How do media mention research papers? Structural analysis of blogs and news networks using citation coupling

José Luis Ortega

Institute for Advanced Social Studies (IESA-CSIC), Córdoba, Spain, jortega@iesa.csic.es

Joint Research Unit Knowledge Transfer and Innovation, (UCO-CSIC), Córdoba, Spain

Abstract

This study aims to explore the connections between news and blogs based on the co-mention of research papers. This approach attempts to generate several network graphs that show the structural topology of blogs and news media, stressing possible differences between blogs and news when they come to co-mention research publications. 3,810 blogs and 3,387 news outlets that cite 100,529 research articles were displayed in network graphs using citation coupling. Country, language and thematic category were added to each medium. The findings show that the network of blogs and news is characterized by scale-free properties. The news network is highly centralized by general-interest news outlets from English-speaking countries, while the blogs network depicts a less centralized and low-density network, shaped by well-defined thematic clusters that rest on prestigious specialized hubs. The study concludes that these structures have important implications for the media impact of scholarly publications. In the case of news, the highly centralized model on general-interest news outlets acts as echo chambers amplifying the attention of publications. In the case of blog, the impact is less and would be borne by specialized blogs in specific thematic areas.

Keywords: blogs, news, altmetrics, citation coupling, social network analysis

1. Introduction

The study of the research impact in bibliometrics has been defined by the exclusive analysis of bibliographic citations. The times that a publication is mentioned in the subsequent literature has become the main proxy to value the impact of scientific publications, authors and organizations (Garfield, 1970). This approach implies a reflected view in which the object of study (cited publications) is also the subject of analysis (citing publications). It means that the environment in that a research output impacts is the same than the one that produced it. This closed surrounding causes that the reach of the bibliometrics is only circumscribed to the academic impact, the impact of scholarly publications in their own community (Aguinis et al., 2014). In this form, the research evaluation according to bibliometrics could be considered an internal evaluation of how the science perceives their own results (Ozanne et al., 2017).

However, the most important difference of altmetrics with regard to bibliometrics, is that the impact studied by altmetrics is produced by external subjects to the academic world (Bornmann, 2014). Social networks, patents, blogs or news are new subjects located in several knowledge spaces that mention research publications, expressing different types of impacts. In this sense, altmetrics quantitatively analyses how research outputs are impacting on non-scholarly environments. We can thus talk of social impact when research papers are discussed on Twitter or Facebook, technological impact if are cited in patents or media impact if mainstream media

talk about scientific findings. Taking this new framework into account, the current challenge of altmetrics is to understand the working of these different environments, the information flows into of these spaces and which are their principal structural characteristics. A previous knowledge about these environments could provide a better understanding of how different types of impacts are produced and what is the reach and meaning of these mentions according to different knowledge spheres.

One of these knowledge spheres is the media world, represented by blogs and news. This space is a fundamental instrument to the transfer of scientific results to the society, being clear intermediaries between academia and the general public. The mention of research articles by these mass media could be understood as an indicator of societal impact (Tahamtan & Bornmann, 2020). Therefore, the study and visualization of how these media co-mention research papers could better inform us on how this impact is generated. This study aims to fill this gap, exploring the connection of blogs and news according to the co-mention of research papers.

2. Literature review

Until now, few articles have centred in the analysis of these external environments and their role in the impact. Twitter has been the source that most attention has arisen due to it is the main platform for disseminating and discussing research results. Demographics and behaviours of Twitter users has been analysed to understand the impact of scientific results in the public opinion. Mohammadi et al. (2018) studied the accounts that mentioned research articles and they found that the almost the half of them did not belong to academia, demonstrating that the observed impact on Twitter transcends the academic sphere. Similar findings were achieved by Joubert and Costas (2019) when they studied South Africa science tweeters. Other studies have explored the role of the scholars on Twitter, discussing or disseminating scientific information. Holmberg and Thelwall (2014) investigated tweets from scientists in five disciplines and the results showed that researchers were more active than the average users. Other social networks such as Facebook (Ringelhan et al., 2015; Xia et al., 2016) and Reddit (McKnight, 2015) has been studied to understand their implications in the scientific impact.

However, studies about the structure and shape of the scientific blogosphere and their connections with the citation of research papers have been less prolific. Blanchard (2011), in one of the first studies about the blog phenomenon, suggests that the principal characteristic of the blogging is the interdisciplinary. Shema et al. (2012) performed the first descriptive analysis of the scientific blogs, analysing the aggregator ResearchBlogging.org and they found a higher presence of life and behavioral sciences. That same year, Fausto et al. (2012) analyzed the same source to uncover that bloggers are more interested in articles from high-impact journals. Shema et al. (2014) were the first ones to study the relationships between journal citations and blog mentions, finding positive correlations. Jamali and Alimohammadi (2015) explored the reasons of 300 blog posts for citing research papers and they stated that discussion and criticism are the main motives.

According to news, much less studies have addressed the mention of research articles on mass media. In this sense, some early works observed the connection between media and journal citations, detecting that the diffusion of research articles on newspapers influences the later citation (Phillips et al., 1991; Kiernan, 2003). Timilsina et al. (2017) analysed the connection between the bibliometric performance and the mention in news outlets and web blogs of scientists, finding a positive association between h-indexes and media mentions. However,

other studies have stressed on qualitative aspect to explain the mention on news. Papworth et al. (2015) explored the role of news in the dissemination of conservative research, their results showed that the probability of reporting on news depended on the journal title. Stryker (2002) found that the issuing of press releases influences in the later mention of papers in media, suggesting a bias toward journals with strong press offices. MacLaughlin et al. (2018) confirmed this last claim, founding that press releases are the feature that better predicts the mention in news outlets. From a network analysis view, Spitz and Gertz (2015) studied the news citation graph, finding similar properties to bibliometric citation networks, such as power laws and preferential attachment. Varlamis and Hilliard (2017) also used network analysis to detect the most influential news media. Lastly, Ortega (2020) studied the mentions of 100k research articles in news and blog media, finding that prevails English-speaking sites and general-interest media.

