

Collapsing Ivory Towers? A hyperlink analysis of the German academic blogosphere

 blogs.lse.ac.uk/impactofsocialsciences/2015/09/29/collapsing-ivory-towers-a-hyperlink-analysis/

9/29/2015

In a substantial analysis of over 500 German-speaking science blogs, [Jonas Kaiser](#) and [Benedikt Fecher](#) look at what hyperlinks are used within prominent science blogs to investigate how scientists link to each other and outside sources. Using visualisation and mapping software, their results show how science blogs form new networks beyond traditional disciplines and interact with the wider general blogosphere.



In times of increasing specialization of academic disciplines (a 2012 [report from the STM Association](#) estimates that over 28,000 journals publish over 1.8 million articles each year), academia faces a communication challenge. It has to uphold its forums for excellent research (mirrored in high ranked journals) while at the same time mediating between (sometimes even closely related) disciplines and justifying itself towards the public. The digitization of communication has a reinforcing effect on this communication challenge when it comes to the accessibility and comprehensibility of publicly funded research.

In the wealth of new communication tools that academia gradually begins to [explore](#), blogging appears to be particularly promising for its capacity to communicate scientific content considerably [quick and in a tentative manner to an often wider audience](#); at least in comparison to a traditional journal application. It is a powerful tool for mediating between disciplines and for engaging in a meaningful dialogue with different societal target groups. If these assumptions of the academic blogosphere hold true, it should constitute a network that is less specialized and more open towards alternative actors than the one represented in the academic journal system. However, some scholars suggested that these options may, in turn, also [lead to a balkanization of science](#).

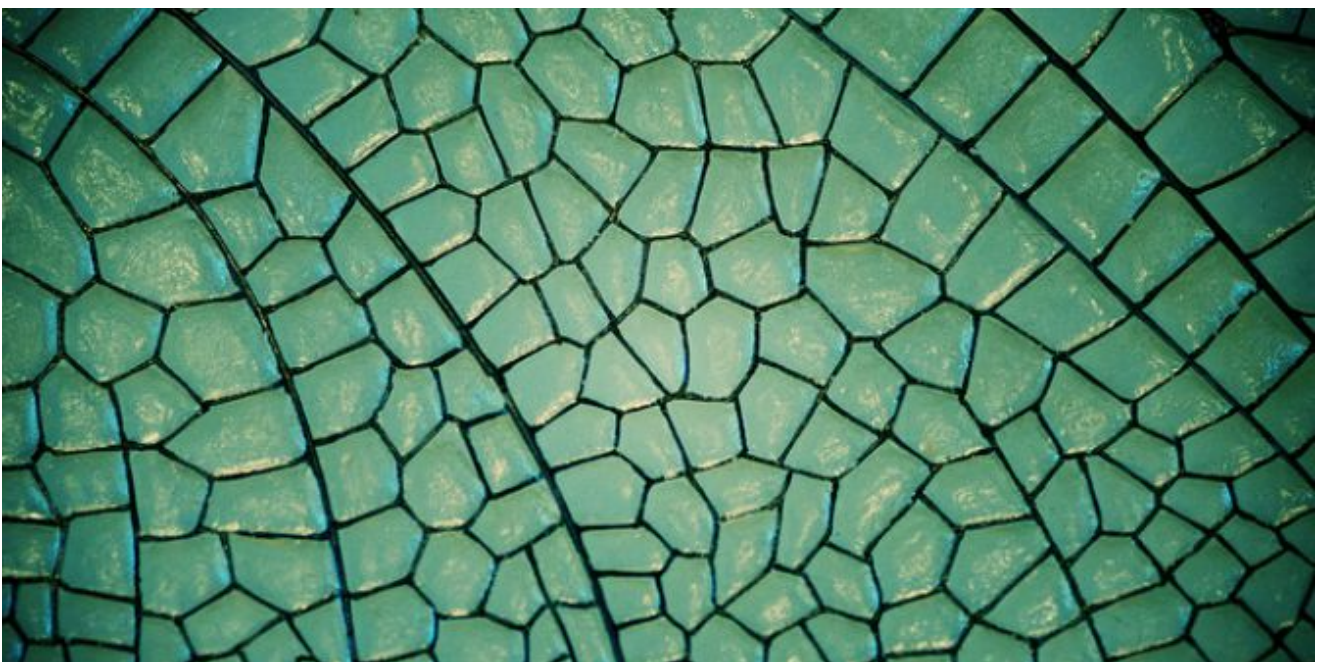


Image credit: [David Goehring](#) (Flickr [CC BY 2.0](#))

