

RESEARCH ARTICLE

How Does National Scientific Funding Support Emerging Interdisciplinary Research: A Comparison Study of Big Data Research in the US and China

Ying Huang¹, Yi Zhang^{1,2}, Jan Youtie³, Alan L. Porter⁴, Xuefeng Wang^{1*}

1 School of Management and Economics, Beijing Institute of Technology, Beijing, 100081, China, **2** Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, NSW 2007, Australia, **3** Enterprise Innovation Institute, Georgia Institute of Technology, Atlanta, GA, 30332, United States of America, **4** School of Public Policy, Georgia Institute of Technology, Atlanta, GA, 30332, United States of America

* wxf5122@gmail.com



OPEN ACCESS

Citation: Huang Y, Zhang Y, Youtie J, Porter AL, Wang X (2016) How Does National Scientific Funding Support Emerging Interdisciplinary Research: A Comparison Study of Big Data Research in the US and China. PLoS ONE 11(5): e0154509. doi:10.1371/journal.pone.0154509

Editor: James Wilsdon, University of Sheffield, UNITED KINGDOM

Received: August 11, 2015

Accepted: April 14, 2016

Published: May 24, 2016

Copyright: © 2016 Huang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This research is supported by grants from the US National Science Foundation (award no. 1527370), the National High Technology Research and Development Program of China (grant no. 2014AA015105), the General Program of National Natural Science Foundation of China (grant no. 71373019). Besides, the authors are grateful for the scholarship provided by the China Scholarship Council (CSC Student ID 201406030005). The funders had no role in study design, data collection

Abstract

How do funding agencies ramp-up their capabilities to support research in a rapidly emerging area? This paper addresses this question through a comparison of research proposals awarded by the US National Science Foundation (NSF) and the National Natural Science Foundation of China (NSFC) in the field of Big Data. Big data is characterized by its size and difficulties in capturing, curating, managing and processing it in reasonable periods of time. Although Big Data has its legacy in longstanding information technology research, the field grew very rapidly over a short period. We find that the extent of interdisciplinarity is a key aspect in how these funding agencies address the rise of Big Data. Our results show that both agencies have been able to marshal funding to support Big Data research in multiple areas, but the NSF relies to a greater extent on multi-program funding from different fields. We discuss how these interdisciplinary approaches reflect the research hot-spots and innovation pathways in these two countries.

Introduction

Dramatic advances resulting from the rapid pace of technological developments and new interdisciplinary fields are on the horizon [1]. Science and engineering research continually evolves beyond the boundaries of single disciplines and offers employment opportunities that require not only depth of knowledge but also breadth of knowledge, integration, synthesis, and an array of skills. Interdisciplinary research (IDR), defined as work which integrates theories, methods, tools and/or concepts from multiple specialized knowledge bases, is often treated as proxy for research to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice [2, 3]. Since IDR is increasingly prominent today, compiling reflections on its premises and issues could serve

and analysis, decision to publish, or preparation of the manuscript.

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

those engaged in R&D analysis, policy and management [4], and accelerate the process of scientific discoveries and societal problem solving.

Interdisciplinarity has three distinct components that are typically taken into account: Variety, Balance and Disparity [5]. Shannon entropy [6] and Simpson [7] offered popular diversity indices in the ecological literature and more recently in the interdisciplinarity literature to capture both variety and balance of referenced disciplines. Furthermore, Rao [8] and Stirling [5] proposed disparity as the third dimension, and Rao-Stirling diversity was a composite measure of the three diversity components [5]. Further indicators, based on the above theoretical foundations, have been proposed to measure IDR, including network coherence [9], specialization scores [10], the integration score [11, 12], the diffusion score [13], and other improved indicators [14, 15]. However, most of bibliometric literatures on measuring IDR focuses on the outputs of science, or publications [16]. Less emphasis has been placed on inputs, such as research proposals, which typically contain broader information than research articles. Proposals reflect on academic individuals, networks, and evolutionary processes of science and directly indicate overarching research programs, which, if funded, can lead to multiple publications [17]. In general, research proposals are granted by national governments and aimed to support academic institutions and R&D departments to conduct basic research, the content of which focuses on new ideas, concepts, and potential innovative actions. Understanding of funding proposals could be considered as an express path to reveal how evolutionary R&D pathways work in given countries, regions, and research fields [18].