3. Objectives

This study aims to explore the connections between news and blogs based on the co-mention of research papers. This approach attempts to generate several network graphs that shows the structural topology of blogs and news media, stressing possible differences between blogs and news when they come to co-mention research publications. Several research questions are proposed:

- What are the main structural characteristics of these networks? What are the most central nodes in the network?
- Are there structural differences between both blogs and news networks?
- What are the implications of these networks for the media impact of scientific publications?

4. Methods

This study is focused on analysing the mass media environment that mentions and discusses research articles, as a way to understand the meaning and reach of these citations. Altmetric data providers are the best way to identify the media that cite research publications. Three of the most relevant services, Altmetric.com, PlumX and Crossref Event Data (CED), were selected to obtain information about citing blogs and news. Three different platforms were used because there are important differences in the coverage of media, therefore it is necessary to select more than one source in order to have the most complete picture (Ortega, 2019a).

4.1. Altmetric providers

PlumX: PlumX (plu.mx/plum/g/samples) is the main product of Plum Analytics, an initiative headed by Andrea Michalek and Michael Buschman in 2012. This provider of alternative metrics initially targeted scholarly organizations, offering a dashboard of altmetrics counts for private institutions at different aggregation levels. But since 2017, when Plum Analytics was acquired by Elsevier (www.elsevier.com), it captures the online footprint of any publication indexed in the Scopus database (Elsevier, 2017). PlumX is now the aggregator that provides the most metrics, including citation and usage metrics (i.e., Views and Downloads) and is the largest altmetrics aggregator, covering more than 52.6 million artifacts (Plum Analytics, 2020).

Altmetric.com: Altmetric.com (www.altmetric.com) was the first aggregator of alternative metrics. It was created by Euan Adie in 2011, with the support of Digital Science. Unlike PlumX,

Altmetric.com is centred on academic publishers, offering the monitoring of the altmetric impact of their scholarly publications. Altmetric.com provides a public Application Programming Interface (API) to extract altmetric counts. Today, close to nine million research outputs are tracked by this service (Altmetric.com, 2020a).

Crossref Event Data (CED): CED was created in 2016, and due to its youth is still in beta format (www.crossref.org/services/event-data). Unlike the other services, CED is supported by a non-for-profit organization and therefore it provides free access to their data through a public API. Another important difference with the previous services is that it does not provide counts, but only displays altmetric events associated to a Digital Object Identifier (DOI). For example, CED collects information about the discussion of a paper in a blog (date, media, link, etc.), but it does not count the number of mentions. For this reason, it is recommended to process CED data previous to any altmetric study.

4.2. Data extraction

During the second fortnight of August 2018, a sample of 100,529 DOIs were randomly extracted from Crossref API. This source was selected because it allows to extract random publications through the *sample* function. Another reason to select this source is that DOI is the most extended identifier and it is used by the altmetric providers to query documents. This allows an unequivocal identification of documents and an exact comparison between providers. This data set was limited to journal articles and publication date from 2012 (https://api.crossref.org/works?sample=100&filter=type:journal-article,from-pub-date:2012-01-01). The reason to limit the sample to articles published from 2012 was because the time window is sufficiently broad to capture the impact of these papers in blogs and news. Next, the sample was searched in the data providers, following specific strategies. In the case of Almetric.com, the Altmetric API (api.altmetric.com/v1/doi/) was used to obtain the Altmetric ID, and then this ID was used to scrape data about blogs and news titles, links and media directly from the web site (www.altmetric.com/details/). This is because the Altmetric API does not include these details about each blog and news mention. Data from PlumX were obtained from the web site of PlumX (plu.mx/plum/a/?doi=) using web scrapping. Lastly, CED API (query.eventdata.crossref.org/events?filter=obj-id:) was used to request information. In the three cases, several SQL-based scripts were written to extract the data from websites and APIs. In the event of a same mention was recorded by two different providers, one of them was discarded to avoid redundant information.

4.3. Media

3,810 blogs and 3,387 news outlet were identified as citation sources (Ortega, 2019a). The definition of each category, the collection process and final coverage are different according to each provider.

4.3.1. Blogs

From 2015, PlumX includes data about blog mentions (Parkhill, 2015). This information is obtained from an internal list. This list was extended from 4,000 sites to more than 10,000 blogs thanks to an agreement with ACI Scholarly Blog Index in 2016 (Parkhill, 2016). However, this company ceased trading and many links were broken (Ortega, 2019b). Altmetric.com has always curated an own list of roughly 15,000 blogs (Altmetric.com, 2020b), although this information is not publicly available. CED does not make a clear differentiation between blogs and news because it uses web domains to group sites. It thus defines three categories: wordpressdotcom,

web, and newsfeeds (Crossref, 2020). The first group corresponds to sites hosted on WordPress, independent if they are or are not blogs. The second groups only websites, which could include blogs or other type of webs. In addition, the reddit-links category includes links from Reddit that connect to external sources such as blogs and news.