Here we investigate the communication network of science blogs and focus on the German “science blogosphere” as our case. The approach we take is based on a [growing body of research](#) which takes hyperlinks to be indicative of social, thematic, and political relationships between online actors (read more on the approach [here](#)). We look at how science blogs link to each other and outside sources in order to get an understanding of their network and clusters. The questions that we are interested in are if and how science blogs form new clusters (e.g., less discipline-specific and more trans-disciplinary) and involve new actors (e.g., “non-scientists”). In how these blogs relate to each other through their hyperlinks we aim to develop a deeper understanding of the public(s) they address and the discourse they engage in.

Our analysis of German speaking science blogs shows that

1. Science blogs are not necessarily clustered around issues or (academic) background but rather form trans-disciplinary clusters which are more likely to be influenced by the hosting platforms.
2. Scientists and non-scientists link to each other and even though scientists are more popular within the scientific blogosphere, they are also part of the bigger “general” blogosphere
3. This means that scientists see blogging as a conscious act and are aware of their surroundings and try not to replicate the “ivory tower” 2.0 but rather connect with others.

The sample

To define a sample for our analysis, we identified five prominent (based on a Google search) German science blog websites and scraped these sites for the relevant blog URLs. Two of the five sites function as general blog directories ([bloggerei.de](#) and [labs.teads.tv/top-blogs](#)) and three are, at least in the German academic blogosphere, well-known platforms that host science blogs themselves ([scienceblogs.de](#), [de.hypotheses.org](#), and [scilogs.de](#)). We tested the sample with random double-checks and assume that the sample covers the great bulk of the German scientific blogs. After deleting blogs that have been named by more than one site we identified $n = 507$ German-speaking science blogs in total (or better “unique links”) with which we conducted the analyses. We chose German speaking blogs due to a) the network’s relative insularity which gives us the opportunity to focus on the resulting clusters more closely than a bigger sample would have and b) because we want to understand our own academic blogosphere better (we are both German).

For those interested: The list of blogs in our sample as well as the other data we used can be found [here](#).

Analysis

To understand the relations between the blogs, we conducted an interlink-analysis and a co-link analysis. In this context, we understand websites as nodes in our network and the hyperlinks connecting the nodes as edges. An *interlink analysis* focuses on the relationship between a set of actors (here blogs) within a defined sample. It gives an indication about clusters within the given sample. For example: Do science blogs form disciplinary or issue clusters or are they trans-disciplinary? Which other kind of clusters do they form? A *co-link-analysis* considers not only the links between the URLs within a set of websites but more importantly it identifies common out-links, i.e. URLs that at least two of the blogs in our sample link to. This helps providing a broader picture of the network’s “neighborhood” as now not only blogs but also other kinds of websites are included. With that analysis we can questions such as: Do science blogs rather link to traditional news outlets, scientific journals and institutions or do they prefer news sources from the blogosphere? ([for more information on interlink and co-link-analysis](#))

We copied the list of websites into the [Issuecrawler](#), a free network mapping software provided by the Govcom.org Foundation, and defined them as the starting points (standard settings). From there the Issuecrawler captured outgoing links, followed them and searched for ties between these sites. As both, an analysis and data visualization tool we used [Gephi](#), an open source software for visualizing networks. It has to be noted, however, that the data transfer from Issuecrawler to Gephi was a bit problematic in the sense that even though the crawler differentiates

between sites (top level domains) and pages ('deep pages' within a site) the further analysis in Gephi is limited due to technological (or user error based) constraints. This led to the reduction of our 507 starting points to 371 since all links to and from Scilogs.de and Scienceblogs.de were aggregated into one node each and could not be attributed to separate blogs. (Side note: if anyone has an idea on how to get the page data into Gephi so that we can also work with the data, we're all ears.)

