



La Ciencia Abierta en América Latina desde el análisis y métricas de redes sociales

Patricia Garrido Villegas¹

Resumen

El principal objetivo de la Ciencia Abierta es fomentar que las investigaciones científicas sean replicables, accesibles, reutilizables para todos los niveles de la sociedad de manera gratuita. Al ser un tema novedoso es interesante saber cuál es la interacción entre la comunidad científica tanto de instituciones académicas como de investigadores, editores, empresas privadas y estudiantes en América Latina. Una de las maneras de saberlo son las redes sociales. Este trabajo se enfoca en Twitter. Construiré una red dirigida de colaboración y para ello trabajaré con esos *tweets* extraídos, con las menciones y con los *retweets* basándome en un modelo de algoritmo, cuyos resultados se llevarán al programa Gephi para analizar las relaciones entre nodos (usuarios de Twitter, por ejemplo, un editor o revista) y aristas (las relaciones entre dichos usuarios). El análisis se llevará a cabo a través de las siguientes métricas de red: *Modularity class*, la cual sirve para saber la detección de comunidades, más la distribución Fruchterman Reingold, ya que deja más claro el conjunto de grupo que discute la temática. El grado de entrada y grado de salida miden interacciones dentro de la red social creada. La centralidad del vector propio mide la importancia de un nodo en la red basada en sus conexiones. *Page Rank* sirve para ver las páginas de los nodos de acuerdo con la frecuencia con la que un usuario sigue enlaces para llegar a la página de dicho nodo de forma no aleatoria.

1 Universidad de Alcalá de Henares.

Palabras clave: Ciencia Abierta; Twitter; América Latina; distribución Fruchterman Reingold; *Modularity class*; centralidad del vector propio; grado de entrada y grado de salida.

Abstract

The main objective of Open Science is to encourage scientific research to be replicable, accessible, reusable for all levels of society free of charge. As this is a new topic, it is interesting to know what is the interaction between the scientific community of both academic institutions and researchers, publishers, private companies, and students in Latin America. One way to know it is through social networks. This work focuses on Twitter. I will build a targeted collaborative network and work with those extracted tweets, mentions and retweets based on an algorithm model, whose results will be taken to the Gephi program to analyze the relationships between nodes (Twitter users, for example, a publisher or magazine) and edges (the relationships between those users). The analysis will be carried out through the following network metrics: Modularity class, which serves to know the detection of communities, plus the Fruchterman Reingold distribution, because it makes clearer the set of groups that discuss the issue. The in-degree and out-degree measure interactions within the social network created. The eigenvector centrality measures the importance of a node in the network based on its connections. Page Rank is used to view node pages according to how often a user follows links to that node's page on a non-random basis.

Keywords: Open Science; Twitter; Latin America; Fruchterman Reingold distribution; Modularity class; eigenvector centrality; in-degree and out-degree.

Introducción

La Ciencia Abierta fomenta que las investigaciones científicas sean replicables, accesibles, reutilizables para todos los niveles de la sociedad de manera gratuita. Permite una difusión mayor y más rápida de las investigaciones e implica un aumento de las redes de colaboración. Dichas redes de colaboración pueden llevarse a cabo a través de las redes sociales. Este trabajo se enfoca en la red social Twitter.

You can make your workflow more open by ...

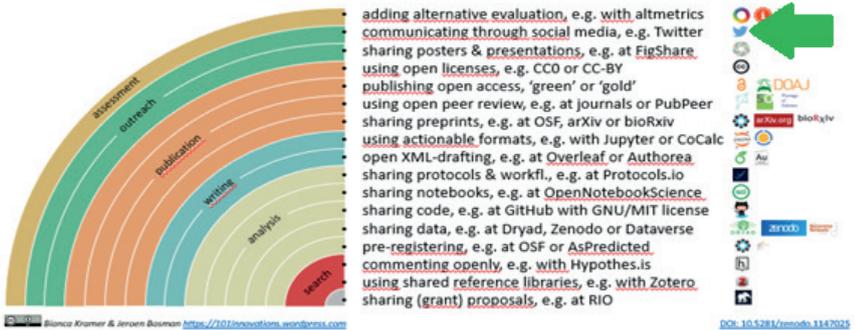


Figura 1. Prácticas científicas abiertas en todo el flujo de trabajo de investigación para hacerlo más abierto (Kramer *et al.* 2018).

