PLoS one

Characterizing Interdisciplinarity of Researchers and Research Topics Using Web Search Engines

Hiroki Sayama[1,2]*, **Jin Akaishi**[1,3]

1 Collective Dynamics of Complex Systems Research Group, Binghamton University, Binghamton, New York, United States of America, 2 Departments of Bioengineering & Systems Science and Industrial Engineering, Binghamton University, Binghamton, New York, United States of America, 3 Faculty of Liberal Studies, Kumamoto National College of Technology, Yatsushiro, Kumamoto, Japan

Abstract

Researchers' networks have been subject to active modeling and analysis. Earlier literature mostly focused on citation or co-authorship networks reconstructed from annotated scientific publication databases, which have several limitations. Recently, general-purpose web search engines have also been utilized to collect information about social networks. Here we reconstructed, using web search engines, a network representing the relatedness of researchers to their peers as well as to various research topics. Relatedness between researchers and research topics was characterized by *visibility boost*—increase of a researcher's visibility by focusing on a particular topic. It was observed that researchers who had high visibility boosts by the same research topic tended to be close to each other in their network. We calculated correlations between visibility boosts by research topics and researchers' interdisciplinarity at the individual level (diversity of topics related to the researcher) and at the social level (his/her centrality in the researchers' network). We found that visibility boosts by certain research topics were positively correlated with researchers' individual-level interdisciplinarity despite their negative correlations with the general popularity of researchers. It was also found that visibility boosts by network-related topics had positive correlations with researchers' social-level interdisciplinarity. Research topics' correlations with researchers' individual- and social-level interdisciplinarities were found to be nearly independent from each other. These findings suggest that the notion of "interdisciplinarity" of a researcher should be understood as a multi-dimensional concept that should be evaluated using multiple assessment means.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: sayama@binghamton.edu

Introduction

The structural and dynamical properties of networks among researchers have been an important research subject for the last several decades [1–15]. Price originally proposed the key idea of preferential attachment and the resulting scale-free degree distributions for networks of scientific publications [1], which is now widely applied and utilized in various kinds of scientific fields [16–18]. Typical data sources for such studies on networks of scientific communities are domain-specific electronic paper or citation archives, such as arXiv.org for physics, DBLP for computer science, and SSRN for social sciences, from which co-authorship or citation networks can be created. Price's predictions have been confirmed in those data, such as scale-free degree distributions and the network growth over time based on the preferential attachment principles [4,5,8].

While earlier studies mostly focused on citation or collaboration networks within a particular domain, there is a growing body of literature on the characterization and measurement of interdisciplinarity of scientific journals and researchers [19–22]. These recent studies used cross-disciplinary citation indexing services, such as ISI Web of Knowledge, and analyzed how multiple disciplines are connected by publications and researchers. Interdisciplinarity has been characterized in several different ways,

e.g., how many different disciplines were represented in the references cited in a single paper, how many different disciplines an individual researcher publishes his/her work in, and so on [19,22].

However, the existing citation indexing services have several limitations. One apparent limitation is the lack of flexibility in their disciplinary classification. Established disciplinary classification structures, such as those used in ISI Journal Citation Reports, are based on traditional notions of scientific disciplines, which may not be up-to-date for capturing emerging fields of cutting-edge research where the characterization of interdisciplinarity is most needed. Also, it is commonly assumed in the citation indexing services that each journal belongs to just a few disciplines (mostly just one), which is not necessarily a valid assumption when analyzing properties of highly interdisciplinary publications. Another limitation is that their indexing coverage may not include non-mainstream journals, conference proceedings, and other online archives, which are often more important in particular disciplines (e.g., in computer science and physics). Finally, the data of the citation indexing services are only available on a subscription basis, which would be hard to obtain for researchers whose institutions do not have a subscription to those services.

Given those limitations of the citation indexing services mentioned above, researchers have recently started to utilize

more general-purpose web search engines as an alternative data source for researchers' network reconstruction [10,14]. For example, Lee et al. [14] defined "Google correlation" (i.e., number of hits obtained by a Google search query for names of two persons) and used it to reconstruct social networks of physicists and politicians. They also evaluated the validity of this data collection method by showing that the reconstructed network by Google searches was indeed correlated with the real social network [14].

There are several significant advantages in this new web search engine-based method over the conventional methods. First, it can exploit a massive amount of information about various forms of association between two researchers (or any pair of keywords) that are collectively produced and maintained by people from all over the world. While web search-based data is less structured and more noisy than data collected from the online archives and citation indexing services, the information created through massive "collective intelligence" is often quite informative and useful [23,24]. Second, the use of general-purpose web search engines liberates researchers from the existing disciplinary classification structures, giving them full flexibility in choosing any relevant set of disciplinary keywords to study. Third, many non-indexed sources of information can be included in the analysis, such as conference proceedings and online archives. And finally, the data is open and accessible to everyone, with no subscription required. These advantages are quite suitable and beneficial for studying interdisciplinarity of researchers, but to our knowledge, there is no study reported yet on the use of such web search engines for that purpose.

Here we conducted a preliminary study on the interdisciplinarity of individual researchers and a variety of research topics using a web search engine-based data collection method. We searched the web for information about connections between individual researchers *as well as* researchers and research topics, each quantified by the number of hits obtained through a search query for two keywords (names of two researchers, or a name of a researcher and a research topic). One of the novelties of our work compared to earlier literature is that we developed a unique measurement called *visibility boost*, defined as an increase of a researcher's visibility brought by focusing on a particular research topic. We propose this as a more meaningful way of quantifying the relatedness between the researcher and the research topic than simply using the number of web search hits for those two keywords. We also characterized each individual researcher's interdisciplinarity by measuring the diversity of research topics related to him/her as well as his/her centrality in the researchers' network. The former measurement represents individual-level interdisciplinarity, i.e., how diverse the research topics the researcher is associated with, while the latter captures social-level interdisciplinarity, i.e., how important the researcher is in connecting other researchers.