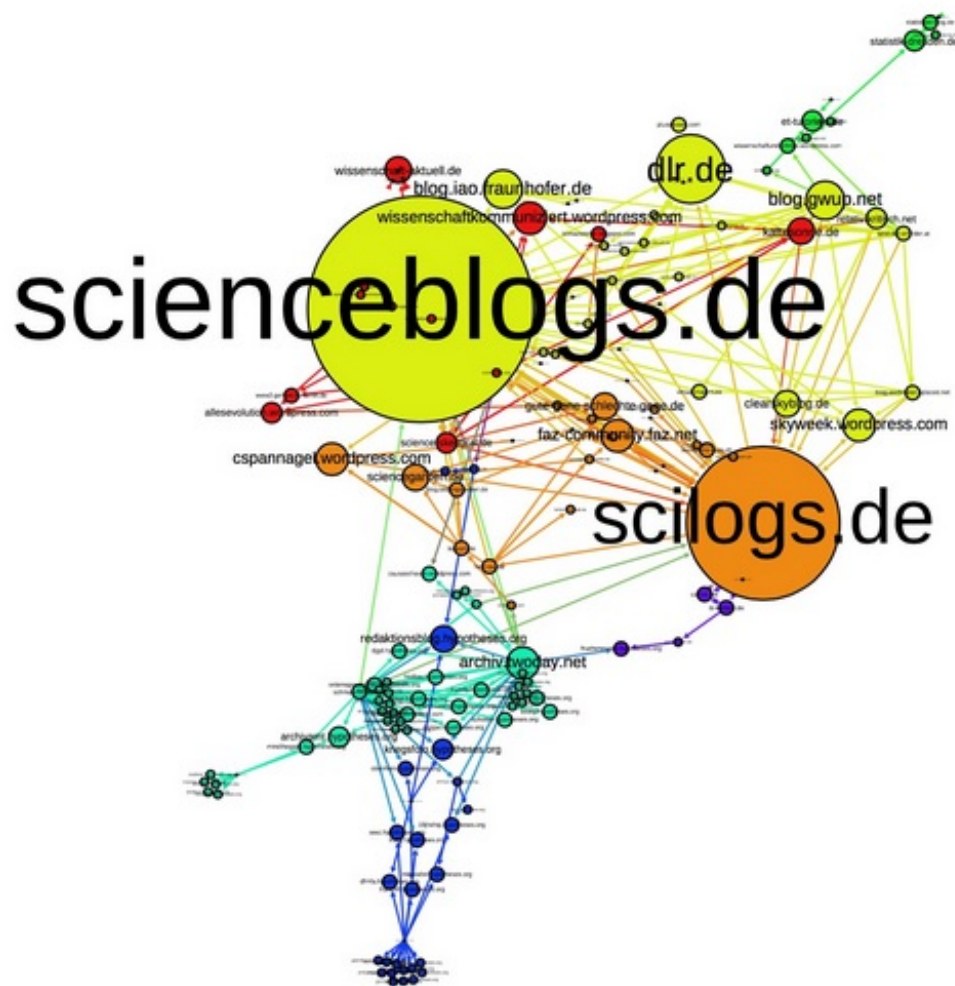
Results

Interlink analysis

When looking for interlinking connections between actors the underlying assumption is of course that the actors do link to each other. However, from the 371 URLs we analyzed, only 166 (44,74%) link to another page within our sample. That means that more than half of the blogs in our sample are not in the "science blog network". This may have several reasons, for example that they only link to other scientific sources that are not in our list (e.g., scientific papers or institutions), that they are not relevant for the sites within our sample, or that they just do not link at all.

Unsurprisingly, scienceblogs.de and scilogs.de are the actors with the highest in-degree (see fig. 1) since these have no URL that the Issuecrawler can identify as unique site (for example both the climate blog [Klimalounge](#) and the history/archeology blog [AbenteuerGeschichte](#) are represented by the yellow Scilogs.de-node). This, unfortunately, allows no clear picture regarding the relevance of single blogs hosted by scienceblogs.de and scilogs.de. It does however highlight the two platforms' importance for hosting science blogs (on par with [hypotheses.org](#) that has unique links for the blogs it hosts). Scienceblogs.de alone has 37 in-links, scilogs.de received 25. In comparison: The third most prominent site, [dlr.de](#) (the blog from the German Aerospace Center), received 11 in-links.

