron participación en todas o casi todas las ediciones. Otro dato a ser observado es que más de 50 instituciones tuvieron autorías en un único trabajo, y más de 25 tuvieron autorías en solamente dos trabajos en la historia del CAI. La Figura 7 presenta los resultados, de las cuatro instituciones, para las siguientes métricas: cantidad de trabajos, grado, intermediación y *pagerank*.

Tabla 4: Instituciones con mayor cantidad de autorías en la historia del CAI

Cantidad de Autorías	Acrónimo	Institución	País
311	INTA	Instituto Nacional de Tecnología Agropecuaria	Argentina
122	CONICET	Consejo Nacional de Investigaciones Científicas y Técnicas	Argentina
113	UNC	Universidad Nacional de Córdoba	Argentina
105	UNL	Universidad Nacional del Litoral	Argentina
66	UTN	Universidad Tecnológica Nacional	Argentina
40	UNSJ	Universidad Nacional de San Juan	Argentina
36	UNIPAMPA	Universidade Federal do Pampa	Brasil
35	UNICEN	Universidad Nacional del Centro de la Provincia de Buenos Aires	Argentina
34	UNS	Universidad Nacional del Sur	Argentina
27	UNR	Universidad Nacional de Rosario	Argentina
889	Total		

Conforme presentado en la Figura 7a, en general INTA tiene la mayor cantidad de publicaciones en las ediciones del evento, como puede ser observado en las 1ª, 2ª, 3ª, 6ª, 8ª, 9ª y 10ª ediciones. Pero esto no ocurrió en la 4ª y 5ª ediciones, que fueron muy atípicas y con poca cantidad de trabajos, donde UNC tuvo la mayor cantidad. De cualquier forma, INTA sobresale con la mayor cantidad de publicaciones en la historia del CAI, consolidándose como la institución más productiva en este evento. Además, en seis de las nueve ediciones (1ª, 3ª, 4ª, 8ª, 9ª y 10ª), CONICET tuvo la segunda mayor cantidad de trabajos, lo que