Scientific funding plays an essential role in individual scientific research, university discipline construction, and national innovation system patterns [19]. Previous research on scientific funding has focused on research investments [20], optimization design for peer review [21], and funding distribution [22]. In addition, evaluations of the social impact [23, 24] and performance [25, 26] of science funding reflect an evaluative perspective on funding research.

As science increasingly deals with boundary-spanning problems, various policy and funding initiatives have been developed to encourage interdisciplinary research to push forward academic capability and accelerate scientific discovery. National scientific funding organizations specifically support research, including investigations of an interdisciplinary nature, through a variety of target methods [24]. As the two leading scientific funding organizations in the world, National Science Foundation (NSF) of the US and National Natural Science Foundation of China (NSFC) are noteworthy. NSFC, as the largest scientific funding agency in China, has made considerable efforts to promote basic and applied research [27]. NSFC not only supports cutting-edge ideas and projects in different disciplines, but also sponsors high-level strategic academic exchange platforms (e.g. Shuangqing Forum) to promote interdisciplinary studies and innovative cultures and create a favorable environment for research [28]. Dr. Wei Yang, the present president of NSFC, defined strengthening interdisciplinary interaction as one key mission of NSFC [29]. The US NSF supports research and education in non-medical fields of science and engineering, not only to produce leading edge research, but also to narrow the gap between science and society through the use of broader impacts criteria [30, 31]. NSF treats fostering IDR as its high priority and encourages researchers to conduct IDR research in novel and emerging areas extending beyond any particular current NSF programs to keep the US at the leading edge of discovery in a wide range of scientific areas [32].

We are especially interested in examining how these two agencies support research in a rapidly growing interdisciplinary domain. The domain of interest for this study is Big Data. In an increasingly complex economic and social environment, accessing vast amounts of data and information can help organizations and governments make better policies, predictions and decisions [33]. Big Data is the application of data technology to obtain valuable information from various types of extremely large data sets, which can come from social networks, images,

sensors, the web, or other unstructured sources [34]. Big Data has its legacy in information technology developments. However, its capacity and analytic capabilities promise to make an essential contribution in areas such as traffic management, logistics, health care, and education [35]. Big Data research can help further these scientific and societal benefits for governments, enterprises and academic institutions.

Aiming to make the most of the fast-growing volume of digital data, the US government established a “Big Data Research and Development Initiative” program on March 29, 2012, which committed over \$200 million in new funding through six agencies to extract knowledge and insights from large and complex collections of digital data [36]. The Chinese government has similarly targeted Big Data as a strategic area of research and development [37]. Since 2012, a large number of Big Data-related projects have been supported by the Ministry of Science and Technology, the National Development and Reform Commission, the Ministry of Industry and Information Technology, and other central governmental departments of China. Porter and his colleagues [38] figured out that Big Data scientific publications grew dramatically in 2013 and 2014, by more than four times the number of the papers published in 2012. They further reported that the leading countries based on author location were the US and China; these two countries accounted for more than half of all Big Data publications and nearly all of the top 30 author organizations.

In this paper, we apply tech mining [i.e., text mining of science, technology & innovation (ST&I) information to generate useful intelligence] [39], network analysis, and interdisciplinary assessment methods to understand how NSF and NSFC have ramped-up their funding awards to address the rapid rise of Big Data research. Our particular focus is on research awards rather than scientific articles [40–43]. Our research addresses the following questions about how NSF and NSFC support research in rapidly emerging areas, based on a case study of the emergence of Big Data research:

1. How do NSF and NSFC differ in terms of the amount of funding resources allocated to Big Data research over time?
2. To what extent do NSF and NSFC fund similar or different subareas of Big Data research?
3. What are the disciplinary ranges and connections in Big Data funded proposals in the two organizations and how do they differ in the degree of interdisciplinarity of funding research?

The remainder of this paper consists of four sections. Following this general introduction, the “Methodology” section describes the framework and methodology. The “Results” section presents comparative results of the research profiling and tech mining analyses. The “Conclusion and Discussion” section reviews our research, identifies research limitations, and indicates promising research opportunities to pursue in the future.

Methodology

Based on the aforementioned research questions, we retrieved a raw dataset of all funded research proposals relating to Big Data from the NSF official website (<http://www.nsf.gov/>) and the NSFC’s ISIS system (Internet-based Science Information System) (<http://isisn.nsf.gov.cn/>). The NSF website offers two different search types: (1) simple search: provides title, abstract, names, institutions, programs and other information associated with an award; and (2) consolidated advanced search: provides access to all fielded searches. In this paper, we chose the consolidated advanced search.