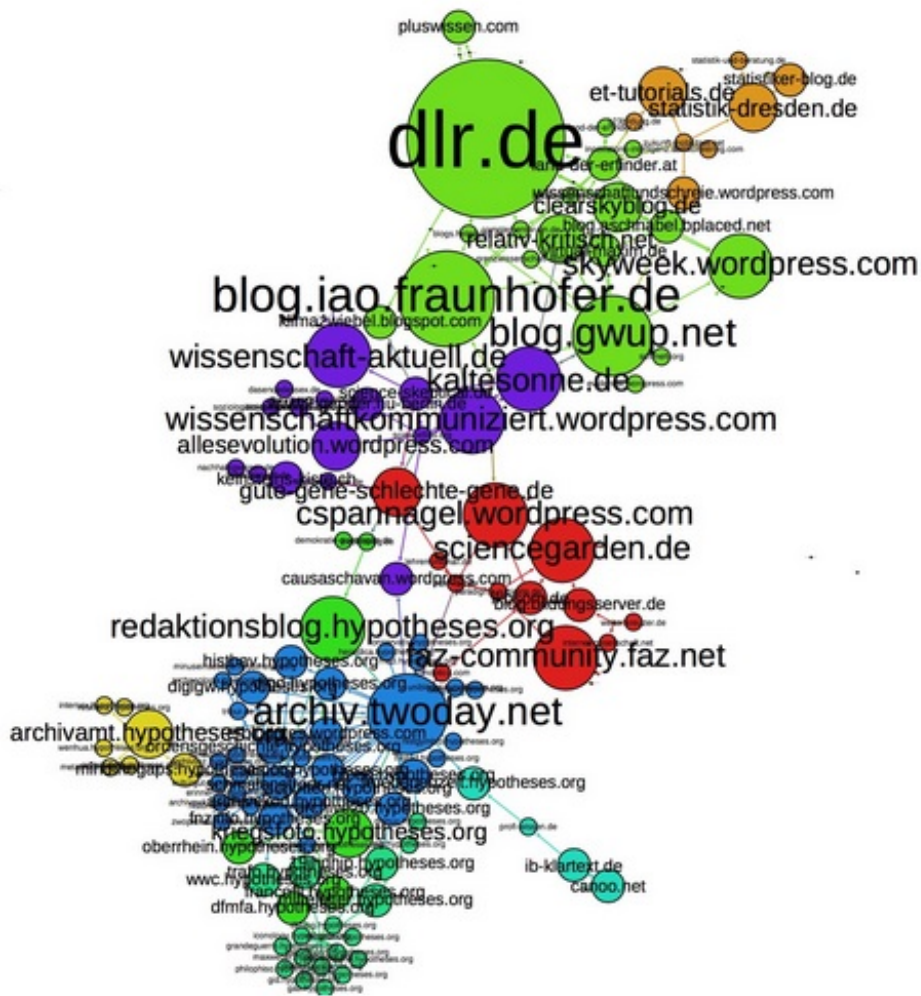
Figure 1: Inter-linking network (node size according to in-degree)



In contrast to the latter two, Hypotheses.org puts the blog names in front of its domain (e.g., soziologieblog.hypotheses.org vs. scilogs.de/klimalounge). This allows us to see how the blogs hosted by hypotheses are linking to each other. Hypotheses blogs mostly link within their own platform. This can be explained in two ways, first through the platform itself and second due to thematic similarities (many of the hosted blogs on hypotheses cover similar issues, e.g., history and archiving). However some hypotheses blogs appear in other clusters. For example: The gender blog dasendedessex.blogspot.de links to soziologieblog.hypotheses.org and forms a small sociology cluster with a few other blogs within the bigger purple cluster.

In order to take a closer look at the network that is not dominated by the two big “players” Scilogs and Scienceblogs we decided to delete these two sites from our network in order to get a better grasp of other clusters (see fig. 2).