¿Qué es al análisis de redes sociales?

El análisis de redes sociales “se basa en la suposición de la importancia de las relaciones entre las unidades que interactúan” (Wasserman *et al.* 1994), implica llevar a cabo una metodología sobre una estructura formada por nodos y aristas. Surgió de la interacción de varias disciplinas: las matemáticas en su rama llamada teoría de grafos enfocada en los datos; las ciencias de la computación debido a que usa algoritmos, bases de datos, así como desarrollo de *software* (Barabási 2016). De igual manera, considera a la sociometría, que es un método cuantitativo para medir las relaciones sociales desarrollado por el psiquiatra y psicopedagogo rumano Jacob Levy Moreno, quien la define como

"la investigación sobre la evolución y organización de grupos y la posición de los individuos dentro de dichos grupos" (Mora-Cantallops 2018).

Pregunta principal

¿Cuál es la interacción entre la comunidad científica tanto de instituciones académicas como de investigadores, editores, empresas privadas y estudiantes en América Latina?

Objetivos

1. Localizar a los usuarios más implicados en la temática de la Ciencia Abierta.
2. Saber cuáles son los principales exponentes en América Latina que nos puedan demostrar el avance en la discusión sobre el tema en la región.
3. Saber cuál es el grado de relación entre los usuarios más interesados en discutir la Ciencia Abierta.

Metodología

- Los datos se extrajeron con la API de Twitter con la temática *#openscience*.
- Se utilizó el programa Python y su módulo NetworkX, así como Tweepy, almacenando los datos en MongoDB.
- Se buscó por el más popular de los *tweets*, se extrajo por *retweet* y por ID del usuario.
- Se construyó una red de colaboración dirigida con los *tweets* extraídos, con las menciones y con los *retweets*.
- Se realizó un modelo de algoritmo con NetworkX para obtener nodos (usuarios de Twitter, por ejemplo, un editor, revista o institución) y aristas (las relaciones entre dichos usuarios).
- Se utilizó el programa Gephi para la visualización y el uso de métricas de las redes de colaboración obtenida.

El análisis se llevó a cabo a través de las siguientes métricas de red, las cuáles son las más habituales en el análisis de redes sociales y están basadas en algoritmos:

- Modularity class: Sirve para saber la detección de comunidades más la distribución Fruchterman Reingold, la cual deja más claro el conjunto de grupo que discute la temática.
- Grado de entrada: Número de conexiones entrantes de un nodo con otros.
- Grado de salida: Número de conexiones salientes de un nodo con otros.
- Centralidad del vector propio: Mide la importancia de un nodo en la red basada en sus conexiones.
- Page Rank: Mide la frecuencia con la que un usuario sigue enlaces para llegar a la página de un nodo de forma no aleatoria.

Procedimiento

Como se puede observar en la Figura 3, de los resultados del algoritmo (Figura 2) se obtuvieron 4,143 nodos (Instituciones, investigadores, editores, revistas) y 4,901 aristas (las relaciones entre ellos), cuyos tuits fueron obtenidos entre el 17 de febrero y el 5 de marzo de 2020.

```

[5]: import tweepy
import json
import pandas as pd
import numpy as np
from pymongo import MongoClient
import networkx as nx

consumer_key = "6v90T2mew1zDk4LCspC641Wq"
consumer_secret = "c92R354d359Wpql4E6mFv4bdrkrU3EEkxU23v170u3M"
access_token = "17255164-U2u0Ks38&9F4Kuae0249SE408qg4q87cF1"
access_token_secret = "guknADCj03ckIpfV5Wu4G1qg5q5VtPu07949vZgk4C"

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)

client = MongoClient()
db = client.test_paty0V

[6]: result = api.search(q="openaccess", count = 1000)

[7]: result

[8]: [Status(api=tweepy.api.API object at 0x000022287EF740, _json={'_type': 'twee
2020', 'id': 1229285248829322, 'id_str': '1229285248829322', 'text': 'Bicent
a y ponencia' in 120000. XII Congreso Iberoamericano de Ciencias
a lista de r., 'truncated': False, 'entities': {'hashtags': [{'text':
ent': 'Tecnología', 'indices': [95, 106]}, {'text': 'Openmo', 'indices':
sentencia': {'screen_name': 'cicapsal', 'name': 'CICAPSA', 'id': 4670054
[3, 11]}], 'urls': []}, 'metadata': {'iso_language_code': 'es', 'result
@Xena17/hashta_rutaw_rwa' ret=4u0f11uP+Puraw= 50k 50w/4s', 'in:

[9]: result = api.search(q="openaccess", result_type =
'location')

[10]: result = api.search(q="openaccess", result_type =
'topical')
result
    
```

Figura 2. Modelo de algoritmo con NetworkX.