4.3.2. News

When PlumX was acquired by Elsevier, this aggregator started to use Newsflo (an Elsevier company) as a news data provider. In this manner, PlumX has accessed to more than 55,000 different news outlets (Allen, 2017). Following a similar strategy, Altmetric.com signed an agreement with Moreover.com to be provided of mentions of articles on news media. This collaboration makes possible that Altmetric.com gathers mentions from more than 80,000 news outlets, in addition to the list of 1,300 news already curated by its own service (Williams, 2015). However, Lexis-Nexis acquired this company in 2014 and the collaboration ended, leaving 19% of the links inactive (Ortega, 2019b). Now, Altmetric.com monitor more than 5,000 news outlets (Altmetric.com, 2020b). New sources in CED are grouped in the newsfeed category, however this category revealed that media and blogs were equally included as web or newsfeed and sometimes in both categories. Due to this, the categorization of blogs and news was established according to the other altmetric providers. mentions were classified manually, in the event of a non-match.

Information on country and language of the blogs and news were extracted to identify possible associations. In addition, all the media were thematically categorized according to All Science Journal Classification Codes (ASJC). The subject classification and country and language codification was used according to Ortega (2020).

4.4. *Social Network Analysis*

The study of media environment and its meaning and reach in the mention of research publications is addressed from a structural perspective. Based on bibliographic coupling (Kessler, 1963), several network graphs were built from the frequency in which two media co-mention a research paper. This model assumes that two media that cite the same publications would mean that they are interested in similar research topics, suggesting that as more mentions they share, more close they are. Plotting these relationships in a network graph would allow to visualize the topology of the media network involved in scientific information and to characterize the main elements of this network. Several metrics from Social Network Analysis are used to describe the role of each node in the graph and the structural features of the networks:

- Degree centrality (k): It measures the number of lines incident to a node (Freeman, 1978). In this study, the degree centrality enables to measure the importance of a node in a graph and, concretely, to value the different number of media that share the mention of the same publications.
- Betweenness centrality (CB): It is defined as the ability of a node to mediate among nodes that are not directly linked between them (Freeman, 1980). This indicator helps us to observe the importance of media to mediate between distant clusters. High betweenness centrality would suggest that a news outlet or a blog shares the mention of papers with very different sites, being between different groups of media.
- Clustering coefficient (Ci): this measure indicates the likelihood of a node to establish a perfect cluster, where all their acquaintances are connected between them. It is computed as the proportion of observed triads by the possible ones, in which triads are complete interconnected clusters of three nodes. This indicator measures the

propensity of media to create close groups with other media. A high clustering coefficient means that media have a dense and interrelated network, while a low clustering coefficient reports a weak network of isolate media.

- Density (D): it is the percentage of the highest possible number of incident lines between nodes. It is used to measure the connection degree of a network. High rates of density show that a strong co-mention of research articles between media, while low values are symptoms of weak sharing of scientific news.
- Modularity: it is a way to identify clusters comparing the relative density of links inside communities with respect to ties outside communities. The Louvain method (Blondel et al., 2008) was selected because it outperforms identifying groups and improving time consumption.

Pajek 5.08 was used to manipulate the networks and Gephi 0.9.2 was used to visualize the networks and calculate the parameters. The resulting nodes and vertices tables are available through https://osf.io/fm5aq/

5. Results
5.1. Global network

This section displays the analysis and visualization of the total network of blogs and news, showing the most relevant media in the network and differences between discipines according to structural indicators.
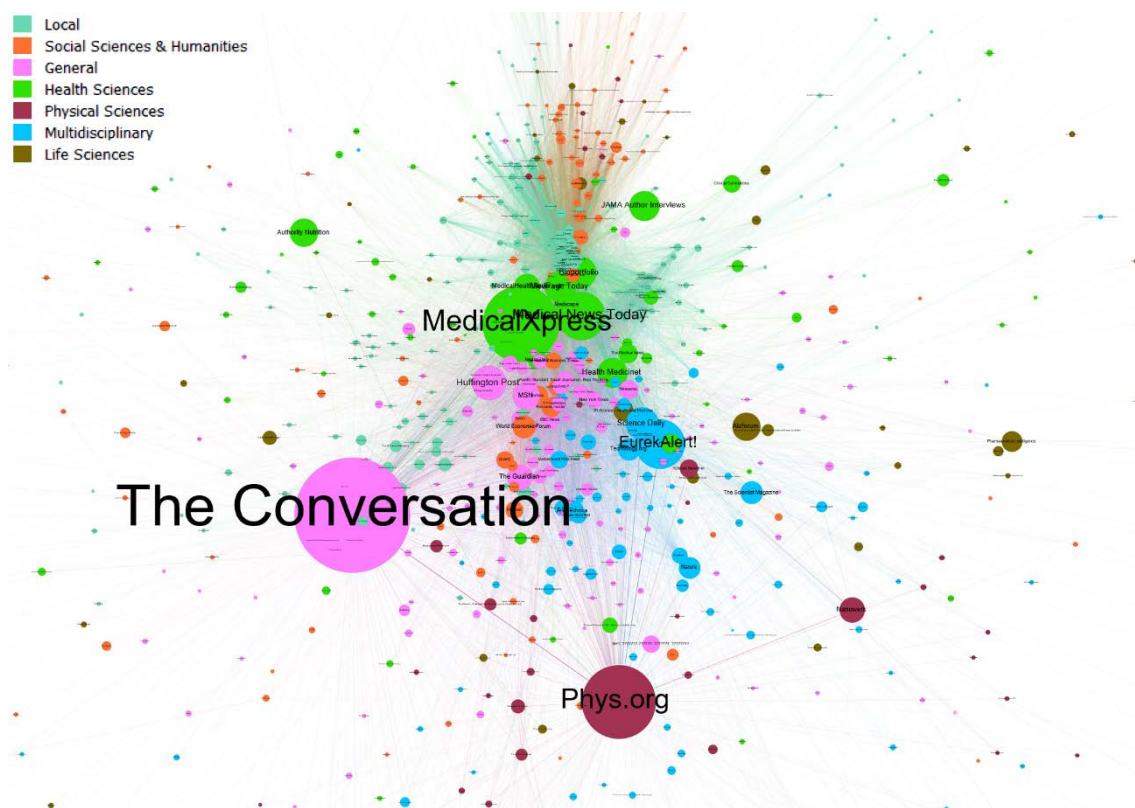