Figure 2: Interlinking network without Scilogs.de and Scienceblogs.de (node size according to in-degree)



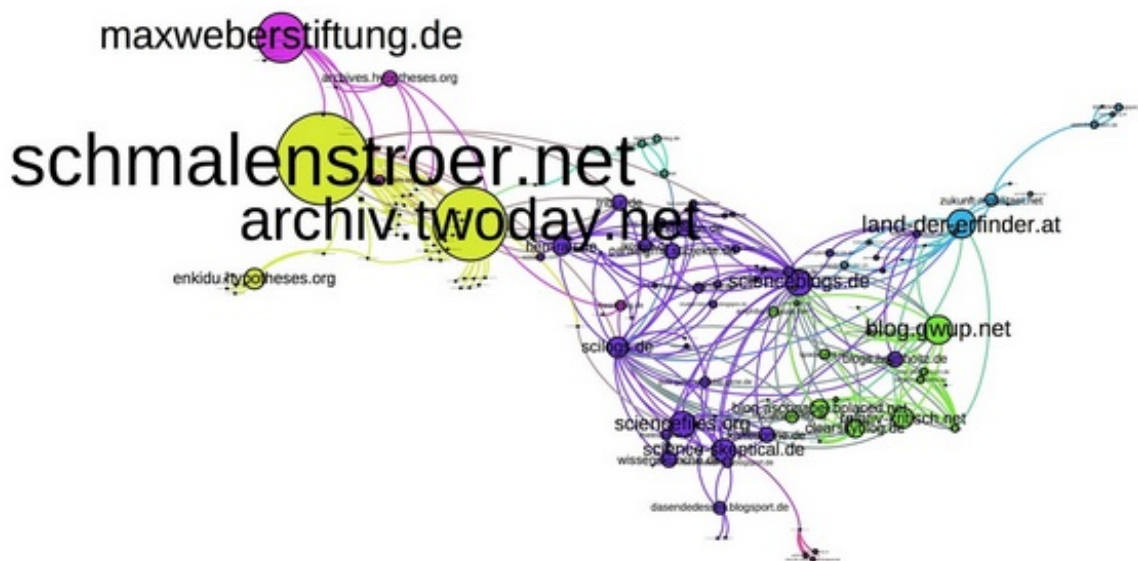
We identified 6 major clusters that covered roughly 80% of all the nodes within our new network. We then examined the top 5 blogs in each cluster with the highest in-degree and coded manually the topic they cover and the actors responsible for the content (the table can be downloaded from our [figshare page](#)). We did this in order to get an idea if a) the clusters have a common issue at “heart” and b) if the blog writers have an academic or non-academic background.

We identified two issue clusters for history/archive blogs, most of them hosted by hypotheses.org (blue and dark green). However, in one of these two clusters, the most prominent blog is archiv.twoday.net which is not hosted by blogosphere and which focuses on old historic archives. Another cluster (green) is tied to the topic of astronomy, however not exclusively. For example: Three of the top URLs deal with astronomy topics, another one ([gwup.net](#)) with the criticism of “pseudoscience” (e.g. homeopathy) and the last one with issues related to labor economics ([iao.fraunhofer.de](#)). When taking a closer look into the cluster it becomes apparent that the mix of skeptical blogs regarding pseudoscience as well as astronomic blogs creates a sort of “STEM” cluster. Thematic issues cannot explain the other top clusters or scientific disciplines alone and even further analyses don’t show more similarities (e.g. climate skeptic blogs are in one cluster with gender blogs and a blog which focuses on science communication). An interesting descriptive result concerns the content creators: Within the 30 blogs we coded, almost 40% (12) are written from non-scientists. Especially the green and purple clusters in fig. 2 have mixed content creators. The two hypotheses.org-clusters are mainly written by scientists.

However, we were not only interested in the websites’ relevance with regards to in-degree but also which sites link the most and especially who is responsible for these sites (see fig. 3). Our analysis shows that the most linked to actors are quite different to the actors which link the most. This is both true when we are looking at the clusters as

well as the content creators. Whereas the network in Fig. 1 was “dominated” by Scilogs and Scienceblogs we now see that both sites mostly are linked to and link little themselves (7 resp. 9 out-links). As most active *link sponsors* we see a mixture of issues as well as actors that are united by one similarity: the top five sites with the highest out-degree are actively linking more than they are being linked to (for the five combined: 15 inlinks, 99 outlinks). When it comes to content creators there is no clear visible trend: within the top 5 there is a scientific actor, two from the civil society (one civilian and one foundation), one entrepreneur and one from the media. There is also no overarching issue which would “unite” these sites. So even though the results remain inconclusive we are able to say that despite scientific actors being more prominent link targets than link sponsors these roles are not exclusive and we can also additionally say that they do also link to blogs which are not written by scientists and are occasionally even dubious (e.g. scilogs.de linking to the climate skeptic site skeptical-science.de).