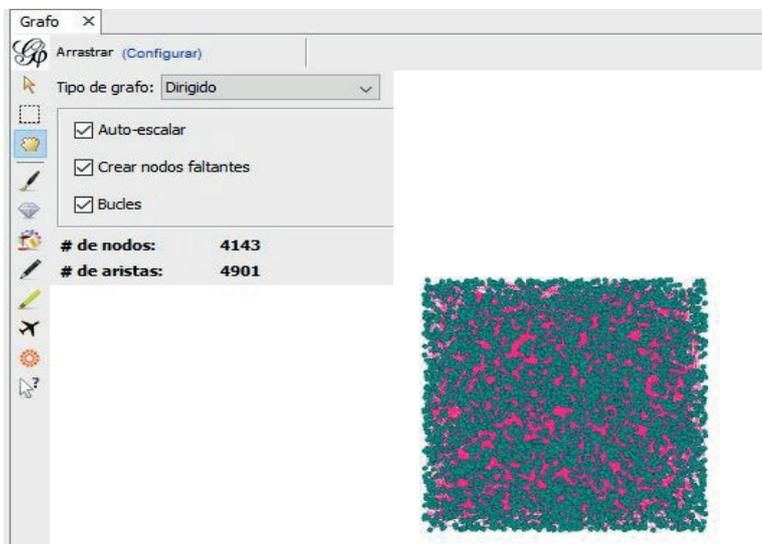


Figura 3. Resultados de la extracción en Gephi.

Análisis y visualización con las métricas

Se llevó el grafo de la red a Gephi, donde se comenzó a darle forma, de tal manera que se pueda responder a las preguntas planteadas al inicio. Se utilizó la distribución Fruchterman Reingold, la cual es un algoritmo de separación por atracción de relaciones entre los nodos y permite que se vean de manera más clara las diferentes comunidades.

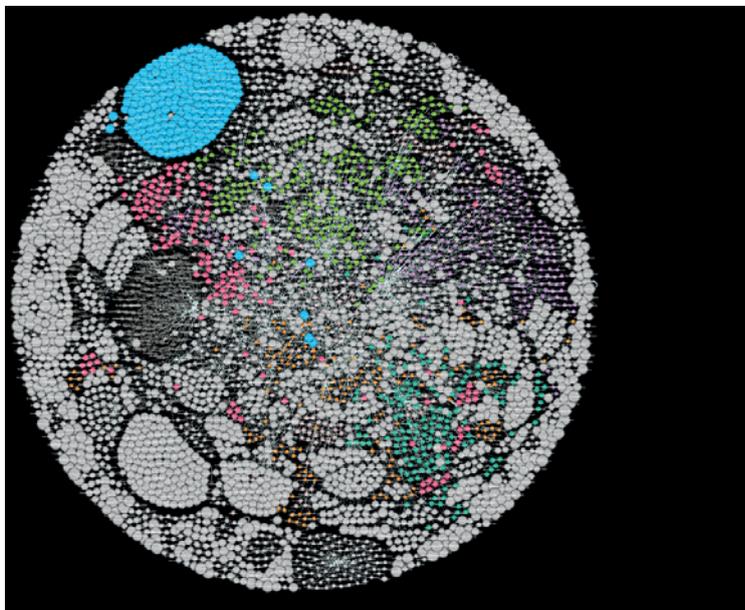


Figura 4. Distribución Fruchterman Reingold de toda la red.

La anterior figura es el panorama mundial del hashtag *#openscience* en Twitter. Sin embargo, este trabajo se enfoca en América Latina, por ello, de esos más de cuatro mil resultados, lo mejor fue realizar una búsqueda de los nodos más importantes que están involucrados en la temática de Ciencia Abierta. Uno de ellos es SciELO México. La comunidad a la que pertenece SciELO México es grande y por ello el análisis se enfoca en sus vecinos más cercanos. La investigación arrojó que se trata de usuarios provenientes de España, en particular de Barcelona, la mayoría son investigadores, profesores, grupos de investigación, bibliotecas, divulgadores de ciencia y tecnología e instituciones financiadoras de la ciencia.