Methods

Web search engine-based data collection methods require a list of keywords to be searched for. In our study, we created the following two separate lists. One is a list of names of 1,000 researchers. This list was compiled by having student volunteers manually collect about 4,000 names from four annual international conference websites for years 2006–2009, and then selecting the top 1,000 significant names based on their numbers of independent web search hits. The four selected conferences were all interdisciplinary ones the authors were already personally familiar with. Although our prior familiarity with the conferences

could be a source of potential biases, it was necessary in order for us to be able to manually check and correct mistakes in the raw data collected by student volunteers. It was also our hope that the interdisciplinary nature of these conferences would allow us to create a representation of broader research communities within limited time and labor available.

The other list of keywords is a list of research topics. One could use the traditional categories established in scientometrics literature (e.g., ISI Journal Citation Reports categories) for this purpose. We did not take that option, however, because the relevant research topics discussed in those four subject conferences did not quite fit into the ISI JCR categories, and also because one of our objectives was to demonstrate the flexibility in keyword selection. Therefore, the list of research topics were collected again from the same set of websites by the student volunteers, and then manually edited and compiled by the authors. In so doing, we paid attention to maintaining a good balance among different disciplines. Specifically, we set four major categories and made sure that the numbers of keywords were similar across those categories. As a result, we had 13 words for biological and medical sciences, 9 words for physical sciences, 9 words for engineering and robotics, and 9 words for general terms (40 research topics in total). Because there was an overall emphasis on biological systems among the four conferences we used, there were slightly more words in the first category. The actual list of research topics can be found in Appendix S1.

In total, we compiled 1,000 (researchers' names) +40 (research topics) = 1,040 keywords to use. The total number of search queries was (1,040 choose 2)−(40 choose 2) = 539,500. The subtraction of "40 choose 2" was because we did not need to measure the relationships between research topics. We recognize that our specific choices of those researchers' names and research topics may have significantly influenced the results described below, which will be discussed in more detail later.

The overview of the network to be reconstructed using these keywords is illustrated in Fig. 1. The data about the relatedness between researchers and research topics (Fig. 1, left) were used to characterize each researcher's unique research areas as well as his/her individual-level interdisciplinarity. The data about the relatedness among researchers (Fig. 1, right) were used to characterize the researcher's social-level interdisciplinarity. Our primary goal is to illustrate how the proposed method works in identifying possible relationships between these two characterizations.

We implemented a computer program in Java for repeated searches using Google Web Search API [25]. Google Web Search API allows one to write a program that can send a search query directly to Google web search engines and then receive search results (with some limitations). Earlier work also used Google for social network data collection [10,14,26–29]. In every single search query we conducted in this study, we always included an additional word "research" in order to narrow search results to those related to scientific research (although this was by no means a perfect filtering technique), following a similar technique used by Lee et al. [14]. To improve the reliability of search results, each query was searched three times at different times in a day, and their average values were used for analysis (see Appendix S1 for details). We inserted sufficient amount of waiting time between queries in order to avoid overloading the search engine. Therefore the actual data collection took place rather slowly over several weeks in June and July 2010.

In characterizing researchers' relatedness with particular research topics, we had to address the following technical problem: More common words tended to result in more search hits
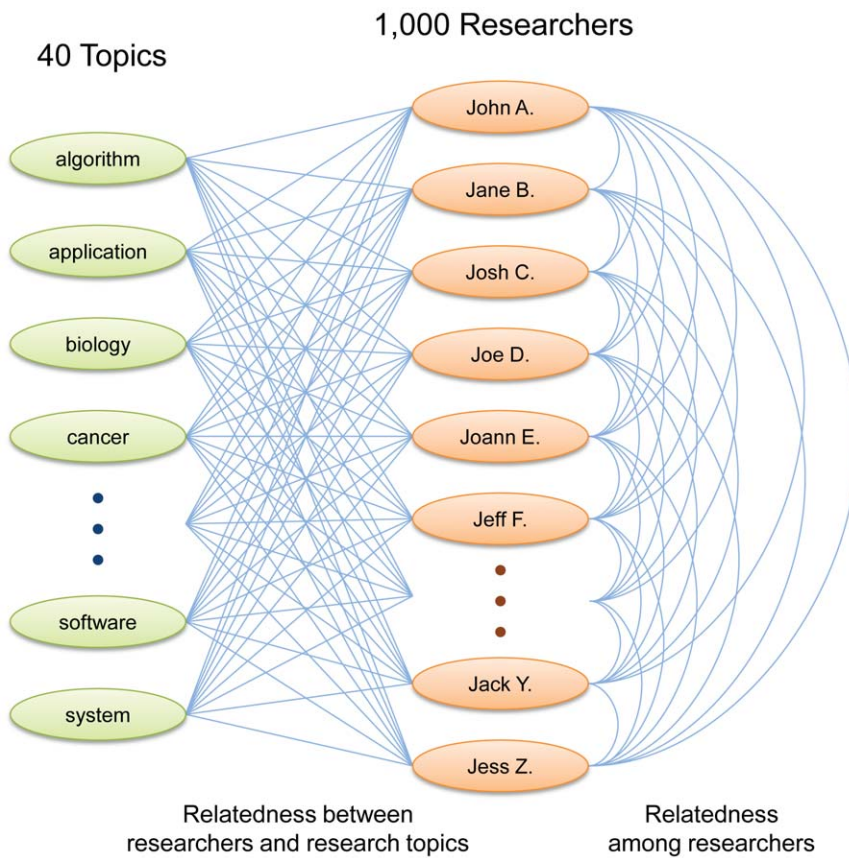
Figure 1. An overview of the network that consists of 1,000 researchers and 40 research topics reconstructed in this study. Each link is weighted by the average number of web search hits for a search query of (research topic) + (researcher's name) + "research" (left), or (researcher's name 1) + (researcher's name 2) + "research" (right).
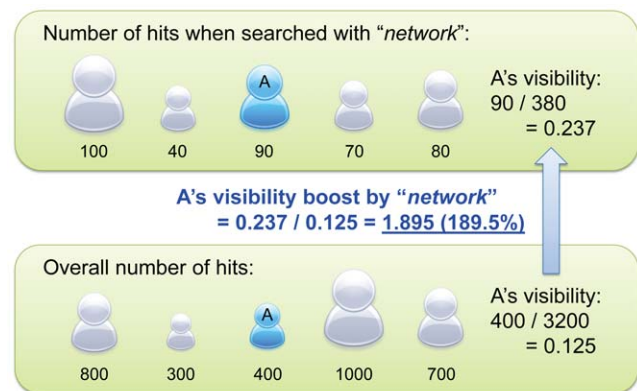doi:10.1371/journal.pone.0038747.g001

regardless of a researcher's actual research domain. For example, the word "biology" is more commonly used than the word "network", and therefore a search query "John Doe" + "biology" + "research" can produce more search hits than "John Doe" + "network" + "research", even if John Doe's research domain is network science and not biology. This means that one cannot simply use the absolute number of search hits for characterizing unique research areas and interdisciplinarity of individual researchers.