The ISIS system was applied nationwide in 2003, and it has been well received and supports NSFC to implement managerial control to cope with China's rapidly growing research productivity [44]. In the ISIS system, users can retrieve the information by typing specific search terms in the query page, but only a few fields can be viewed: Awards Number, Disciplinary Application Code, Title, Principal Investigator, Affiliation, Sponsored Funds, Start Date and End Date.

A key question is when to begin the search. First, we originally set the starting year at 2008, which is when a special issue of *Nature*, "what Big Data sets mean for contemporary science" [45], was published. However, there are no awards for the year 2008 in our retrieval of funded proposals. Thus, 2009 became our starting year. Second, we originally planned to use a topical search, which captures the occurrences of "Big Data" in titles, abstracts and keywords [46]. This approach worked for searching the NSF awards database. Ultimately, we obtained 921 records from NSF and 434 records from NSFC during the period of 2009 to 2015. However, because the NSFC data are in Chinese, we found that use of standard codes such as the NSFC's Disciplinary Application Code (DAC), was more effective than direct translation of topics during the analyzing process. We validated the DACs further by arranging native speakers to examine our approach and select NSFC funded proposal records to guarantee its comparability.

We conducted a retrieval of publications from Web of Science (<http://www.webofknowledge.com/>) to verify some conclusions and judgments about searching the funded proposal databases using an elementary search strategy of TS = "Big Data". This approach resulted in 2815 records during the period of 2008 to 2015 (retrieved on January 8, 2016). We imported the data into the text-mining tool VantagePoint (<https://www.thevantagepoint.com/>), and ascertained from the funding acknowledgements section of these papers that 272 publications were supported by NSFC, and 198 by NSF. They ranked as the top 2 national funding agencies, and occupied 22.48% and 16.36% respectively in all funded publications (1210 publications). All raw data of Big Data proposals funded by NSF and NSFC can be founded in "S1 File".

In addition to seeking to measure basic activity information, we also attempted to interpret the reasons and rules behind the activities of these two funding agencies. To this end we drew on the work of tech mining research [47, 48] to extract metadata for research profiling. We also introduced two indicators to measure the inequality of funding organization distribution and the degree of interdisciplinarity for a single granted project.

The framework of this paper, comprising the data resources, research questions, and analytical methods, is diagramed below (Fig 1).

Results

Research Profiling Analysis

Fig 2 shows grant activity trends for Big Data supported by NSF and NSFC from 2009 to 2015. Looking at the counts of grant awards, both NSF and NSFC supported very limited Big Data research before 2012. For NSF, we can see an increase in 2012; a notable rise in both counts of Big Data awards and dollar amounts associated with these counts can be observed. For NSFC, the number of granted projects increased from only 1 record in 2009 to 61 records in 2013; these figures for NSF are 2 and 204. In regards to funds granted, NSF sponsored universities and other organizations with \$45.82 million in 2012, 42 times more than NSFC did (\$1.07 million dollars). NSF and NSFC both continued to increase funding, reaching \$108.84 million and \$37.45 million in 2014, respectively. There is a slight downturn in NSFC award activity in 2015 which is not presented in the NSF award data. This downturn in NSFC award activity may not necessarily represent a decline in funding Big Data research in China. In China, nearly every

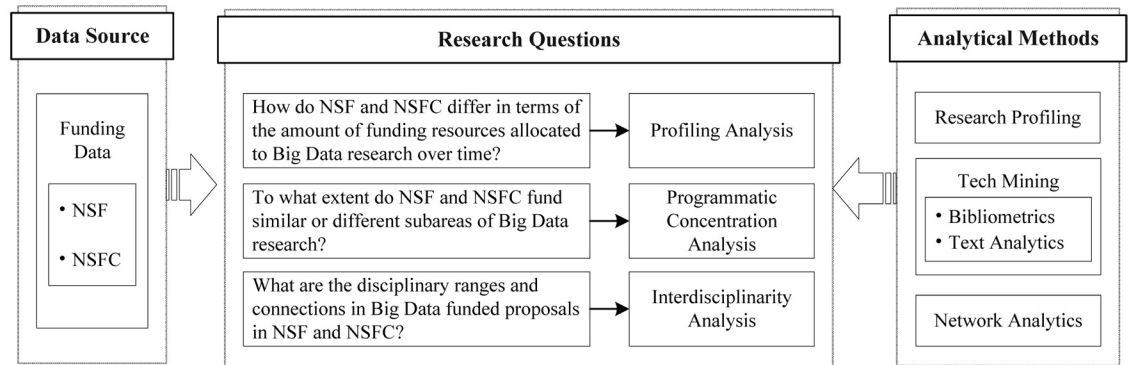


Fig 1. The Framework of Paper.