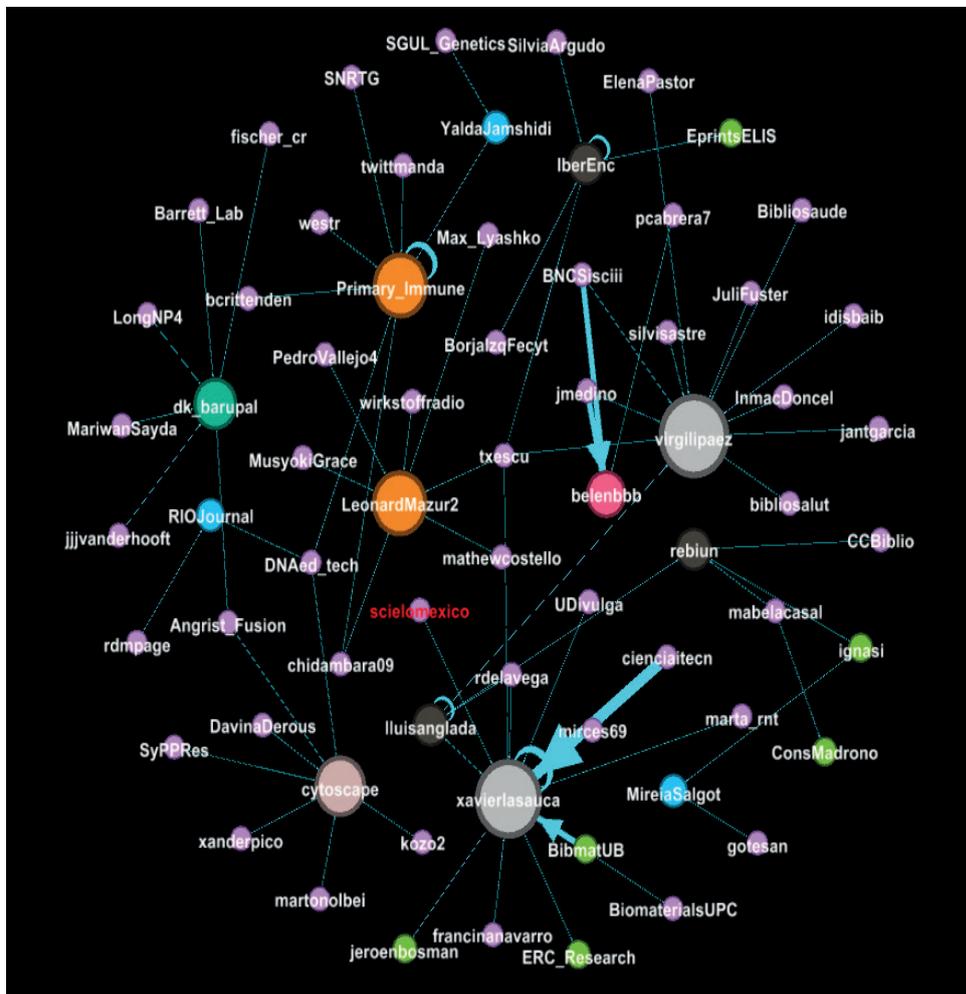


Figura 5. Grado de entrada del grupo en el que está relacionándose SciELO México.

En la Figura 5 se puede observar el grado de entrada, el cual es el número de conexiones entrantes de un nodo con otros, esto significa que entre más grande sea el tamaño del nodo (circulo), más usuarios interactúan con él, es decir, el usuario A es seguido por B y C. De igual modo, el color del nodo significa que esos usuarios tienen el mismo grado de entrada. Por ejemplo,

xavierlausaca², es seguido por SciELO México, así como por otros once usuarios. SciELO México a su vez tiene una interacción de menor grado y pertenece a un mismo color de grupo de usuarios como rdelavega, quien es profesor de la Universidad de Barcelona. También en la imagen se puede ver que entre más gruesa es la línea de las aristas (flechas) más fuerte es la relación entre los nodos.