Figure 3: Interlinking network (node size according to out-degree)



Co-link-analysis

In order to fully analyze the co-link-network we first had to delete the social media sites Twitter and Facebook due to their all-outshining standing within the network (mostly due to plug-ins and self-cross-promotion). Nonetheless we can conclude that both are relevant for scientific bloggers and are possible ways of self-promotion but also follow-up communication. After having deleted these two, we were able to identify four clusters within the network that are not related to scientific topics (purple, blue, pink, turquoise; fig. 3). The two most visible clusters (purple and blue) show the prominence of blog directories within our sample. Indeed, three of the five websites with the highest in-degree are blog directories (bloggerei.de, bloggeramt.de, topblogs.de). For comparison: Bloggerei.de alone has 109 inlinks, Wikipedia.de has 6. Most of the directories’ are linked to by science blogs with little to no relevance (in-degree) within the network and which were not part of the interlinking network (e.g., the philosophical blog irrwegen.info has no in-link and only one common out-link with other sites from our sample). This leads us to conclude that these sites possibly link to directories for SEO-considerations or a form of appreciation.

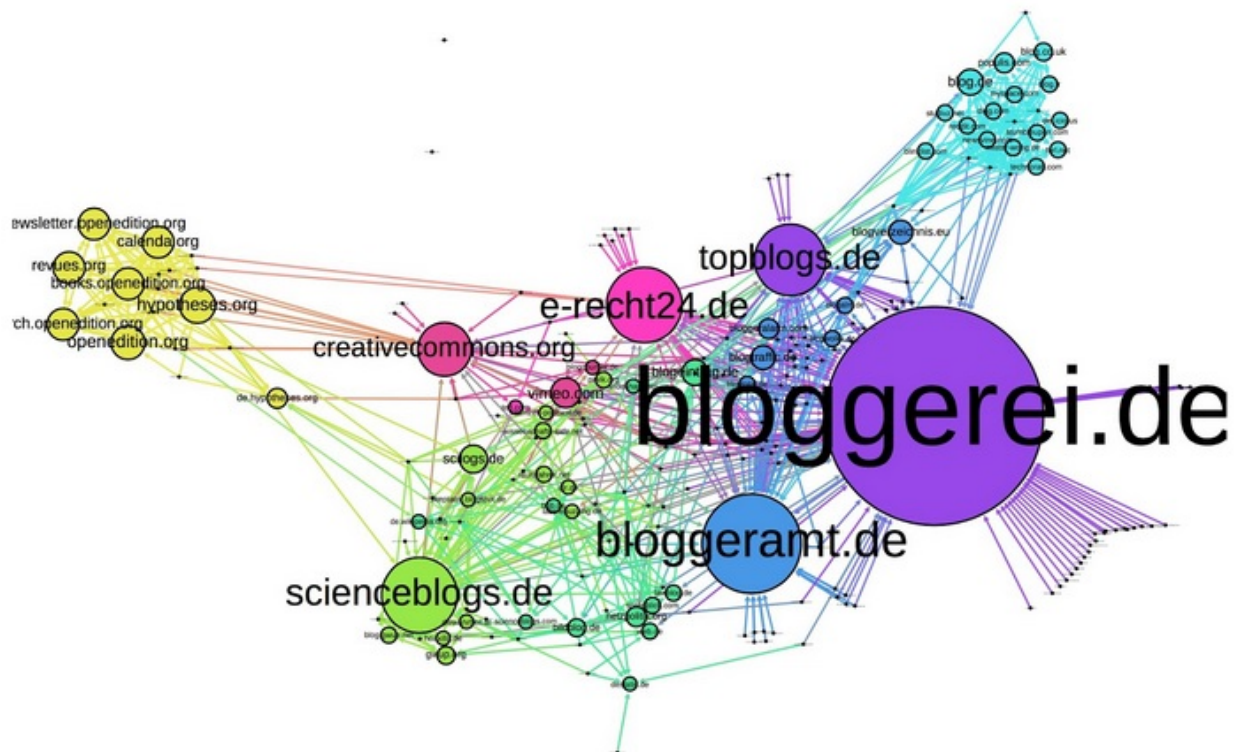
When taking a closer look at the clusters that include the prominent scientific blogs from our initial sample (like Scilogs, Scienceblogs or Gwup) we can see that these are closely connected to a cluster which is formed by several popular German blogs. For example, they are linked to bildblog.de (a media watchblog), netzpolitik.org (an independent Internet policy news outlet), spreeblick.de (an entertainment and culture blog) or lawblog.de (a blog on German law). In this context, it is important to note that traditional news outlets (e.g., opinion leading magazines and

newspaper), the websites of scientific institutions, archives or journals are practically non-existent in the network.