Figure 1. Citation coupling map of media (n=945, layout= Yifan Hu Multilevel)

Figure 1 shows the whole network. To improve the visualization, only media that cite more than five publications were selected. Yifan Hu layout was used because it is the most appropriate for large networks. Thus, 945 (34.2%) nodes and 27,555 arcs make up the principal component. Size

of the nodes is according to the number of mentioned publications and the colour represents the thematic classification of each medium in seven main subjects (Ortega, 2020). The graph presents a skewed degree distribution, being a clear symptom of a scale-free network (Spitz and Gertz, 2015). In these cases, a small fraction of nodes connects with most of the nodes, while the remaining ones just have one (44.2%) or two (25.1%) links.

Table 1. The ten most connected (degree) media

Medium	Mentioned articles	Type	Country	Language	Subject	Degree	Degree %
The Conversation	627	News	Australia	en	General	627	66.3%
EurekAlert!	261	News	United States	en	Multidisciplinary	487	51.5%
MedicalXpress	409	News	United Kingdom	en	Health Sciences	469	49.6%
Huffington Post	179	News	United States	en	General	395	41.8%
MSN	139	News	United States	en	General	391	41.4%
Bioportfolio	168	News	United Kingdom	en	Health Sciences	380	40.2%
Medical News Today	253	News	United Kingdom	en	Health Sciences	369	39.0%
Health Medicinet	153	News		en	Health Sciences	360	38.1%
Business Insider	95	News	United States	en	Social Sciences & Humanities	357	37.8%
The Guardian	128	News	United Kingdom	en	General	353	37.4%

Table 1 depicts the first ten media by the number of times (degree) in which one medium has co-mentioned a research article with other one. For instead, *The Conversation*, which has commented 627 articles, has done it along 627 media, 66.3% of the whole nodes of the sample. The fact that a medium shares the citation of a research article with a high proportion of media means that it is mentioning very popular articles with a high media impact. These media shape the core of the network and present similar characteristics. The first one is that the nodes with the highest degree are news outlets, which suggests that this type of media tends to co-mention more articles. The second one is that these media are in English language and come from English-speaking countries, mainly from the United States and the United Kingdom. This finding points out the hegemony of the English-speaking countries in the spreading of scientific information and the importance of this language to the dissemination of scientific results. From a thematic point of view, general-interest media and specialized media in health information are the nodes that share most articles.

Table 2. Number of nodes, mean degree and betweenness by thematic category for global network

Categories	Nodes	Nodes %	Mean Degree	Mean Betweenness

General	175	18.5%	62.23	1143.78
Health Sciences	103	10.9%	43.98	933.06
Life Sciences	51	5.4%	21.02	236.79
Local	255	27.0%	92.40	256.20
Multidisciplinary	113	12.0%	43.00	748.96
Physical Sciences	99	10.5%	15.82	500.58
Social Sciences & Humanities	143	15.1%	59.62	399.27

Table 2 confirms this last result, showing the average degree and betweenness of all the nodes grouped by thematic category. *Local* –media focused on the local everyday life– (27%) and *General* –media not limited to any scope– (18.5%) media are the groups with the greatest presence and the highest centrality degree (*Local* k=92.4, *General* k=62.23), whereas *Life Sciences* (k=21.02) and *Physical Sciences* (k=15.82) media has an outlying position in the network. The reason of this central position of general-interest media could be that they mention papers from different disciplines with a high social impact, therefore it is more likelihood to co-cite these articles along with more different media. On the contrary, specialized media tend to cite specialized articles for specific audiences, for which they are co-cited by a limited range of thematic media. However, it is interesting to notice the low mean betweenness centrality of *Local* media (CB=256.2). This means that this group of nodes are highly integrated among them, but they do not occupy a central position in the global network. A possible reason could be that these media are co-citing similar papers (high degree), but different ones to those mentioned by the core of the network (low betweenness).