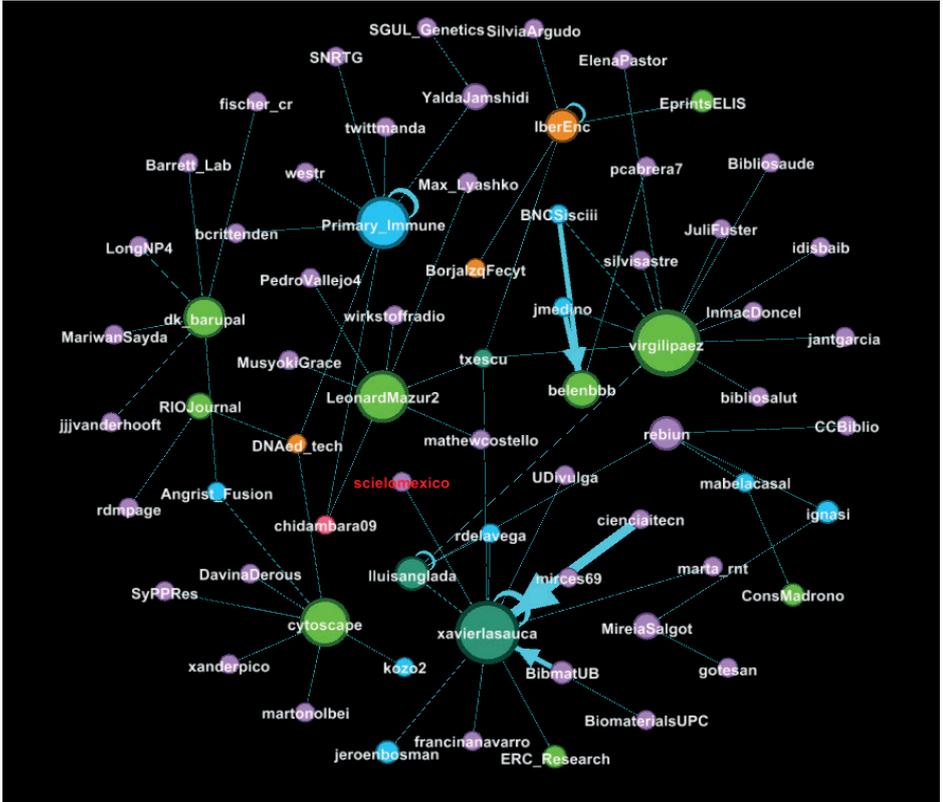


Figura 6. Grado de salida del grupo en el que está relacionándose SciELO México.

2 “Responsable de la gestión del conocimiento y sistemas de información I+D, encargado de la coordinación técnica de Recercat, el boletín electrónico de la investigación en Cataluña, y de coadministrar los perfiles sociales de la Dirección General de Investigación de la Generalitat de Catalunya en Twitter y Facebook” (Gil 2015).

En la Figura 6 se puede observar el grado de salida, el cual es el número de conexiones salientes de un nodo con otros, es decir, el usuario A sigue a B y C. En este caso, entre más grande sea el tamaño del nodo significa que los usuarios siguen a otros en mayor medida. Al igual que en el caso anterior, el mismo color del nodo significa que tienen un idéntico grado de salida; mientras que, en las aristas, entre más gruesa es la línea, más fuerte es la relación entre los usuarios. Por ejemplo, SciELO México sigue a dos nodos que tienen el mismo color y por lo tanto el mismo grado de salida, se trata de lluisanglada³ quien a su vez sigue a xavierlausaca y ambos son quienes más siguen a otros nodos dentro de este grupo.