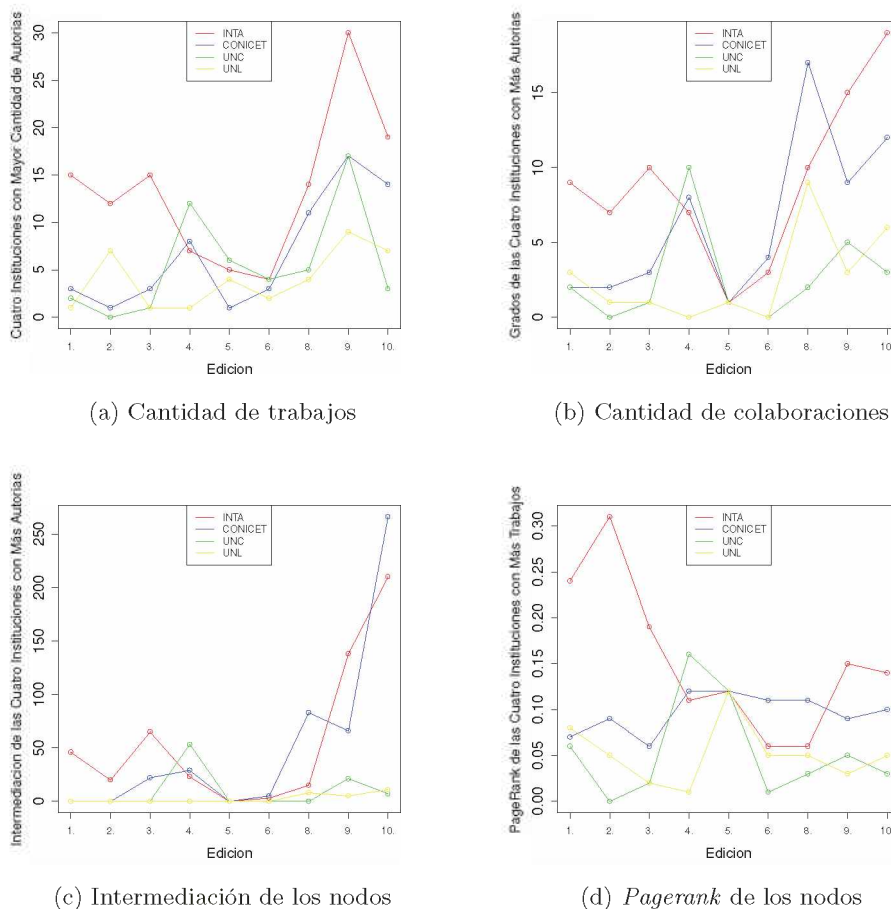


Fig. 7: Métricas de los principales nodos en nueve ediciones del CAI.

contribuyó para su segunda posición como institución más productiva. Dentro de las cuatro instituciones más productivas, UNL típicamente ocupa la cuarta posición en cantidad de trabajos, como ocurrió en 1^a, 3^a, 4^a, 6^a, 8^a y 9^a ediciones.

Los grados de los nodos, presentados en la Figura 7b, representan la cantidad de colaboraciones. En seis de las diez ediciones (1^a, 2^a, 3^a, 5^a, 9^a y 10^a), INTA tuvo la mayor cantidad de colaboraciones, lo que muestra su importancia como colaborador de muchas instituciones en la Argentina en la investigación y desarrollo en agroinformatica. Hay que observar también, que en las últimas cuatro ediciones, INTA presenta una tendencia creciente de cantidad de colaboraciones linealmente creciente, muy bien definida. Por su parte, CONICET fue la primera institución en cantidad de colaboraciones en la 5^a, 6^a y 8^a ediciones,

y la segunda en 2^a, 3^a, 4^a, 9^a y 10^a ediciones. UNC y UNL tuvieron picos de cantidad de colaboraciones en 4^a y 8^a ediciones, respectivamente. Además, en la 4^a edición, UNC fue la institución con más colaboraciones.

Acerca de la métrica de intermediación, presentada en la Figura 7c, se percibe que INTA y CONICET tienen un incremento muy grande de su valor de intermediación en las últimas ediciones. Esta evidencia nos permite concluir que están tornándose elementos centrales en la red de investigación agropecuaria en la Argentina. Por otro lado, en muchas ediciones, los valores de intermediación muy bajos de UNC y UNL, muestran que estas instituciones tienen la tendencia de estar en los bordes de las redes de colaboración, con un rol secundario, y no en el centro con un papel destacado.

En relación al *pagerank*, INTA y CONICET fueron las dos instituciones con mayor valor para esta métrica en la historia del CAI. La excepción fue en el 4^o CAI, con el mayor *pagerank* de la UNC. En las tres primeras ediciones, INTA tuvo una importancia fundamental. A partir de la 5^a edición, pasó a alternar el papel de institución más importante con CONICET.

Todos los gráficos en la Figura 7 muestran una participación muy destacada de INTA en las tres primeras ediciones. En la 4^a edición, llevada a cabo en Córdoba, UNC tuvo un papel destacado. A partir de la 5^a edición, INTA y CONICET se alternan como instituciones más importantes en el contexto del CAI.

3.5 Análisis de Tendencias

A fin de identificar las tendencias, fue hecho un análisis estadístico de las métricas de la Tabla 3. Este análisis fue realizado en tres pasos para cada una de las métricas: 1) prueba de normalidad de la métrica, 2) cálculo de correlación entre la métrica y la edición del evento, 3) creación de un modelo de regresión lineal para identificación de la tendencia. Para estos pasos, fue delimitado un nivel de significancia de 0,05 ó 5%. Los resultados del análisis son presentados en la Tabla 5.

Como resultado de la prueba de normalidad, utilizando un grado de confianza de 95%, fue detectado que las métricas Componentes Conectados y Trabajos Publicados tienen una distribución normal, con *valor p* de 0,0365 y 0,0144, respectivamente. Las demás métricas no tienen una distribución normal, dado que su *valor p* fue superior a 0,05.