We solved this problem by introducing a new quantity, named *visibility boost (VB)*, defined as

$$VB(r,t) = \frac{h(r,t)}{\sum_i h(i,t)} \bigg/ \frac{\sum_j h(r,j)}{\sum_{i,j} h(i,j)} = \frac{h(r,t) \sum_{i,j} h(i,j)}{\sum_i h(i,t) \sum_j h(r,j)}, \quad (1)$$

where r is the researcher, t the research topic, and $h(r, t)$ the number of search hits for search query "r" (researcher's name) + "t" (research topic) + "research". This formula mathematically describes how much change occurs to the visibility of researcher r (i.e., ratio between r's own hits and the total hits over all researchers) by limiting the focus to research topic t (Fig. 2). A visibility boost greater than (or less than) 1 means that researcher r is more (or less) associated with research topic t on the web. A similar link weight normalization method was also proposed by Lee et al. [14], though their method produces values that are influenced significantly by frequencies of two keywords searched for. In contrast, our visibility boost gives a more consistent,

intuitive measure of association. Specifically, $VB = 1$ always means a neutral level of association between two keywords, while such a constant reference value for neutrality does not exist in Lee et al.'s method. This property allows one to use visibility boost values comparatively for multiple different topics.

Figure 2. An example of "visibility boost" calculation. This figure shows how to calculate the visibility boost by research topic "network" for the researcher A in the middle (blue).
doi:10.1371/journal.pone.0038747.g002

Results

Figure 3A shows the network of 1,000 researchers reconstructed from the data obtained above (which corresponds to the researchers' network illustrated on the right in Fig. 1). Nodes and undirected links represent researchers and their relatedness, respectively, where the average numbers of search hits were used as link weights. Figure 3B shows a complementary cumulative distribution of total link weights of nodes, which does not follow a power-law but still shows a remarkably long tail even with this small data set. Note that the following analysis also used a bipartite network made of connections between 1,000 researchers and 40 research topics (illustrated on the left in Fig. 1), which is not visualized here.

To evaluate the utility of the proposed visibility boost, we counted how many unique researchers would be ranked within top 20%, at least once, according to their visibility boosts by any of the 40 research topics. We then conducted the same counting task using the raw search hit counts instead of visibility boosts. In addition, as a control, we also counted how many unique researchers would be selected, at least once, if 20% of researchers were purely randomly sampled 40 times. The results are summarized in Fig. 4A. Based on the raw search hit counts (yellow), only less than 60% of the researchers had a chance to be ranked within top 20%. This indicates that relying on raw search hit counts would cause unwanted concentration of analysis on fewer researchers with greater general popularity. In contrast, using the visibility boosts for the same task (red) resulted in nearly every researcher having a chance to be ranked within top 20% for some topic, which is comparable to the random sampling case that showed perfect coverage (blue). This result demonstrates that the proposed visibility boost measure is useful in extracting information about individual researchers' unique specialties, without being dominated by general popularity differences among them.

We found, both visually and statistically, that researchers who had high visibility boosts by the same research topic tended to aggregate in their relatedness network. Figure 4B presents statistical evidence supporting this observation, in which average shortest path lengths among the selected 20% researchers were calculated for 40 cases and their smoothed histograms were plotted, under two conditions used in Fig. 4A: selection by visibility boost (red) and pure random selection (blue). The average shortest path lengths among the top 20% researchers under the former

condition (red) were significantly shorter than their random counterparts (blue), implying that researchers strongly associated with a particular topic were indeed located closer to each other, possibly forming a research community on that topic.

We calculated correlations between the visibility boosts by research topics for a researcher and his/her overall popularity and individual-level interdisciplinarity. The popularity was measured by *total topic hits (TTH)*, defined as

$$TTH(r) = \sum_j h(r,j), \qquad (2)$$

i.e., how many search hits the researcher r had in total across all the research topics. Individual-level interdisciplinarity of a researcher was defined in this study as the diversity of research topics associated with him/her. Following similar metrics used in the literature [21,22], we characterized the individual-level interdisciplinarity by *topic hit entropy (THE)*, defined as

$$THE(r) = -\sum_j \frac{h(r,j)}{TTH(r)} \log \frac{h(r,j)}{TTH(r)}, \qquad (3)$$

which is a Shannon entropy applied to the frequency distribution of search hits over all the topics. It is small if the researcher is strongly associated with a small number of research topics but not to others, or large if he/she is associated broadly with many research topics.