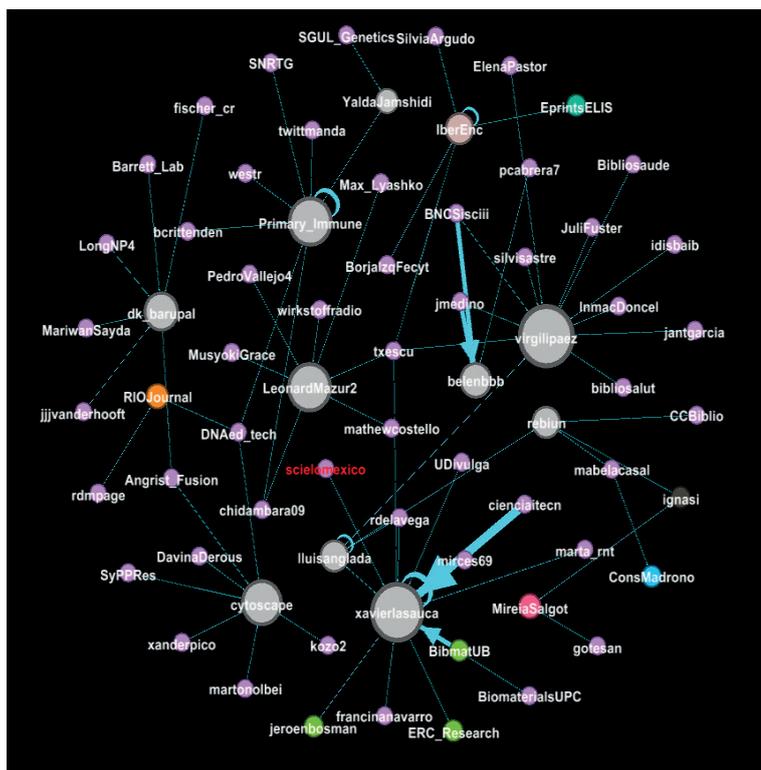


Figura 7. Centralidad del vector propio del grupo en el que está relacionándose SciELO México.

3 Es el director de Open Science del Consorcio de Servicios Universitarios de Cataluña.

En la Figura 7 vemos la centralidad del vector propio, el cual mide la importancia de un nodo en la red basada en sus conexiones, es decir, A es el nodo que más siguen los usuarios de la comunidad. Aunado a lo anterior se puede observar que entre más grande es el grosor de las aristas (flechas), más fuerte es la relación entre los nodos, mientras que el mismo color representa a los usuarios más importantes dentro de la comunidad. En este caso, nuevamente xavierlasauca y lluisanglada son parte de los nodos más importantes de la comunidad y son los usuarios con quién más interactúa SciELO México.

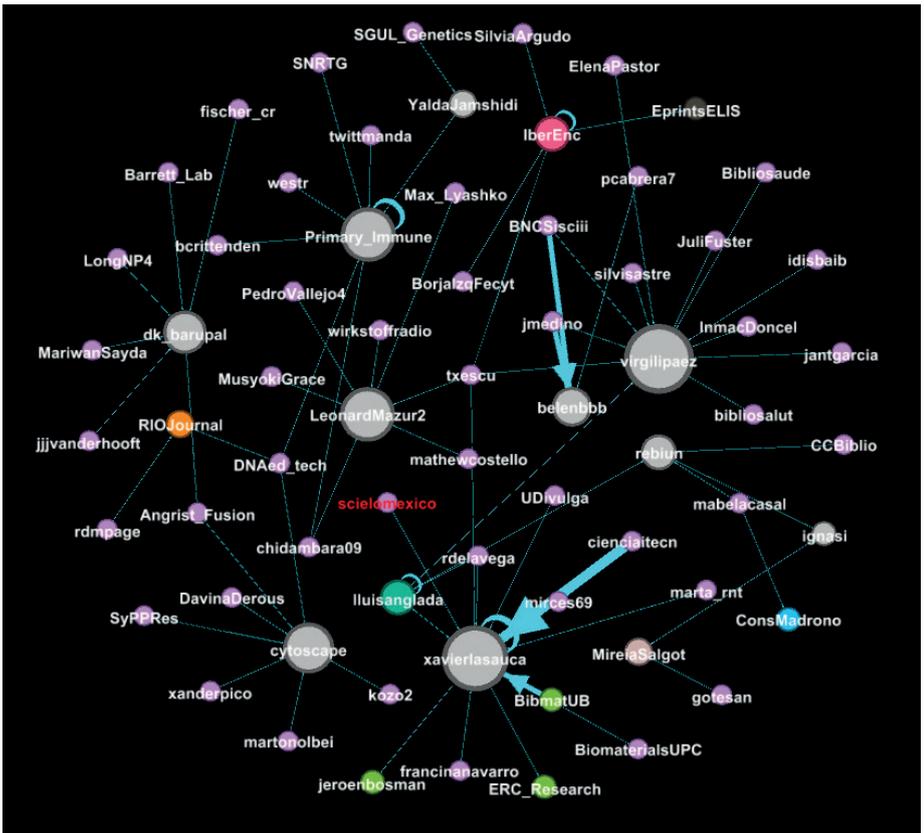


Figura 8. Page Rank del grupo en el que está relacionándose SciELO México.