Para el cálculo de la correlación de las métricas con las ediciones, a pesar que dos métricas tienen una distribución normal, se utilizó el método no-paramétrico de Spearman, visto que las ediciones tampoco tienen una distribución normal. Como resultado del análisis de correlación, se detectó una fuerte correlación positiva (*valor p* > 0,7) entre la Cantidad de Instituciones y las ediciones del evento, con grado de confianza de 95%. Así, como resultado de la regresión lineal, se detectó un pendiente de 2,1083, es decir que, a cada nueva edición, se puede esperar la participación de autores de más 2,1083 instituciones. De esta forma, hay una comprobación estadística del crecimiento del CAI, que involucra autores de más instituciones en cada edición.

Hay una correlación positiva moderada ($valor\ p > 0,4$) de la edición con las métricas de Cantidad de Colaboraciones, Trabajos Publicados, Grado Promedio y Grado Promedio Ponderado. Hay también una correlación negativa moderada ($valor\ p < -0,4$) de la edición con la métrica de Densidad de la Red. De esta forma, los datos presentan, a cada edición del CAI, una tendencia de crecimiento de 4,1125 colaboraciones y 2,2333 trabajos. También hay una leve tendencia de crecimiento de 0,1392 en el grado promedio y 0,2104 en el grado promedio ponderado, lo que muestra que las instituciones tienden a colaborar más. Acerca de la densidad de la red, hay una tendencia casi cero, o sea, de permanencia en el mismo nivel. Pero en estas correlaciones moderadas, tanto en las positivas como en las negativas, el $valor\ p$ no permite asegurar la significancia estadística de estas afirmaciones, con nivel de confianza de 95%.

Las demás correlaciones (Diámetro de la Red, Modularidad, Componentes Conectados y Tamaño Promedio del Camino) son muy débiles ($valor\ p < 0,2$) y tampoco son estadísticamente significativas, visto que sus $valores\ p$ son mayores que 0,05.

Tabla 5: Análisis Estadístico de las Tendencias

Métrica	Normalidad		Correlación con Edición		Tendencia	
	P-Value	Coefficiente	P-Value	Intercepto	Pendiente	
Cantidad de Instituciones	0,4864	0,7059	0,0336	10,2000	2,1083	
Cantidad de Colaboraciones	0,2931	0,6333	0,0760	2,7333	4,1125	
Grado Promedio	0,6337	0,5000	0,1777	1,3120	0,1392	
Grado Promedio Ponderado	0,7544	0,5000	0,1777	1,3767	0,2104	
Diámetro de la Red	0,3961	0,0345	0,9298	2,8667	0,0250	
Densidad de la Red	0,9398	-0,5294	0,1427	0,1247	-0,0036	
Modularidad	0,3615	-0,0418	0,9149	0,4260	-0,0107	
Componentes Conectados	0,0365	0,1486	0,7027	5,2000	0,1500	
Tamaño Prom. del Camino	0,8866	0,1590	0,6828	1,6740	0,0368	
Trabajos Publicados	0,0144	0,5594	0,1173	19,2000	2,2333	

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 de todas las ediciones del Congreso Argentino de Agroinformática (CAI), llevados a cabo de 2008 hasta 2018.

Los análisis realizados en este trabajo muestran las evidencias del constante crecimiento de la importancia del INTA en el contexto de la agroinformática en Argentina. En este mismo entorno, el CONICET también viene presentando

un papel muy representativo. Las universidades que aparecen como actores importantes se encuentran localizadas en la Pampa Húmeda (con excepción de la UNSJ). Esta situación, en parte, se debe a: a) la preponderancia e importancia económica de las producciones agropecuarias en la Pampa Húmeda y b) la realización del congreso siempre con sede en alguna ciudad de esta región de Argentina. Relacionado con el punto anterior, 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 y apertura a nuevos grupos de trabajos e instituciones. La 11^o edición del CAI se realizará en la ciudad de Salta, una de las más importantes del NOA. Se tiene como hipótesis que al llevar la sede del evento a esta región permitirá que los grupos de investigación y desarrollo de la región puedan participar con mayor facilidad.