Figure 5 shows correlation coefficients between the two measurements introduced above and researchers' visibility boosts by various research topics. The topics are sorted from positive to negative correlations. It is observed in Fig. 5A that common words tend to correlate positively with the overall popularity of a researcher, while technical terms tend to correlate negatively. This is not surprising, because popular researchers who frequently appear on news and other online media (i.e., those who have high total topic hits) would tend to be associated more with common words on the web. The word order changes, however, when correlations with topic hit entropy are plotted instead (Fig. 5B). There was no correlation found between Figs. 5A and 5B regarding the word positions in the rankings, which means that each research topic has unique, independent effects on popularity and individual-level interdisciplinarity of a researcher. Of partic-
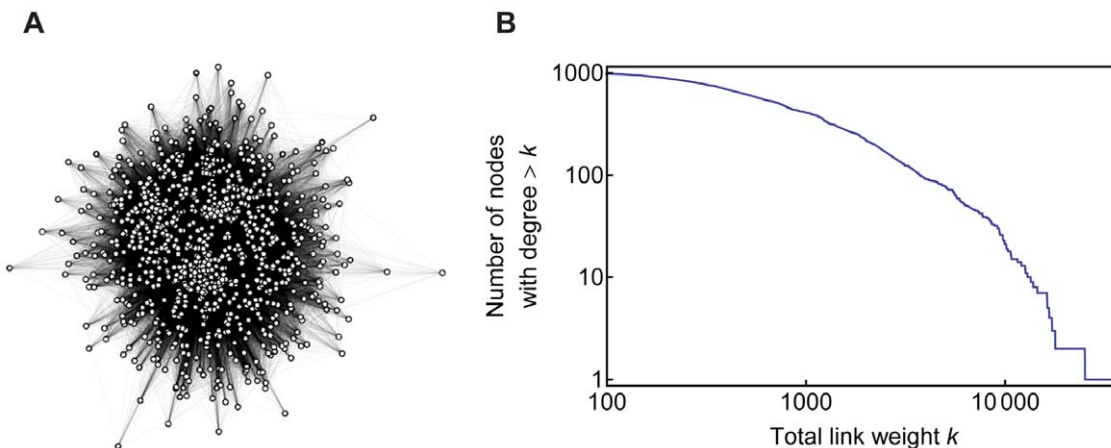
Figure 3. Reconstructed network of 1,000 researchers. (A) Visualized network. Each node represents a researcher. (B) Distribution of total link weights of nodes, plotted as a complementary cumulative distribution function (CCDF).
doi:10.1371/journal.pone.0038747.g003
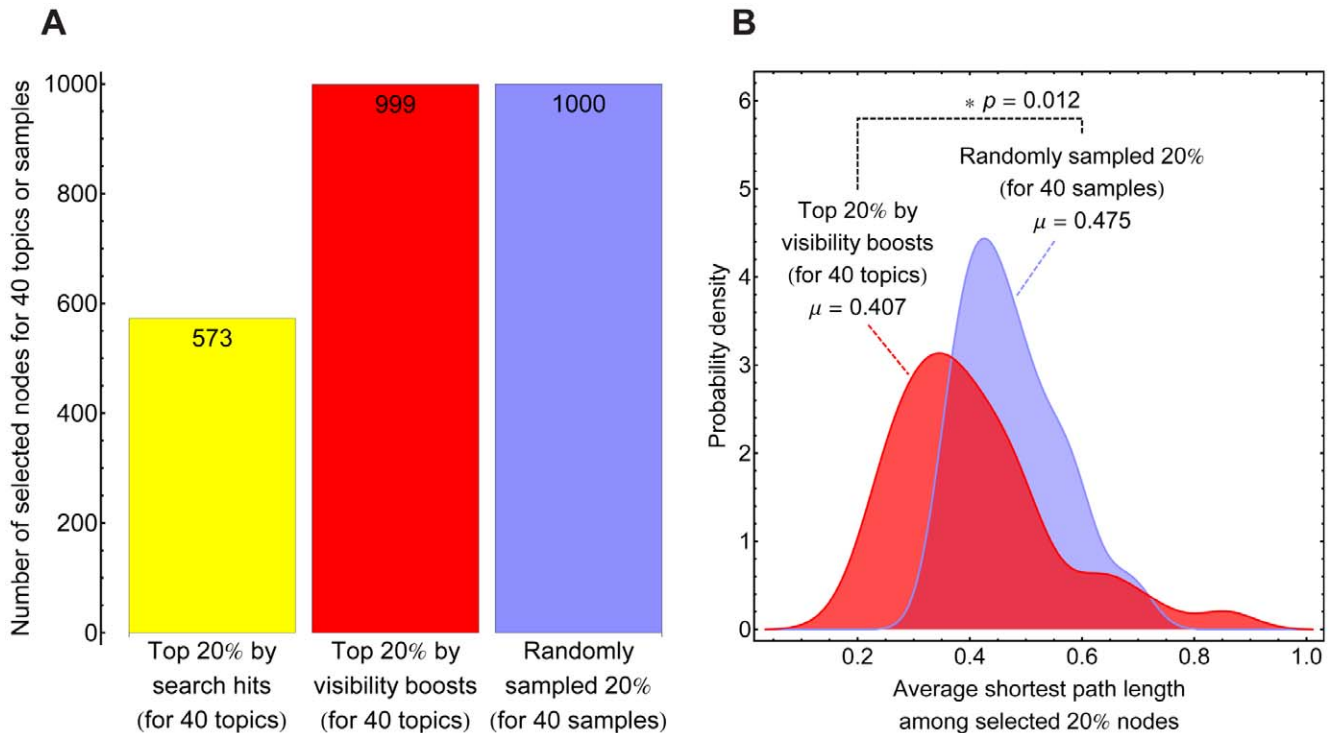
Figure 4. Use of visibility boosts. (A) Comparison of the numbers of nodes (researchers) that appeared at least once in 40 times of selection trials. Yellow: Top 20% nodes selected based on their original search hit counts for each of the 40 research topics. Red: Top 20% nodes selected based on their visibility boosts by each of the 40 research topics. Blue: Random selection of 20% nodes repeated 40 times. (B) Smoothed histograms of average shortest path lengths among the selected 20% nodes in the researchers' network ($N = 40$ for each histogram; each sample point corresponds to one measurement of average shortest path length among the selected 20% nodes). To calculate path lengths, the reciprocals of link weights were used as edge distances. The average shortest path lengths among the top 20% nodes selected based on their visibility boosts were significantly smaller than random counterparts ($p < 0.05$ by standard t-test), showing that researchers who share high visibility boosts by the same topic tended to come closer to each other in the network.
doi:10.1371/journal.pone.0038747.g004

ular interest are the research topics that moved significantly from a negative side in Fig. 5A to a positive side in Fig. 5B, such as "evolution", "biology", "neuron", "cognition", "dynamics", "simulation", and "modeling". This implies that researchers who are strongly associated with these topics tend to be less popular overall but associated with diverse topics at an individual level.