En la Figura 8 se puede observar la métrica Page Rank, la cual mide la frecuencia con la que un usuario sigue enlaces

para llegar a la página de un nodo de forma no aleatoria. Entre más grande sea el nodo y más gruesa la línea de la arista significa que tienen una mayor frecuencia con la que otros nodos más pequeños llegan a los nodos más grandes. SciELO México tiene el mismo color, así como el mismo tamaño, que BiomaterialsUPC,⁴ por lo tanto tienen la misma frecuencia *Page Rank* y ambos usuarios son atraídos por xavierlasauca.

Otro caso en América Latina que está relacionándose con la temática de Ciencia Abierta en Twitter es la Biblioteca CEPAL.

4 Grupo de investigación español cuyo principal objetivo es desarrollar biomateriales para la regeneración/repación funcional de tejidos y órganos.

fosterscience,⁶ y ambos son proyectos que impulsan la Ciencia Abierta a nivel mundial.

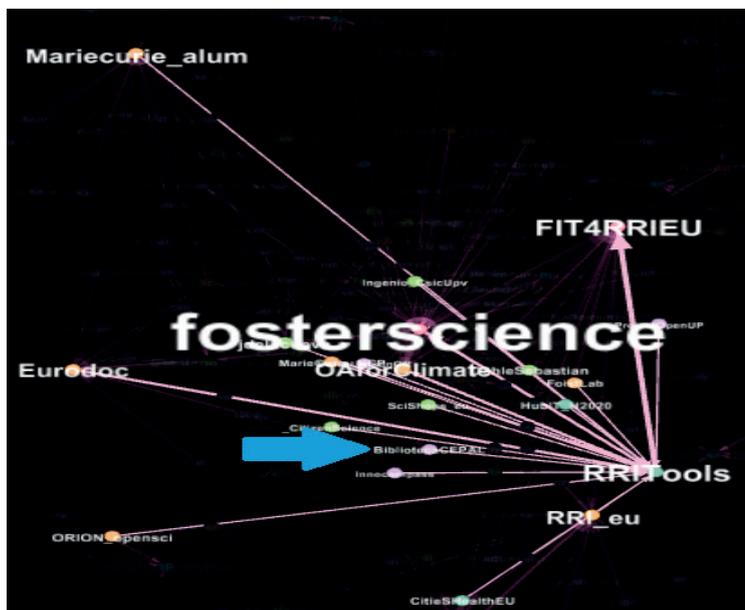


Figura 10. Grado de salida del grupo en el que está relacionándose la Biblioteca de la CEPAL.

En la Figura 10 podemos ver el grado de salida del grupo de la Biblioteca de la CEPAL, misma que está resaltada con la flecha color azul y su tamaño es muy pequeño, lo cual significa que el número de conexiones salientes con otros nodos de la comunidad es poca. Sin embargo, está relacionado con otros nodos importantes de la comunidad: RRITools y fosterscience.

6 El portal FOSTER es una plataforma de aprendizaje electrónico que reúne los mejores recursos de capacitación dirigidos a aquellos que necesitan saber más sobre Open Science, o que necesitan desarrollar estrategias y habilidades para implementar prácticas de Open Science en sus flujos de trabajo diarios (FOSTER portal s. f.).

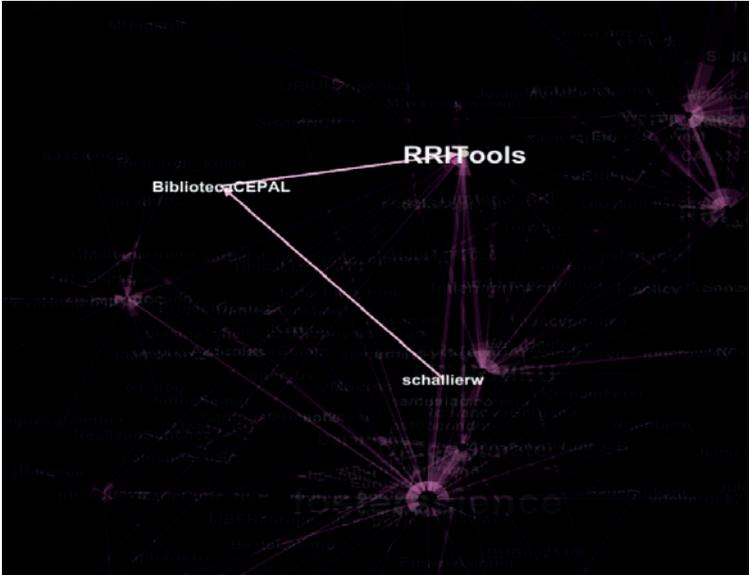


Figura 11. *Page Rank* del grupo en el que está relacionándose la Biblioteca de la CEPAL.