Conocer a las instituciones relevantes para el CAI y como se relacionan entre ellas, es de utilidad, no solo para los grupos de trabajo de estas instituciones y para la organización del congreso, sino también para otros actores del sector agropecuario y agroindustrial que deseen investigar o desarrollar temáticas de agroinformática y actores del sector científico tecnológico para identificar grupos de trabajos en estas temáticas, tanto sea para consulta, asesoramiento o financiamiento.

Como restricción del abordaje aplicado, fueron percibidos datos incorrectos. Dentro de estos errores, se pueden citar el nombre de las instituciones, provincias y ciudades escritos de formas distintas. Algunas veces se identificó que el nombre del departamento o facultad fue utilizado en el lugar del nombre de la institución. Otro ejemplo es el nombre de la ciudad en lugar del nombre de la provincia. En algunos casos, no fueron completados o corregidos los datos de filiación. Estos problemas ocurrieron especialmente en aquellos trabajos enviados en las primeras ediciones donde no se contaba con el sistema de envíos de trabajos y tampoco con la plantilla unificada para el formato de las presentaciones de las contribuciones. Desde la edición 2013 del CAI se cuenta con el sistema Open Conference System y se utiliza la plantilla de LNCS (Lecture Notes in Computer Science) de Springer. Igualmente, una comisión de admisión de trabajos que se ocupe de chequear que todos los datos en el sistema estén completos es importante.

Los trabajos de instituciones extranjeras se restringieron a contribuciones puntuales, en ediciones específicas. Uruguay tuvo trabajos presentados en cuatro ediciones; Estados Unidos y España en tres; Brasil, Chile y Venezuela en dos; Alemania, Colombia, Costa Rica y Paraguay en una edición. La participación de otros países como miembros del comité científico y organizador podría ayudar a darle un contexto internacional permanente y también podría mejorar la difusión de los llamados para presentación de trabajos.

Continuar con este tipo de análisis sumando las nuevas ediciones permitirá chequear hipótesis planteadas en este trabajo, detectar la utilidad de las mediadas tomadas por la organización del congreso y mantener actualizado

el conocimiento sobre las instituciones relevantes en la temática de AgroInformática.

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*. pp. 16–23. 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. Camargo, S.S., de Pinho, L.B., Saibene, Y.B.: Congreso argentino de agroinformática: Un análisis bibliométrico. In: *Anales del X Congreso Argentino de Agroinformática (CAI)*. pp. 434–445. Buenos Aires, Argentina (2018)
6. 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)
7. Diestel, R.: *Graph Theory*. No. 173 in *Graduate Texts in Mathematics*, Springer (1997)
8. Foulds, L.: *Graph theory applications*. Springer-Verlag, New York (1992)
9. 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)
10. Newman, M.E.J.: *Networks: An Introduction*. Oxford University Press, New York, NY, USA (2010)
11. Otte, E., Rousseau, R.: Social network analysis: a powerful strategy, also for the information sciences. *Journal of information Science* 28, 441–453 (2002)
12. Sampaio, R.B., Sacerdote, H.C.d.S., Fonseca, B.d.P.F., Fernandes, J.H.C.: A Colaboração científica na pesquisa sobre coautoria: um método baseado na análise de redes. *Perspectivas em Ciência da Informação* 20, 79 – 92 (Diciembre 2015)
13. Su, Y.F.: *Study on the Cooperation Network of Sport Research Papers*. Shanghai University of Sport, Shanghai (2010)
14. Wang, R., Zhang, W., Deng, H., Wang, N., Miao, Q., Zhao, X.: Discover community leader in social network with pagerank. In: Tan, Y., Shi, Y., Mo, H. (eds.) *Advances in Swarm Intelligence*. pp. 154–162. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
15. Wu, Y., Duan, Z.: Social network analysis of international scientific collaboration on psychiatry research. *Int J Ment Health Syst* 9(2) (Enero 2015)
16. 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)

17. 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)