doi:10.1371/journal.pone.0154509.g001

province has its own science foundation (including Natural Science Foundation and Social Science Foundation). At the same time, several central departments also offer funding to support scientific research, such as the Ministry of Science and Technology of China, the Ministry of Education of China, and so on. While a parallel system exists for the US (through research funding by state governments, other federal agencies, and private non-profit organizations), the US system is not as extensive as is the case in China.

Why was 2013 such a bellwether year for Big Data research proposals in both countries? We are not privy to all the reasons for this growth, but one factor is likely to be the White House’s announcing the “Big Data Research and Development Initiative” in 2012. NSF subsequently announced its support of new research to extract knowledge and insights from large and complex collections of digital data, including developing new methods of deriving knowledge from data; constructing a new infrastructure to manage, curate and deliver data to communities; and forging new approaches for associated education and training [49]. Four specific programs were set up through the NSF’s Computer and Information Science and Engineering Directorate

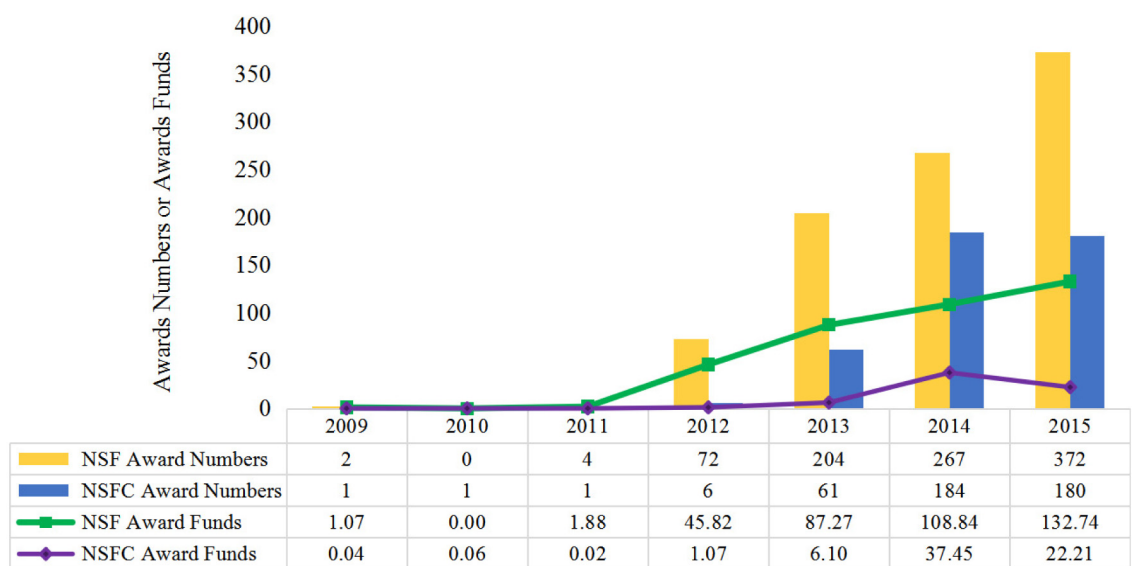


Fig 2. Summary of Big Data Awards Supported by NSF and NSFC. Note: NSFC funds data are transformed by using the exchange rates in current year.

doi:10.1371/journal.pone.0154509.g002

Table 1. The Profile of Big Data- Related Shuangqing Forum.

Workshop ID	Topic	Time	Place	Organizers
89th	Challenging scientific problems in the fields of technologies and applications of Big Data	March 5–7, 2013	Shanghai	NSFC, Tongji University
109th	Big data and major basic problems on management and decisions	March 26–27, 2014	Beijing	NSFC
132th	Multi-disciplinary methods for Big Data analysis and processing	March 26–27, 2015	Guangzhou	NSFC, Sun Yat-sen University

doi:10.1371/journal.pone.0154509.t001

(CISE). One of the most important programs is the Critical Techniques and Technologies for Advancing Foundations and Applications of Big Data Science & Engineering (BIGDATA Program). Multiple NSF directorates and other federal agencies participated in this program. The research directions of NSFC are influenced by NSF to some extent. At the same time, NSFC began to organize the high-level Big Data-related “Shuangqing Forum” academic workshop in 2013 to focus on national strategic development. This workshop was repeated in subsequent years, which suggested NSFC’s persistent value towards the Big Data area for strategic research investment (Table 1). These forums proposed several frontiers for crucial scientific problems and suggested policies and solutions associated with research on technologies and applications of Big Data [50]. Moreover, the topical focus progressed from an emphasis on problems in the Shanghai workshop in 2013 to methods two years later at the Guangzhou workshop.