Next, we investigated correlations of a researcher's visibility boosts by research topics with his/her social-level interdisciplinarity, i.e., how "central" he/she is in the researchers' network. We considered three typical centrality measurements: degree, betweenness and closeness [30]. Some elaboration was required in measuring degree centrality because links in our network were weighted and the weights might be heterogeneously distributed. We used two approaches in measuring degrees. One was to calculate the Shannon disparity of link weights on a node, introduced by Lee et al. [14], which is given by

$$D(r) = \exp\left(-\sum_j \frac{w(r,j)}{TNH(r)} \log \frac{w(r,j)}{TNH(r)}\right) = \prod_j \frac{w(r,j)}{TNH(r)}^{-\frac{w(r,j)}{TNH(r)}}, \quad (4)$$

where $w(r, j)$ is the number of search hits for search query "r" (researcher's name) + "j" (another researcher's name) + "research", and $TNH(r)$ the *total name hits* defined as

$$TNH(r) = \sum_k w(r,k). \quad (5)$$

Note that this Shannon disparity $D(r)$ is an exponential of a Shannon entropy of the link weight distribution on node r, which is the effective number of links of researcher r if link weights were all equal. To make the terminology more intuitive, we call $D(r)$ an *effective degree* of researcher r.

The other approach we took in measuring degree centrality is to calculate *total normalized incoming link weights* of a node, defined as

$$I(r) = \sum_j \frac{w(j,r)}{TNH(j)}. \quad (6)$$

This is a sum over j of how much portion of j's link weights comes in to r, which characterizes how important node r is to other nodes.

For the other two centrality measures (betweenness and closeness), the reciprocals of link weights were used as edge distances. We used Python NetworkX's [31] built-in functions to calculate these centralities. We note that the use of betweenness as a measure of interdisciplinarity was already proposed by Leydesdorff [20], but it was the betweenness of a journal in a
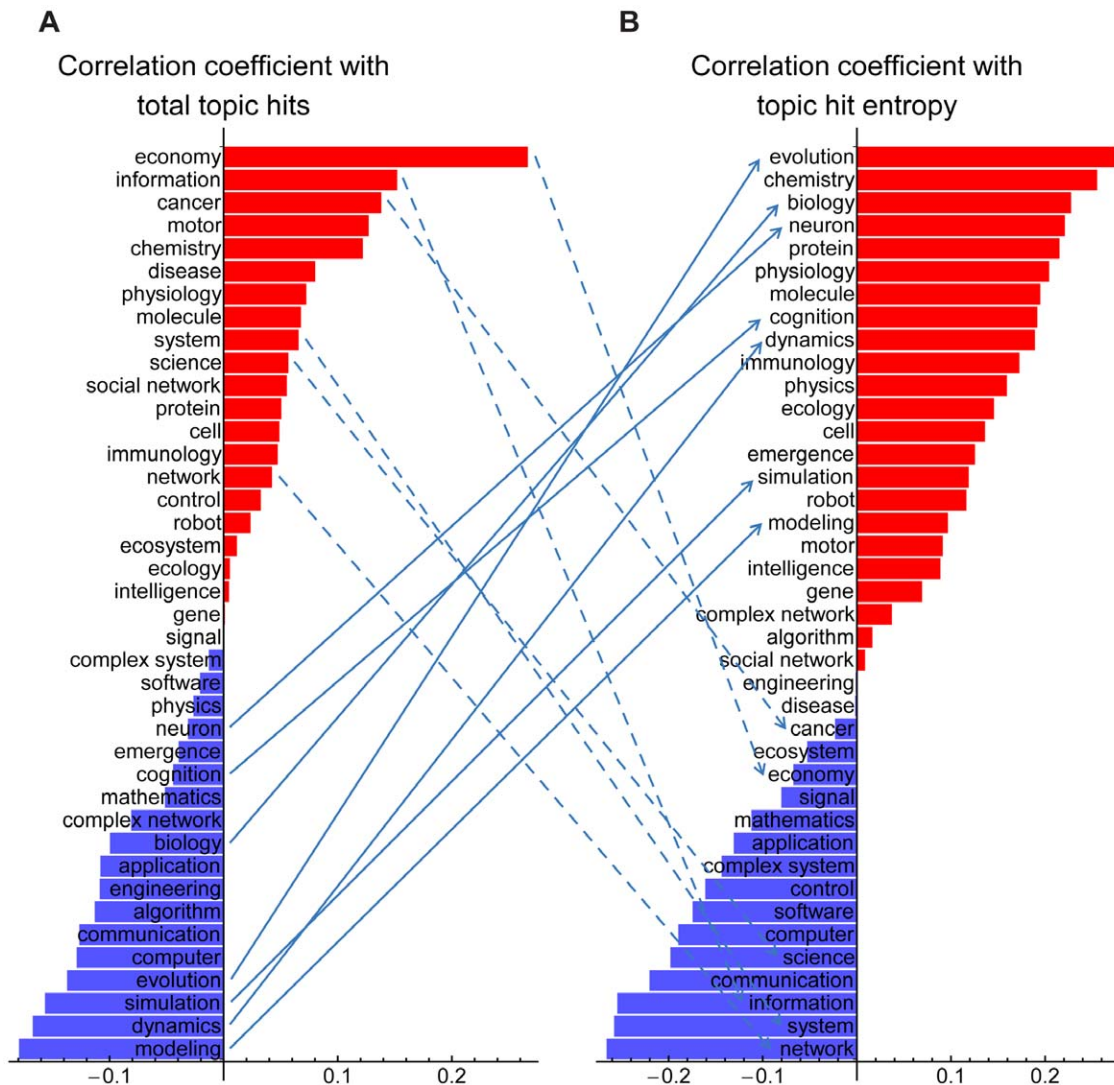
Figure 5. Correlations between the visibility boost of each research topic and a researcher's overall popularity (total topic hit, A) and individual-level interdisciplinarity (topic hit entropy, B). Upward or downward moves of topics from A to B by 20 or more places in the ranking are indicated by solid and dashed arrows, respectively.
doi:10.1371/journal.pone.0038747.g005

citation network while ours is the betweenness of an individual researcher in the researchers' relatedness network.

The results are summarized in Fig. 6. We found that the rankings of the topics "complex network" and "social network" jumped up drastically from Fig. 6A to Figs. 6B, 6C and 6D (and a single-word topic "network" also showed similar behavior, but in a slightly different way). This implies that the researchers who are strongly associated with these topics tend to be important to other researchers (Fig. 6B) and occupy central positions in the network (Figs. 6C, 6D) *without being associated with a broader range of other researchers (Fig. 6A) or topics (Fig. 5B)*. In other words, researchers who are more strongly associated with network-related topics may have higher social-level interdisciplinarity in their network without having too broad social relatedness.