Con la métrica *Page Rank* se corrobora la relación que tiene la Biblioteca de la CEPAL con RRITools, así como con schallierw (quien es el Jefe de la Biblioteca Hernán Santa Cruz, ONU/CEPAL Chile), es decir, schallierw llega a RRITools a través de la Biblioteca CEPAL, y por ello esta última tiene un mayor grado de *Page Rank*.

Conclusiones

De acuerdo con lo visto en este análisis de la red social Twitter, la interacción entre la comunidad científica, tanto de instituciones académicas como de investigadores, editores y estudiantes en América Latina, respecto a la discusión de la Ciencia Abierta es poca hasta el momento.

Los usuarios más implicados en la temática de la Ciencia Abierta son portales y proyectos europeos como fosterscience o RRITools, quienes estuvieron por encima del resto de los nodos en las métricas. Hasta el momento, en América Latina, SciELO

México y la Biblioteca de la CEPAL han estado involucrados en la discusión debido a su interacción con nodos importantes.

La mayoría de los nodos del grupo donde está SciELO México son españoles y todo es debido a su interacción con el usuario Xavier Lasauca Cisa, quien es gerente de conocimiento en el Sistema de Universidades, Investigación e Innovación de Cataluña y un importante difusor de la Ciencia Abierta.

Si bien, no se lograron obtener muchos nodos participantes de la discusión de la Ciencia Abierta en América Latina, lo que sí se pudo saber fue quiénes son los más involucrados a nivel mundial y de esta manera seguirlos para la innovación de nuestras prácticas; como FOSTER, uno de los portales principales sobre Ciencia Abierta, y RRITools, proyecto europeo que ofrece herramientas de recursos abiertos.

Las visualizaciones que se presentaron permitieron saber el grado de relación entre los usuarios más interesados en discutir la Ciencia Abierta y que lleva a concluir que los más involucrados son investigadores, así como proyectos que incluyen a universidades.

Finalmente, se puede concluir que las redes sociales permiten incrementar el impacto y la visibilidad de las publicaciones de acuerdo con la comunidad con la que se relacionan.

Referencias

- Barabási, Alber-László. 2016. *Network science*. EUA: Cambridge University Press. <http://networksciencebook.com/>
- FOSTER portal. s. f. "About the FOSTER portal". Consultado el 8 de diciembre de 2023. <https://www.fosteropenscience.eu/about#theportal>.
- Gil, Lydia. 2015. "Charlando con Xavier Lasauca sobre comunicación científica 2.0". *Social Media en Investigación*. Publicado el 11 de mayo de 2015. <https://socialmediaeninvestigacion.com/xavier-lasauca-comunicacion-cientifica-20/>.
- Kadushin, Charles. 2012. *Understanding social networks: Theories, concepts, and findings*. EUA: OUP.
- Kramer, Bianca y Jeroen Bosman. 2018. "Arcoiris de las prácticas de ciencias abiertas". *Zenodo*. Publicado el 14 de enero de 2018. <http://doi.org/10.5281/zenodo.1147025>

- Mora-Cantalops, Marçal . 2018. *Análisis de Redes Sociales. Introducción y Teoría de Grafos*. España: Universidad de Alcalá.
- Wasserman, Stanley y Katherine Faust. 1994. *Social network analysis: Methods and applications*. Vol. 8. EUA: Cambridge University Press.

Reseña curricular

Patricia Garrido Villegas. Máster en *Data Science* por la Universidad de Alcalá de Henares, España. Desde enero de 2010 a la fecha (2020) se desempeña como editora de revistas electrónicas en XML, coordinadora de grupos de trabajo, así como especialista en reportes bibliométricos en la hemeroteca científica SciELO-México de la Dirección General de Bibliotecas-Universidad Nacional Autónoma de México.