What type of projects resulted from these research investments? Table 2 indicates the proportion of granted projects by organization type in the US and China. Both NSF and NSFC tended to emphasize academic research, which accounted for 92.83% and 88.25% of the total number of funded proposals. This finding is not surprising given NSF and NSFC are the agencies with an orientation toward providing support for academic research [51]. At the same time, NSFC awards are less intensely focused on universities than are NSF’s. This lesser focus lies in the importance of research institutes in the Chinese research and innovation system, particularly the Chinese Academy of Sciences. The Chinese Academy of Sciences is seen as the linchpin of China’s drive to explore and harness high technology and the natural sciences for the benefit of China. Thus, it is not surprising that the Chinese Academy of Science would be prominent among institutions receiving awards from NSFC. About 11.29% of NSFC’s grants were granted to research institutes compared to only 2.28% for NSF. However, NSFC does not support research funding by individuals and corporations, so none of NSFC’s funding went to

Table 2. Grant Recipient Organizations: NSF and NSFC.

Organization Type	Numbers of NSF Awards	% of Total NSF Big Data Awards	Numbers of NSFC Awards	% of Total NSFC Big Data Awards
Academic University	855	92.83	383	88.25
Research Institution	21	2.28	49	11.29
Corporation	39	4.23	0	0.00
Other	6	0.65	2	0.46

Note: *Academic University* is the organization granting academic degrees in various subjects; *Research Institution* is an independent, nonprofit research institute; *Corporation* is a company or group of people authorized to pursue profit-making rather than non-profit; *Other* includes individuals, hospitals and other research organizations.

doi:10.1371/journal.pone.0154509.t002

the private sector. Nearly 4.23% of awards were conferred on individuals and corporations by NSF through grants, and cooperative agreements.

We provided a further measure of the extent of concentration of the distribution of funding by NSF and NSFC. In this context, we applied a normalized Gini coefficient to measure the inequality of funding organization distribution, both in project counts and funding amounts. The Gini index was originally proposed to measure income inequality, but it also has been used to capture the inequality, unevenness, and imbalance of the distribution of references across involved disciplines. Higher Gini coefficients represent greater inequality, in this case, of scientific funding distribution across organizations. The normalized Gini coefficient can be calculated as [52]:

$$G_n = \frac{\sum_{i=1}^n (2i - n - 1)X_i}{(n - 1) \sum_{i=1}^n X_i} \tag{1}$$

In the eq (1), n indicates the number of sponsored organizations and $n > 1$, i is the ranking of n , and x_i presents the number of projects or the amount of funds obtained by the i^{th} organization.

We applied this measure to the distribution of organizations receiving grants from the two agencies over the 2012-to-2015 period because the time after 2012 comprised the bulk of the grants. The results indicated different distributions for the number of grants (FundNum) and the amount of sponsored money (FundMoney) (Fig 3). In terms of the FundNum, NSF has a relatively lower Gini coefficient than does NSFC. NSFC’s Gini coefficient was much lower in 2015 than in 2012. Furthermore, when one looks at the distribution of money, we can see that NSFC’s Gini coefficient was slightly higher than NSF’s in 2012, but coefficients associated with the two sponsors were relatively similar by 2015. NSF’s coefficient associated with the distribution of money exhibited a slight decrease over the three years while NSFC’s coefficient was decreasing more sharply.