Moreover, we also found that the correlation strengths of a research topic with researchers' individual-level and social-level interdisciplinarities were nearly independent from each other. Figure 7 shows the distributions of research topics in a two-dimensional correlation coefficient space based on the data in
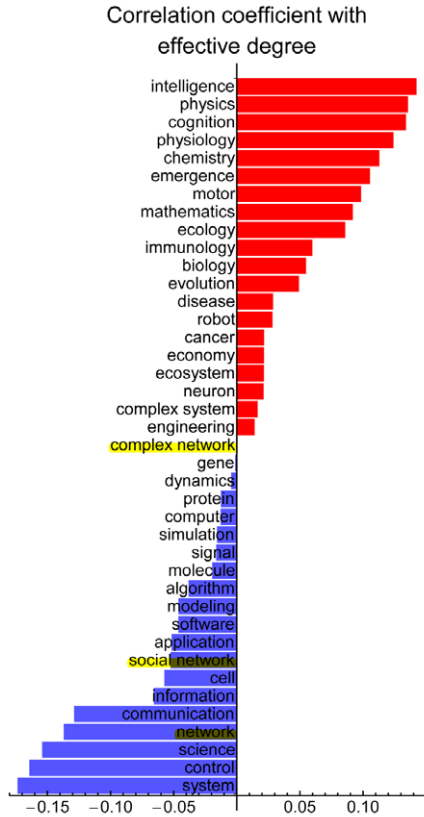
Figs. 5B, 6B, 6C and 6D, where research topics are widely scattered with no clearly identifiable tendency. This implies that the notion of "interdisciplinarity" should be understood as a multi-dimensional concept and should be evaluated using multiple assessment means, which is consistent with what has been suggested in the interdisciplinarity research literature [19,21,22].
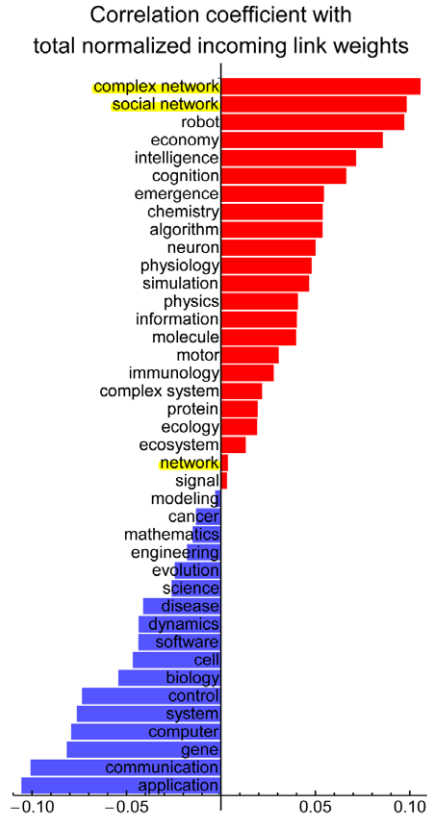
Discussion

In this paper, we illustrated our web search engine-based method to reconstruct a network of relatedness between individual researchers and research topics through preliminary data collection and analysis using a small set of keywords. Our novel contributions include the proposal of visibility boost, a new quantity defined for a pair of a researcher and a research topic, which indicates how the research topic helps increase the researcher's visibility.

Our results showed that visibility boosts by research topics correlated in various ways with other metrics. Most notably, the network-related topics increased their rankings in the order of
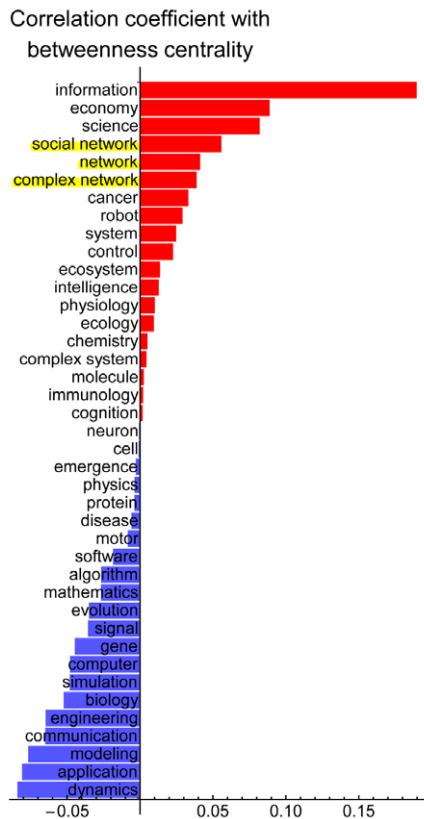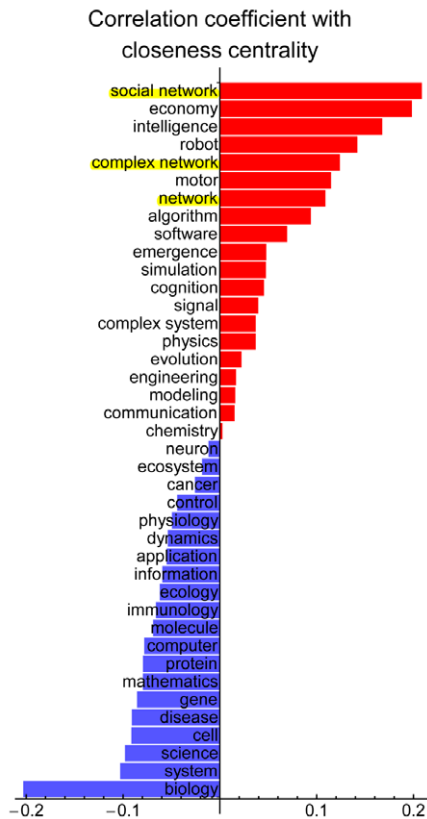
A