5.2. Blogs

The global network has showed that blogs have a secondary role in the citation of research papers. This section attempts to visualize and analyse the particular characteristics of the citation coupling network of blogs.
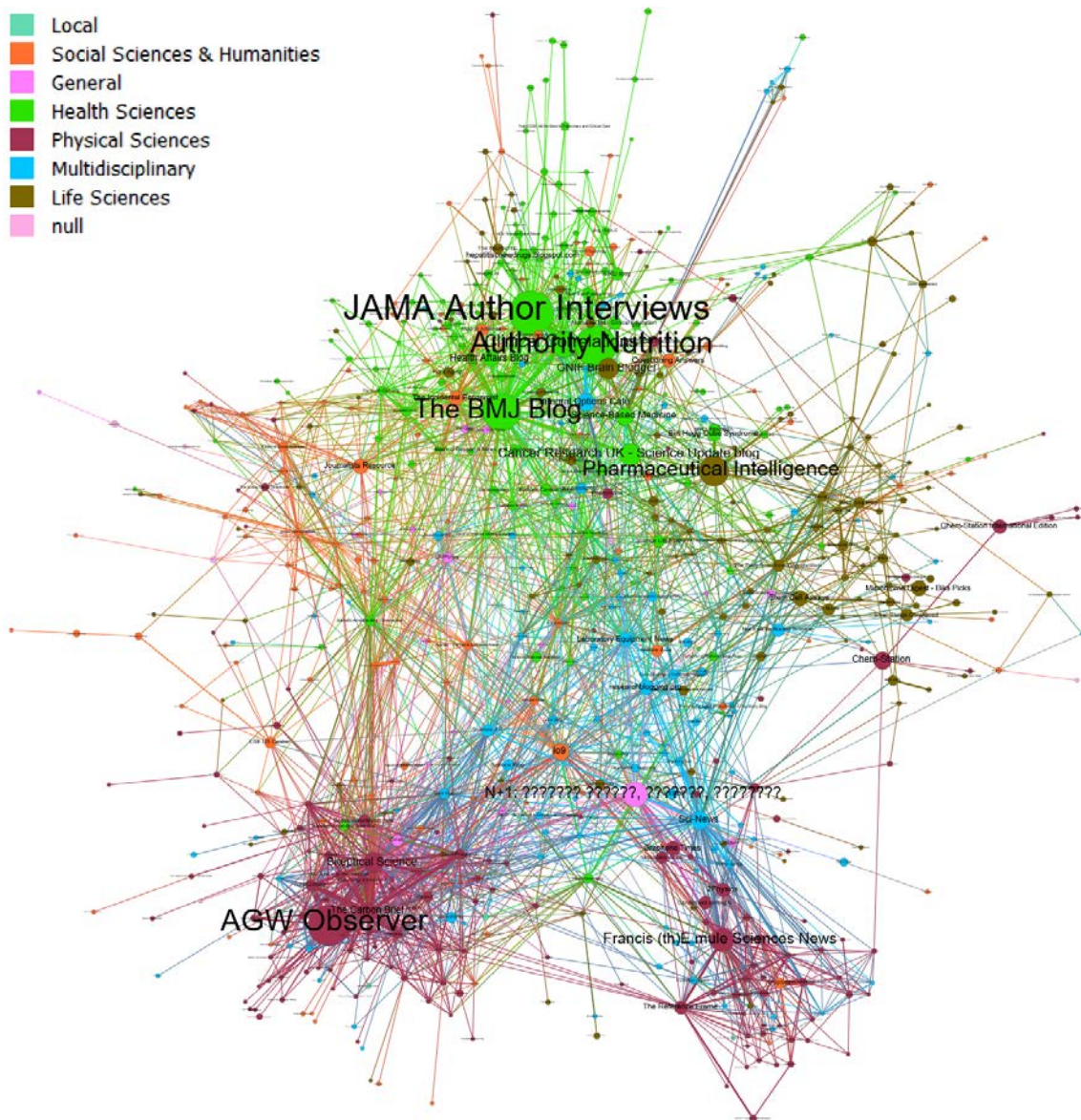
Figure 2. Citation coupling map of blogs (n=666, layout= ForceAtlas2)

Figure 2 shows the full blogosphere network based on citation coupling of research papers. ForceAtlas2 layout was use to emphasize the thematic clusters. In this case, the network includes 666 (24.1%) nodes and 2,786 arcs, with a much lower density (D=0.013) than the global network (D=0.062). This means that the blog network is sparser and more unconnected than the news network and this would explain why blogs do not have a central role in the global network.

Table 3. The ten most connected (centrality degree) blogs

Blog	Mentioned articles	Country	Language	Subject	Degree	Degree %
The BMJ Blog	129	United Kingdom	En	Health Sciences	101	15.2%
Skeptical Science	64	Australia	En	Physical Sciences	83	12.5%
The Carbon Brief	53	United Kingdom	En	Physical Sciences	77	11.6%
Sci-News	56	United States	En	Multidisciplinary	72	10.8%
JAMA Author Interviews	156	United States	En	Health Sciences	62	9.3%

Altmetric.com	17	United Kingdom	En	Social Sciences & Humanities		61	9.2%
Grist	30	United States	En	General		59	8.9%
io9	52	United States	En	Social Sciences & Humanities		59	8.9%
Scientific American Blog: Observations	25	United States	En	Health Sciences		59	8.9%
Guest Work	26	New Zealand	En	Multidisciplinary		58	8.7%

Table 3 presents the first ten blogs by centrality degree. As the global network, the most central media are in English and come from English-speaking countries, mainly from the United States and the United Kingdom, which confirms the importance of this language and countries in the spreading and discussion of research articles (Shema et al., 2012; Fausto et al., 2012). However, the most connected blog (*The BMJ Blog*) hardly share their mentions with more than 15% of the nodes of the blog network, which proves again that this network is barely connected. Perhaps, the reason of this low cohesion would be that the thematic classification of the principal nodes is very varied, with highly connected blogs from *Health Sciences*, *Physical Sciences*, *Social Sciences & Humanities* and *Multidisciplinary*. This suggests that there is not a disciplinary core, rather but the blog network could be set up by disciplinary sub-networks.