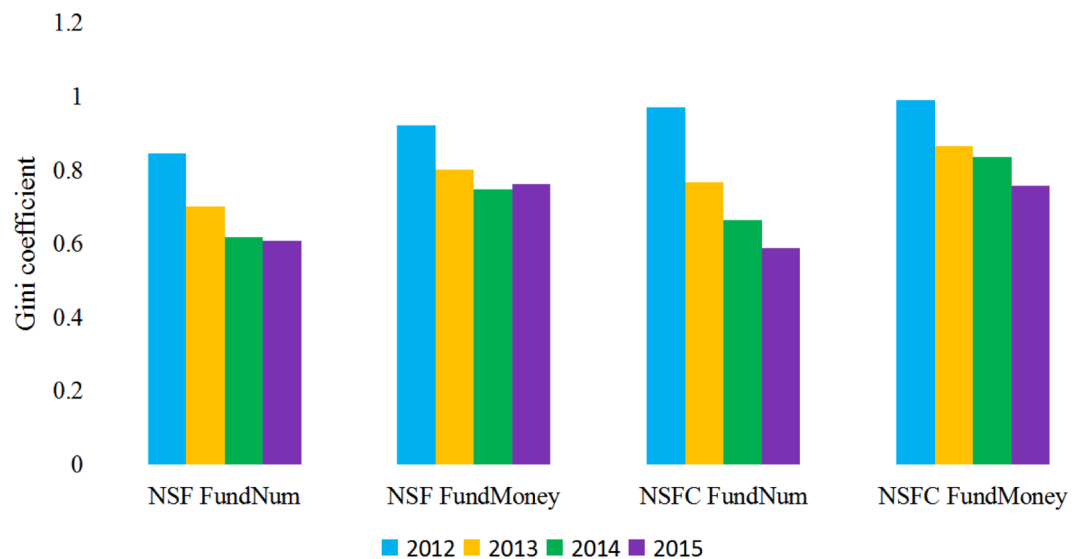


Fig 3. Gini Coefficient in Terms of the Number of Grants (FundNum) the Amount of Sponsored Money (FundMoney) of NSF and NSFC.

doi:10.1371/journal.pone.0154509.g003

Based on the Gini coefficient analysis, we can see that the Gini coefficients were very high in the early stages, in terms of granted numbers and funds, at both sponsoring agencies. It is inferred that in the beginning stage of emerging interdisciplinary research, relatively few organizations care about, or have considerable influence on, such newly-formed topics. From the overall trends of the Gini coefficient, the awards distribution became broader with the advance of Big Data studies, and the number of grants was more balanced between different organizations than the amount of sponsored money bestowed.

Proposal Programmatic Concentration Analysis

The extent to which funding is concentrated or spread among multiple disciplinary programs in a funding agency may well be an important factor in understanding the development of research in a rapidly emerging field. The NSF is organized into directorates that align with broad scientific disciplines: Biological Sciences (BIO), Computer & Information Science & Engineering (CISE), Education & Human Resources (EHR), Engineering (ENG), Geosciences (GEO), Mathematical & Physical Sciences (MPS), Social, Behavioral & Economic Sciences (SBE) and Other (non-disciplinary specific). Because MPS houses several markedly disparate disciplines, the MPS divisions (Astronomical Sciences, Chemistry, Material Sciences, Mathematics, and Physics) are often used instead [17]. In this paper, we treat the divisions under these directorates as different research areas and consider the NSF’s detailed programs as a proxy for research fields. Similarly, in the NSFC, there are eight scientific departments: Mathematical and Physical Science (A), Chemical Sciences (B), Life Sciences (C), Earth Sciences (D), Engineering and Materials Sciences (E), Information Sciences (F), Management Sciences (G) and Medical Sciences (H) (The department of Medical Sciences in NSFC established in 2010). For each project, NSFC applicants must provide a DAC in order to select suitable peer reviewers and help classify the project in its evaluation. DAC is a three-level code indicating the detailed discipline to which an application belongs; it is composed of English characters and Arabic numerals. The English character is the code of a scientific department. The three levels of Arabic numerals denote research areas, research fields and research directions, respectively. The detailed research areas and their corresponding codes are shown in S1 Table.

Although NSF and NSFC have different disciplinary categorizations, most of these categories can be matched. Table 3 indicates that the categories of Information Sciences and

Table 3. Disciplinary Composition of the Big Data Awards in NSF and NSFC.

NSF Discipline	Records	% of Total NSF Big Data Awards	NSFC Discipline	Records	% of Total NSFC Big Data Awards
Computer & Information Science & Engineering (CISE)	511	55.48	Information Sciences (IS)	236	54.38
Engineering (ENG)	134	14.55	Management Sciences (MaS)	111	25.58
Mathematical & Physical Sciences (MPS)	127	13.79	Engineering & Materials Sciences (EMS)	30	6.91
Social, Behavioral & Economic Sciences (SBE)	61	6.62	Earth Sciences (ES)	18	4.15
Education & Human Resources (EHR)	38	4.13	Mathematical & Physical Sciences (MPS)	17	3.92
Biological Sciences (BIO)	34	3.69	Medical Sciences (MeS)	10	2.30
Geosciences (GEO)	12	1.30	Chemistry Sciences (CS)	2	0.46
Other	4	0.43	Life Sciences (LS)	1	0.23
			Other	9	2.07

doi:10.1371/journal.pone.0154509.t003

Table 4. Research Areas and Main Research Fields of CISE in NSF.