Correlation coefficient with
effective degree

B

Correlation coefficient with
total normalized incoming link weights

C

Correlation coefficient with
betweenness centrality

D

Correlation coefficient with
closeness centrality

Figure 6. Correlation between each research topic and a researcher's social-level interdisciplinarity. (A) Effective degree. (B) Total normalized incoming link weights. (C) Betweenness centrality. (D) Closeness centrality. Network-related topics are highlighted.
doi:10.1371/journal.pone.0038747.g006

correlations with researchers' social-level interdisciplinarity even though they were not strongly correlated with the researchers' effective degrees. This finding poses an intriguing future research question about potential causal relationships between topics a researcher works on and his/her position and role in a social context. A straightforward interpretation is that network science is currently a hot topic and therefore network researchers may be referred to more often in public media and other online documents, naturally increasing their centrality in our data set. The opposite explanation is also plausible, though, in that researchers who work at the boundaries of different disciplines may tend to choose networks as part of their research subjects because of their generality and broad applicability to many domains. Yet another, somewhat behavioral, explanation would also be possible, in that those who are aware of properties of complex networks may be able to utilize their knowledge and strategically optimize their positions in a social network. The data used in this study did not contain any causal information and therefore no conclusion can be derived at this point. More systematic studies on temporal changes of researchers' networks will help explain the underlying mechanisms responsible for the patterns observed in this study.

Our work also showed that the individual-level and social-level interdisciplinarities may not necessarily behave similarly in terms of their correlations with specific research topics. This fact suggests that the interdisciplinary nature of a research field should be considered as a multi-dimensional construct ranging over multiple levels, taking into account how many different concepts/disciplines are involved, how important the roles played by the researchers in that field are in connecting different research communities, and so on. This insight may be informative for those who work on formal or informal assessments of academic activities of researchers and research institutions.

We must emphasize that our study is still preliminary and it still has several fundamental limitations. First and foremost, the data collection takes a lot of time in the current form of the proposed

method. The number of keyword pairs one has to search for grows quadratically with the number of keywords; each pair must be searched for multiple times in order to improve the reliability of results; and search queries must be sent to web search engines at sufficient intervals in order to avoid interfering with their regular operations. This problem put significant constraints on the scalability of our method in this study. A closer collaboration with web search and other IT industries will likely offer technical solutions to this limitation.

The second limitation is the relatively low reliability of data. It is known that numbers of web search hits are often unreliable because of the lack of incentives for web search providers to give an accurate estimate of search hits (for example, see [32]). Moreover, the search results can contain anything on the web, possibly including wrong, irrelevant, and redundant webpages in search hits. We used an additional keyword "research" in every search query to reduce such risks, but it is still far from optimal. Another critical issue is the possibility of multiple people who share an identical name. We manually checked to make sure there were no such names included in our list of researchers. However, this may not be perfect because of inherent difficulty in identifying/distinguishing researchers only by their names without using metadata. In this regard, we must be cautious to note that our data and results still remain quite preliminary. To improve their reliability, one should integrate other data sources and utilize them for better filtering and analysis. Semantic analysis of search results would also be of great help in this regard, though at the cost of computational complexity.

Finally, we note that our data set and results may have been influenced significantly by the particular choices we made when collecting primary data of researchers' names and research topics. We selected highly interdisciplinary conferences that were familiar to us as the source of information because of technical reasons described earlier, but we cannot eliminate the possibility of potential biases made by these choices we made. Conducting much larger-scale data collection and analysis, starting with

Correlation with social-level interdisciplinarity
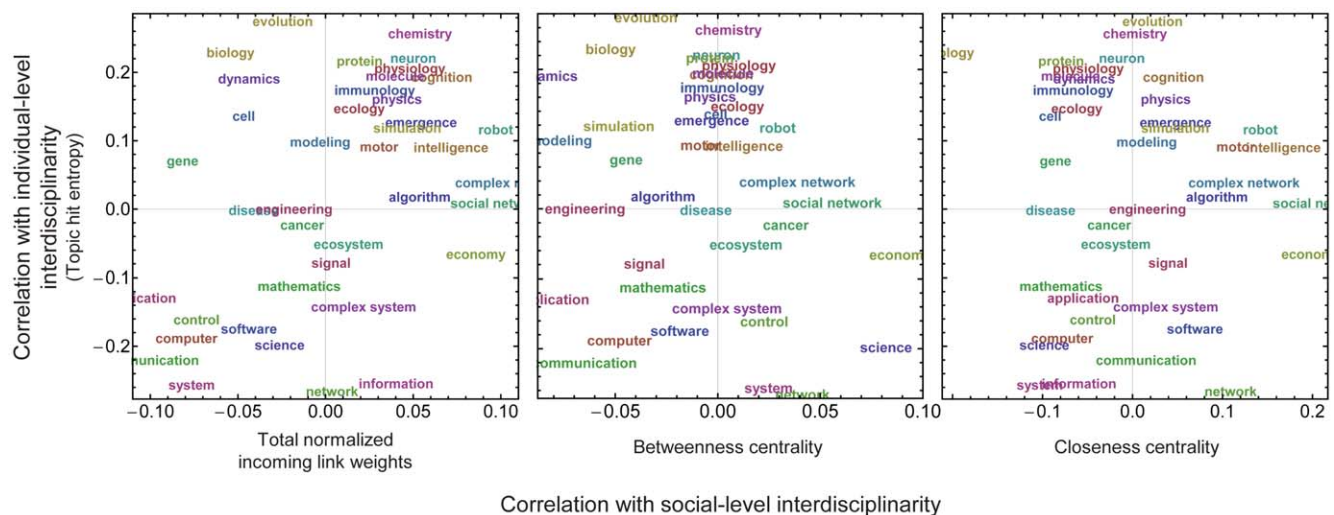
Figure 7. Two dimensional maps summarizing each research topic's correlations with a researcher's individual-level interdisciplinarity (vertical, topic hit entropy) and social-level interdisciplinarity (horizontal, three centrality measurements).
doi:10.1371/journal.pone.0038747.g007

different sets of conferences/researchers/research topics, will be necessary to reduce the effects of potential biases, which is beyond the scope of this paper.

Through this work, we aimed to illustrate a novel methodology for characterizing interdisciplinarity of researchers and their research topics by reconstructing individual-level network data about relationships among them using a general-purpose web search engine. While our technique is still preliminary with significant room for improvement and further validation, we believe it has at least two major advantages. First, web search engine-based methods like ours make it possible for everyone to have access to a large amount of data through very simple interfaces with great flexibility. Second, the methodology can be generalized from researchers' networks to virtually any kind of networks of things based on their conceptual or cognitive similarity, as long as nodes can be represented by keywords. We have published on other applications of this method elsewhere [33,34]. We believe that such web search-based research methods will become more commonly used for scholarly research in the coming years.