Table 4. Distribution of thematic categories according to clusters (algorithm=Louvain method)

Module	General	Health Sciences	Local	Life Sciences	Multidisciplinary	Physical Sciences	Social Sciences & Humanities
0	20.0%	62.94%	12.5%	13.64%	10.68%	2.00%	17.65%
1	10.0%	0.70%	12.5%	6.06%	25.24%	34.00%	8.40%
2	30.0%	10.49%	12.5%	15.91%	13.59%	2.67%	5.04%
3	15.0%	2.80%	62.5%	6.82%	22.33%	45.33%	20.17%
4	0.0%	6.99%	0.0%	50.76%	21.36%	7.33%	10.92%
5	25.0%	16.08%	0.0%	6.82%	6.80%	8.67%	37.82%

Table 4 shows the thematic classification of blogs and percentages in each module. This allows to observe to what extent the modules are set up by disciplinary criteria. Specialized categories such as *Health Sciences*, *Life Sciences* and *Physical Sciences* depict high proportions of blogs in particular modules. Thus, 62.9% of the *Health Sciences* blogs are clustered in the module 0, 50.7% of *Life Sciences* in the module 4, and *Physical Sciences* are split in two groups: module 1 (34%) on *Astrophysics and Mathematics* and module 3 (45.33%) about *Environmental Sciences*. Otherwise, *General* and *Multidisciplinary* are more uniformly distributed among the modules. This finding demonstrates that the blog network is shaped by thematic sub-network and suggests that non specialized categories act as bridges connecting cliques.

Table 5. Number of nodes, mean degree, betweenness and clustering by thematic category for blogs network

Subject	Nodes	Nodes %	Mean degree	Mean betweeness	Mean clustering
General	20	3.0%	11.45	1436.80	0.33
Health Sciences	125	18.8%	9.26	1072.20	0.38

Life Sciences	132	19.8%	6.15	499.94	0.36
Multidisciplinary	103	15.5%	9.01	896.08	0.42
Physical Sciences	150	22.5%	9.85	693.45	0.49
Social Sciences & Humanities	119	17.9%	7.67	806.91	0.35

This differentiate role of disciplines in the configuration of blogs network, could be better appreciated in Table 5. It depicts the average of the structural parameters (centrality degree, betweenness and clustering coefficient) for each thematic category in the blogs network. *Local* category was removed because it contributes only three blogs and their metrics were not representative. *Physical Sciences* (22.5%) and *Life Sciences* (19.8%) are the thematic areas with most blogs in the network, while *General* media have a little presence (3%). This result differs from the global network (Table 1), in which *Local* (27%) and *General* (18.5%) media have a highlighted presence, which confirms that scientific blogosphere is dominated by specialized media. However, and despite of this, *General* media still have a central position with the highest centrality degree (k=11.45) and betweenness (CB=1436.8), while its mean clustering coefficient is the lowest one (0.33). These parameters come to verify that general interest blogs act as intermediaries (high centrality degree and betweenness) between groups, but they do not form cohesive groups (low clustering).

6. Discussion

The structural analysis of the citation coupling network of media has brought to the forefront some interesting patterns about the mention of research articles in blogs and news. The first significant result is that news outlets occupy the core of the network, being the type of media that co-mention more research documents. Perhaps, this central position could be due to there are more general-interest news than generalist blogs. Results have showed that this type of media have the highest centrality indicators (degree and betweenness) both in the global and the blogs networks. The reason could be that these media tend to cite the same research articles, that is, publications with a great social impact (medical advances, astronomy discoveries, funny studies, etc.). This fact increases the density of the network and the emergence of great hubs (*The Conversation*, *Huffington Post*, *MSN*) typical of scale-free networks (Spitz and Gertz, 2015). Mukerjee et al. (2018) confirmed this fact, showing that the core of the news network in the United Kingdom and the United States is set up by important general-interest media.

Other important result is that the core of this network is exclusively formed by English language media from English-speaking countries. The 85% of media written in English fits with the 86% of Shema et al. (2012) and the 82% of Fausto et al. (2012) about blogs. This result could mean that media from these countries could be more interested in scientific news, with greater mention of research publications (Schmidt et al., 2013). On the other hand, the fact that more than the 80% of the scientific literature is written in English (Weijen, 2012) could reinforce that English-speaking media cite more research articles than media written in other languages.

It is also worth to mention the particular position of *Local* media. The results have showed that this category has in average a high centrality degree but a low betweenness, and the graph (Figure 1) depicts a dense cluster formed by these media but isolated from the central nucleus. The reason of this lack of intermediate power could be that many of these media come from the United States and it is possible that they are citing specific articles with a higher local impact. For instead, within this group there are some newspapers with a clear political sign (*Belleville News-Democrat*, *The Tribune-Democrat*) that could cite specific research papers that support

their arguments. Another possible explanation could be that these local media are part of media conglomerates (*Fox*, *ABC*, *CBS*), which promote the sharing of news among the media group.

According to the sub-group of blogs, the main difference is that the presence of generalist media is fewer, which causes that there is not a strong core. Instead, this graph is dominated by specialized media that configure a network based on specialized clusters around local hubs. Shema et al. (2012) and Fausto et al. (2012) also observed the majority presence of specialized blogs, with a significant contribution of biology and life science blogs. In our case, *Physical Sciences* surpasses *Life Sciences* in number of blogs, due, perhaps, these findings are not limited to only one source (*ResearchBlogging.org*). This structure defines the scientific blogosphere as a specialist and decentralized environment in which a small number of media act as thematic hubs discussing and spreading specific research results, creating isolates communities.