Research Areas and Main Research Fields	Records	% of Total NSF Big Data Awards
Information & Intelligent Systems	189	20.52
*Big Data Science & Engineering	60	-
*Info Integration & Informatics	51	-
*Information Technology Research	17	-
Computing & Communication Foundations	114	12.38
*Communication & Information Foundations	26	-
*Algorithmic Foundations	21	-
*Software & Hardware Foundation	18	-
Computer & Network Systems	107	11.62
*Computer Systems	31	-
*Computing research infrastructure	10	-
Advanced Cyberinfrastructure	101	10.97
*Campus Cyberinfrastructure	37	-

* indicates the research fields.

doi:10.1371/journal.pone.0154509.t004

Computer & Information Science & Engineering account for the largest number of proposed projects. Big Data has conventionally been considered as a part of information sciences since it is the process of mining potential information from voluminous amounts of structured, semi-structured and/or unstructured data. Big Data also has a very close relationship with information and computer technologies, including data collection, storage, processing, and analysis/ visualization [53]. For NSF, the second ranking discipline is Engineering, but for NSFC, it is Management Sciences. This shows that US researchers cared more about practical applications in specialized engineering fields while Chinese scholars are more interested in strategic planning to improve decision-making in critical development areas, such as healthcare, social administration, environment protection and resource management. One factor underlying this difference is that China and US are at different stages of development, so Big Data is sometimes treated as a powerful tool to solve practical issues in the US but as a tool for management reform in China.

In order to obtain insights into the specific research areas and their main research fields (more than 15 projects belong to a certain area) between Computer & Information Science & Engineering (CISE) of NSF and Information Sciences (IS) of NSFC, we organized them in Table 4 (for NSF) and Table 5 (for NSFC). For NSF, the main four research areas in CISE were Information & Intelligent Systems (189 records, 20.52%), Computing & Communication Foundations (114 records, 12.38%). Computer & Network Systems (107 records, 11.62%) and Advanced Cyberinfrastructure (101 records, 10.97%) and. The top research fields within these areas were Big Data Science & Engineering, Info Integration & Informatics, Campus Cyberinfrastructure and Computer Systems.

Although we could not create a side-by-side comparison on research fields between NSF and NSFC, it was still possible to compare and contrast the top funding fields and hot-spots in Tables 4 and 5. As the comparison in topics relies heavily on the accuracy of the translations, we attempted to conduct this analysis for NSFC based on the DAC that included subject classification information, and then translated the main research fields corresponding to the DAC. The results indicate that NSFC primarily funds Big Data research out of the Computer Science area, which comprises nearly 39.17% of all Big Data awards. Another 10.83% of awards are conferred by the Automation area and 4.38% by the Electronics & Information System. Within

Table 5. Research Areas and Main Research Fields of IS in NSFC.

Research Areas and Main Research Fields	Records	% of Total NSFC Big Data Awards
Computer Science	170	39.17
*Computer Applications Technology	65	-
*Computer Software	41	-
*Computer Network	23	-
*Computer Architecture	17	-
Automation	47	10.83
*Artificial Intelligence & Knowledge Engineering	18	-
Electronics & Information System	19	4.38

* indicates the research fields.

doi:10.1371/journal.pone.0154509.t005

the Computer Science area, NSFC has balanced the number of awards across several different scientific directions. Computer Applications Technology, Computer Software, Computer Network and Computer Architecture are important fields in the Computing Science research area. Additionally, other prominent research field is Artificial Intelligence & Knowledge Engineering (18 records) in the Automation area (47 records).