References

1. de Solla Price DJ (1965) Networks of scientific papers. Science 149: 510–515.
2. Friedkin NE (1978) University social structure and social networks among scientists. Amer J Sociol 83: 1444–1465.
3. Braam RR, Moed HF, Van Raan AFJ (1991) Mapping of science by combined cocitation and word analysis I. Structural aspects. J Amer Soc Info Sci 42: 233–251.
4. Newman MEJ (2001) The structure of scientific collaboration networks. PNAS 98: 404–409.
5. Newman MEJ (2001) Clustering and preferential attachment in growing networks. Phys Rev E 64: 025102.
6. Eaton JP, Ward JC, Kumar A, Reingen PH (1999) Structural analysis of co-author relationships and author productivity in selected outlets for consumer behavior research. J Consumer Psych 8: 39–59.
7. Shiffrin RM, Börner K, eds (2004) Special issue on mapping knowledge domains. PNAS 101 Suppl 1.
8. Newman MEJ (2004) Coauthorship networks and patterns of scientific collaboration. PNAS 101 Suppl 1: 5200–5205.
9. Börner K, Maru JT, Goldstone RL (2004) The simultaneous evolution of author and paper networks. PNAS 101 Suppl 1: 5266–5273.
10. Asada Y, Matsuo Y, Ishizuka M (2005) Increasing scalability of researcher network extraction from the web. Trans Japan Soc Artif Intell 20: 370–378.
11. Åström F (2007) Changes in the LIS research front: Time-sliced cocitation analyses of LIS journal articles, 1990–2004. J Amer Soc Info Sci Tech 58: 947–957.
12. Janssens F, Glänzel W, De Moor B (2008) A hybrid mapping of information science. Scientometrics 75: 607–631.
13. Liu X, Yu S, Janssens F, Glänzel W, Moreau Y, et al. (2010) Weighted hybrid clustering by combining text mining and bibliometrics on a large-scale journal database. J Amer Soc Info Sci Tech 61: 1105–1119.
14. Lee SH, Kim PJ, Ahn YY, Jeong H (2010) Googling social interactions: Web search engine based social network construction. PLoS ONE. e11233 p.
15. Klavans R, Boyack KW (2011) Using global mapping to create more accurate document-level maps of research fields. J Amer Soc Info Sci Tech 62: 1–18.
16. Barabási AL, Albert R (1999) Emergence of scaling in random networks. Science 286: 509–512.
17. Barabási AL (2003) Linked. New York: Plume. 304 p.
18. Newman M, Barabási AL, Watts DJ, eds. (2006) The Structure and Dynamics of Networks. Princeton: Princeton University Press. 624 p.
19. Porter AL, Cohen AS, Roessner JD, Perreault M (2007) Measuring researcher interdisciplinarity. Scientometrics 72: 117–147.
20. Leydesdorff L (2007) Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. J Amer Soc Info Sci Tech 58: 1303–1319.
21. Rafols I, Meyer M (2010) Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. Scientometrics 82: 263–287.
22. Wagner CS, Roessner JD, Bobb K, Klein JT, Boyack KW, et al. (2011) Approaches to understanding and measuring interdisciplinarity scientific research (IDR): A review of the literature. J Informetrics 5: 14–26.
23. Levy P (1997) Collective Intelligence: Mankind's Emerging World in Cyberspace. New York: Basic Books. 277 p.
24. Malone TW, Laubacher R, Dellarocas C (2009) Harnessing crowds: Mapping the genome of collective intelligence. MIT Sloan Research Paper No. 4732-09.
25. Google Inc (2011) Google Web Search API website. http://code.google.com/apis/websearch/docs/. Accessed 4 November 2011. Now deprecated and replaced by Google APIs Console.
26. Kautz H, Selman B, Shah M (1997) The hidden web. AI Magazine 18(2): 27.
27. Mika P (2004) Bootstrapping the FOAF-web: An experiment in social network mining. Proc 1st Workshop on Friend of a Friend, Social Networking and the Semantic Web. Available: http://www.w3.org/2001/sw/Europe/events/foaf-galway/papers/fp/bootstrapping_the_foaf_web/. Accessed 16 May 2012.
28. Mori J, Matsuo Y, Ishizuka M, Faltings B (2004) Keyword extraction from the web for person metadata annotation. ISWC Workshop Notes VIII (W8) - 4th Intl Workshop on Knowledge Markup and Semantic Annotation. pp 51–60.
29. Matsuo Y, Mori J, Hamasaki M, Nishimura T, Takeda H, et al. (2007) POLYPHONET: An advanced social network extraction system from the Web. Web Semantics 5: 262–278.
30. Wasserman S, Faust K (1994) Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press. 857 p.
31. Hagberg AA, Schult DA, Swart PJ (2008) Exploring network structure, dynamics, and function using NetworkX. Proc 7th Python in Science Conference. pp 11–15. Also see the NetworkX website http://networkx.lanl.gov/.
32. Sullivan D (2011) Why Google can't count results properly. A blog post on SearchEngineLand.com. October 21, 2010. http://searchengineland.com/why-google-cant-count-results-properly-53559 Accessed 13 December 2011.
33. Akaishi J, Sayama H, Dionne SD, Chen X, Gupta A, et al. (2010) Reconstructing history of social network evolution using web search engines. Proc. BIONETICS 2010, #10617. Available: http://bingweb.binghamton.edu/~sayama/papers/BIONETICS2010-akaishi.pdf. Accessed 16 May 2012.
34. Dionne SD, Akaishi J, Chen X, Gupta A, Sayama H, et al. (2012) Retrospective relatedness reconstruction: Applications to adaptive social networks and social sentiment. Organizational Research Methods DOI: 10.1177/1094428112442572. In press.

Supporting Information

Appendix S1 Detailed methods of web search engine-based data collection used in this study.
(DOC)

Acknowledgments

Author Contributions

Conceived and designed the experiments: HS JA. Performed the experiments: HS JA. Analyzed the data: HS. Contributed reagents/materials/analysis tools: HS JA. Wrote the paper: HS.