These particular configurations of the network of blogs and news could inform us about the media impact of research publications. The majority presence of news and their central role in the global network thus favors that research articles have more mentions from news than from blogs (Ortega, 2019a). Even more, the fact that the global network shows a skewed degree distribution with a dense core of generalist media, makes possible great differences in the citation of articles. The high connection of general-interest media in the core could act an echo chamber augmenting the impact of articles, causing that a small portion of articles gain a disproportionate impact, while the rest of articles that are not mentioned by this core, barely receive a few of mentions. This great unbalance in the news mentions was noticed by previous altmetric studies (Thelwall et al., 2013; Costas et al., 2015). This importance of generalist media in the impact is also important from a disciplinary view because there are subjects (medicine, astrophysics) with a higher exposure to media (Bucchi and Mazzolini, 2003; Clark and Illman, 2006). This fact could foster their scientific impact because the media coverage of research articles is associated with more frequent citations (Kiernan, 2003; Manisha and Mahesh, 2015). In the case of the blogs network, this phenomenon could be less significant and the key agent to improve the impact of research articles would be specialized blogs with a high prestige inside their community.

7. Conclusions

This study has been the first approach to draw the underlying structure of blogs and news media when they cite research publications, with the aim of knowing how they could influence the media impact of science. The topological differences between the blogosphere and news media, the role of generalist media in the amplification of the impact, and the importance of the specialized websites uniting the scientific blog network are results that provide new insight on the gestation of the media impact.

From the obtained results, three main conclusions can be derived. The network of blogs and news is characterized by scale-free properties (skewed degree distributions, high density). This network is highly centralized by general-interest news outlets from English-speaking countries. Local media present particular characteristics, shaping groups distant from the core. This suggests that local media, perhaps, cite certain publications of local nature or because they are part of specific media conglomerates (*Fox*, *ABC*, *CBS*).

However, blogs network significantly differs from the news network. In this case, blogs network depicts a less centralized and low-density network, shape by well-defined thematic clusters that rest on prestigious specialized hubs. This absence of generalists blogs influences the low density and the weak connections.

The configuration of the news and blogs network has important implications for the media impact. In the case of news, the highly centralized model on general-interest news outlets could act as echo chambers amplifying the attention of publications. This suggests that the impact of a publication could be favored when it is mentioned by important generalist English-speaking media, and this could be significant for disciplines with a higher media coverage (Astronomy, Medicine). In the case of blogs, the impact is less and would be borne by specialized blogs in specific thematic areas.

These results have important implications for several stakeholders. The mapping of web media that comment research outputs allows the researchers to obtain a better understanding about how the altmetric impact is generated, which sources constitute the core of the media environment and what thematic areas attract more the attention of news outlets and blogs. These findings would support popularization strategies by research organizations, detecting what type of media are more suitable for communicating their discoveries. For the other side, journalists and bloggers could verify in which context are located their media and what is their role in the dissemination of academic publications. This information would contribute to adopt editorial approaches that improve the commercial competitiveness of media houses. Policy makers could understand the complex environment that produces the media impact, observing the knowledge transfer mechanism from the scholarly world to the public opinion. In general, this study provides valuable findings on the structure of the media environment that cites research publications, the topological differences between blogs and news and how these characteristics influence the altmetric impact.

8. References

Aguinis, H., Shapiro, D. L., Antonacopoulou, E. P., & Cummings, T. G. (2014). Scholarly impact: A pluralist conceptualization. *Academy of Management Learning & Education*, *13*(4), 623-639.

Allen, P. (2017). World-Wide News Coverage in PlumX. Retrieved June 5, 2020 from: https://plumanalytics.com/world-wide-news-coverage-plumx/

Altmetric.com (2020a). How it works. Retrieved June 5, 2020 from: https://www.altmetric.com/about-our-data/how-it-works/

Altmetric.com (2020b). What outputs and sources does Altmetric track? Altmetric.com Support. Retrieved June 5, 2020 from: https://help.altmetric.com/support/solutions/articles/6000060968-what-outputs-and-sources-does-altmetric-track-

Blanchard, A. (2011). Science blogs in research and popularization of science: Why, how and for whom. *Common knowledge: The challenge of transdisciplinarity*, PFL Press, 219-232.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, *2008*(10), P10008.

Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of informetrics*, *8*(4), 895-903.

Bucchi, M., & Mazzolini, R. G. (2003). Big science, little news: science coverage in the Italian daily press, 1946-1997. *Public understanding of science*, *12*(1), 7-24.

Clark, F., & Illman, D. L. (2006). A longitudinal study of the New York Times Science Times section. *Science Communication*, *27*(4), 496-513.

Costas, R., Zahedi, Z., & Wouters, P. (2015). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, *66*(10), 2003-2019.

Crossref (2020). Crossref Event Data User Guide. Retrieved June 5, 2020 from: https://www.eventdata.crossref.org/guide/

Elsevier. (2017). Elsevier Acquires Leading 'Altmetrics' Provider Plum Analytics. Retrieved June 5, 2020 from: https://www.elsevier.com/about/press-releases/corporate/elsevier-acquires-leading-altmetrics-provider-plum-analytics

Fausto, S., Machado, F. A., Bento, L. F. J., Iamarino, A., & Nahas, T. R. (2012) Research Blogging: Indexing and Registering the Change in Science 2.0. *PLoS ONE, 7*(12), e50109.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks*, *1*(3), 215-239.

Freeman, L. C. (1980). The gatekeeper, pair-dependency and structural centrality. *Quality and Quantity*, *14*(4), 585-592.

Garfield, E. (1970). Citation indexing for studying science. *Nature*, *227*(5259), 669-671.

Holmberg, K., & Thelwall, M. (2014). Disciplinary differences in Twitter scholarly communication. *Scientometrics*, *101*(2), 1027-1042.

Jamali, H. R., & Alimohammadi, D. (2015). Blog citations as indicators of the societal impact of research: Content analysis of social sciences blogs. *International Journal of Knowledge Content Development & Technology*, *5*(1), 15-32.