In a next step we coded the five most in-linked sites of the top six clusters (these covered ca. 70% percent of all the nodes in the sample). In contrast to the interlink-analysis the clusters were thematically more diverse. There was, for example, only one clear science cluster (light green). As stated above the science cluster is closely connected to the “blog prominence” cluster (pastel green). Also, the hypotheses.org cluster (yellow) “transcended” its scientific blogs and even though these are still present, the most linked to nodes are part of the organizational hypotheses network and not scientific by nature (openedition.org, etc.).

When it comes to the content creators, only 13% of the 30 sites could be attributed to scientists (we also coded Scilogs and Scienceblogs as scientists due to the aggregation issue mentioned earlier). In comparison the same amount can be traced back to citizens, 20% to journalists and 40% to businesses. Even though this, of course, is not representative of the clusters and is to be expected when we look at the network’s neighborhood, it is still interesting to note that the network is not getting more “sciency” but rather more general.

Figure 4: Co-Link Network (node size according to in-degree)



Conclusion

When taking both analyses into account there are some results which need to be emphasized:

1. The interlink-analysis shows that there are both, thematic and trans-disciplinary clusters in the network of science blogs. The fact that clusters are not necessarily disciplinary can be regarded as sign for a more diverse and less specified communication network than the one represented in journal publications.
2. A considerable part of science blogs are written by non-scientists which can be regarded as a sign that the academic blogosphere is more inclusive than traditional journal publications.
3. However, blogs which are written by scientists are more often linked to than those written by citizens.

4. The example of hypotheses.org suggests that blogs that are hosted by the same platform rather link to each other than to outside sites (this, of course, is also due to the platform's innerlinkings which the blog creators have little to no influence on). It should be noted, however, that this could also be explained due to hypotheses.org's issue focus on humanities and cultural sciences.
5. One interesting result of the co-link-analysis is that we are able to identify the "blog prominence" cluster that is in the immediate neighborhood of the science cluster. We interpret this as the corrosion of the ivory tower's "walls" in favor of a more open, egalitarian and especially interactive way of communicating.
6. What is even more interesting, however, is the fact that traditional news media (e.g., online presence of newspapers) and scientific journals are practically non-existent. The presence of thematic irrelevant but well-known blogs suggests that science blogs, at least partly, address a new online target audience and rather than turning away they turn towards society.

Even though this brief analysis focuses on German science blogs and has several shortcomings (small sample, no closer look at the edge relations, methodological issues with/and regarding the Issuecrawler, and no clearer coding of the sites) we are able to show that German science blogs are embedded within and contribute to the more general German blogosphere and are not isolated into specific issue clusters and thus we can reject the fear of [science's balkanization](#). We are also able to show that most of the prominent German blogs are based on platforms that specifically host scientific blogs. However, most of the science blogs in our starting list only play a marginal role within both networks, thus suggesting that either the bloggers don't link very much or that they link to other sites which we were not able to display. In order to show a more complete and international network of science blogs we suggest replicating this analysis on a larger scale and with a focus on the ivory tower's crumbling "walls".

Note: This article gives the views of the author, and not the position of the Impact of Social Science blog, nor of the London School of Economics. Please review our [Comments Policy](#) if you have any concerns on posting a comment below.

About the Authors

Jonas Kaiser is a doctoral researcher at Zeppelin University Friedrichshafen and a Fellow at Humboldt Institute for Internet and Society. His research focus lies on the expression of unpopular opinions in the networked public sphere.

Benedikt Fecher of the German Institute for Economic Research (DIW Berlin), is a doctoral researcher at the Humboldt Institute for Internet and Society. The focus of his dissertation is the participation in Open Science Projects.

- Copyright © The Author (or The Authors) - Unless otherwise stated, this work is licensed under a Creative Commons Attribution Unported 3.0 License.