Joubert, M., & Costas, R. (2019). Getting to Know Science Tweeters: A Pilot Analysis of South African Twitter Users Tweeting about Research Articles. *Journal of Altmetrics*, *2*(1), 2

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American documentation*, *14*(1), 10-25.

Kiernan, V. (2003). Diffusion of news about research. *Science communication*, *25*(1), 3-13.

McKnight, J. (2015). *"The new reddit journal of science": Public evaluation and understanding of scientific information based on source factors in social media* (Doctoral dissertation, University of Delaware).

MacLaughlin, A., Wihbey, J., & Smith, D. A. (2018). Predicting news coverage of scientific articles. In *Twelfth International AAAI Conference on Web and Social Media*.

Manisha, M., & Mahesh, G. (2015). Citation pattern of newsworthy research articles. *J. Sci. Res.*, *4*(1), 42-45.

Mohammadi, E., Thelwall, M., Kwasny, M., & Holmes, K. L. (2018). Academic information on Twitter: A user survey. *PloS one*, *13*(5).

Mukerjee, S., Majó-Vázquez, S., & González-Bailón, S. (2018). Networks of audience overlap in the consumption of digital news. *Journal of Communication*, *68*(1), 26-50.

Ortega, J. L. (2019a), The coverage of blogs and news in the three major altmetric data providers, In *17th International Conference of the International Society for Scientometrics and Informetrics*, Rome, Italy https://osf.io/pd27h/

Ortega, J. L. (2019b). Availability and Audit of Links in Altmetric Data Providers: Link Checking of Blogs and News in Altmetric.com, Crossref Event Data and PlumX. Journal of Altmetrics, 2(1), 4

Ortega, J. L. (2020), Blogs and news sources coverage in altmetrics data providers: a comparative analysis by country, language, and subject, *Scientometrics*, 122(1), 555-572

Ozanne, J. L., Davis, B., Murray, J. B., Grier, S., Benmecheddal, A., Downey, H., & Seregina, A. (2017). Assessing the societal impact of research: The relational engagement approach. *Journal of Public Policy & Marketing*, *36*(1), 1-14.

Papworth, S. K., Nghiem, T. P. L., Chimalakonda, D., Posa, M. R. C., Wijedasa, L. S., Bickford, D., & Carrasco, L. R. (2015). Quantifying the role of online news in linking conservation research to Facebook and Twitter. *Conservation Biology*, *29*(3), 825-833.

Parkhill, M. (2015). PlumX Expands "Mentions" Category by Tracking Blogs. Plum Analytics. Retrieved May 20, 2019 from: https://plumanalytics.com/plumx-expands-mentions-category-by-tracking-blogs/

Parkhill, M. (2016). Research Impact Tracking from Scholarly Blog Posts Expands with Partnership Between Plum™ Analytics and ACI Information Group. Plum Analytics. Retrieved May 20, 2019 from: https://plumanalytics.com/press/research-impact-tracking-scholarly-blog-posts-expands-partnership-plum-analytics-aci-information-group/

Phillips, D. P., Kanter, E. J., Bednarczyk, B., & Tastad, P. L. (1991). Importance of the lay press in the transmission of medical knowledge to the scientific community. *The New England journal of medicine*, *325*(16), 1180-1183.

Plum Analytics (2020). Coverage: Expanding the World of Altmetrics. Retrieved June 5, 2020 from: https://plumanalytics.com/learn/about-metrics/coverage/

Ringelhan, S., Wollersheim, J., & Welpe, I. M. (2015). I like, I cite? Do Facebook likes predict the impact of scientific work?. *PloS ONE*, *10*(8).

Schmidt, A., Ivanova, A., & Schäfer, M. S. (2013). Media attention for climate change around the world: A comparative analysis of newspaper coverage in 27 countries. *Global Environmental Change*, *23*(5), 1233-1248.

Shema, H., Bar-Ilan, J., & Thelwall, M. (2012). Research blogs and the discussion of scholarly information. *PloS ONE*, *7*(5).

Shema, H., Bar-Ilan, J., & Thelwall, M. (2014). Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics. *Journal of the Association for information science and technology*, *65*(5), 1018-1027.

Spitz, A., & Gertz, M. (2015). Breaking the news: Extracting the sparse citation network backbone of online news articles. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 274-279). IEEE.

Stryker, J. E. (2002). Reporting medical information: Effects of press releases and newsworthiness on medical journal articles' visibility in the news media. *Preventive medicine*, *35*(5), 519-530.

Tahamtan, I., & Bornmann, L. (2020). Altmetrics and societal impact measurements: Match or mismatch? A literature review. *El profesional de la información (EPI)*, *29*(1), e290102.

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PloS one*, *8*(5), e64841.

Timilsina, M., Khawaja, W., Davis, B., Taylor, M., & Hayes, C. (2017). Social impact assessment of scientist from mainstream news and weblogs. *Social Network Analysis and Mining*, *7*(1), 48.

Varlamis, I., & Hilliard, D. F. (2017). Finding influential sources and breaking news in news media using graph analysis techniques. *International Journal of Web Engineering and Technology*, *12*(2), 143-164.

Weijen, D. (2012). The Language of (Future) Scientific Communication. Research Trends, 31

Williams, C. (2015) Altmetric to provide expanded mainstream media tracking. Altmetric.com blog. Retrieved June 5, 2020 from: https://www.altmetric.com/blog/moreover/

Xia, F., Su, X., Wang, W., Zhang, C., Ning, Z., & Lee, I. (2016). Bibliographic analysis of nature based on Twitter and Facebook altmetrics data. *PloS ONE*, *11*(12).