When proposers apply for funding from the NSFC, they are required to provide project terms or keywords. As far as we know, keywords are not required for NSF proposals and are therefore not comparably available on the NSF website. To address the lack of investigator-provided keywords, we conducted Natural Language Processing (NLP) on the proposal title field. We used the title field, rather than the abstract field, because the terms in the title were more distinctive. We extracted phrases from the title by applying NLP with the support of the text-mining tool suite- VantagePoint. Phrases and terms retrieved in this way are large and "noisy," making them difficult to manually categorize. Using bibliometric and text mining techniques, this paper applied semi-automated "Term Clumping" to generate better term lists for achieving competitive technical intelligence [54]. For the NSFC awards information, we first extracted the title terms and uploaded them to the LTP-Cloud (Language Technology Platform Cloud) (<http://www.ltp-cloud.com/>) to process Chinese word segmentation. After obtaining a list of phrases, we imported these keywords into our own Chinese text analysis tools—ItgInsight—to help us conduct text cleaning. This process was comprised of four steps: (1) Common and basic term removal, e.g., instance, technology; (2) Fuzzy word matching (to combine terms with similar structures based on pattern commonality, such as stemming—e.g., institute and institution, and combining singular and plural forms of English words, e.g., technology and technologies); (3) Extreme word removal [to remove very common (top 5%) and very rare (occurrence only in single records) terms]; (4) Combine term networks (to combine selected low-frequency phrases with the high-frequency phrases that appear in the same records, and sharing terms). We then translated the top 50 high-frequency Chinese phrases and invited some postgraduates with English language background or bachelor's degree to validate the translation.

Before visualizing the semantic networks based on these keywords, we calculated the frequency of specific terms. The semantic networks of the 30 most frequently occurring terms in Big Data are shown in Fig 4 for NSF and Fig 5 for NSFC, which are mapped in the visualization and exploration software- Gephi (<http://gephi.github.io/>). The node size represents the frequency of a term in proposal titles, and the linkages among different terms are based on their co-occurrence in proposal titles. The color of the node is used to distinguish whether these nodes belong to the same cluster or not, and the cluster membership was obtained from the community detection algorithm [55].

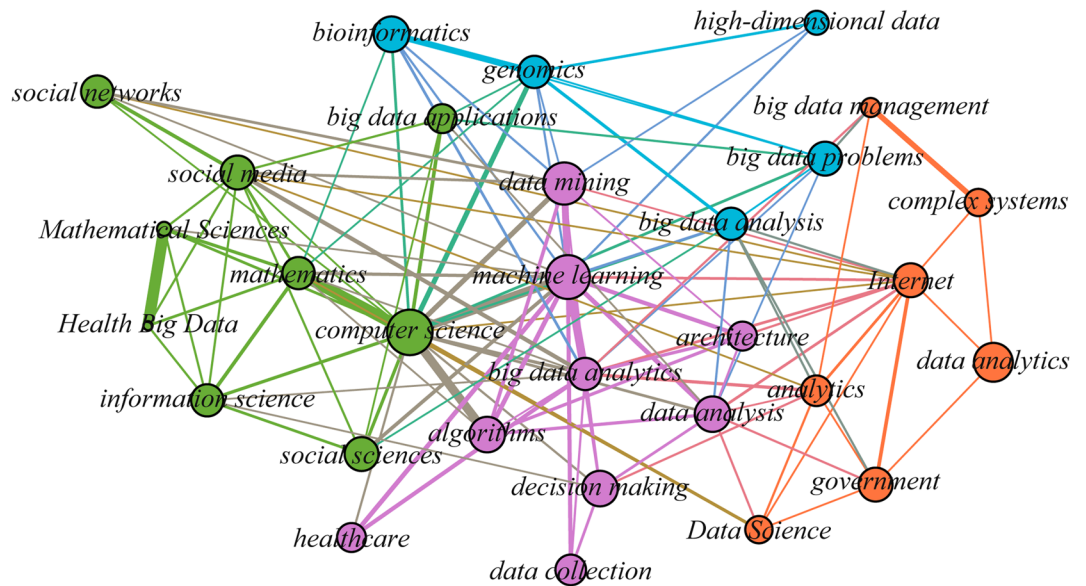


Fig 4. Semantic Network of 30 Most Frequently Occurring Terms in NSF Big Data Proposals.

doi:10.1371/journal.pone.0154509.g004

The 30 most frequently occurring terms in NSF granted proposals were clustered into four groups as follows (Fig 4):

- Green group represents the use of social network analysis in Big Data area; some related branch subjects and keywords are “social networks,” “social sciences,” “social media,” “mathematical sciences,” “mathematics,” “computer science,” “big data applications”; “Health Big Data” and “information science”;
- Purple group indicates some Big Data analytics techniques; keywords include: “data mining,” “machine learning,” “architecture,” “big data analytics,” “data analysis,” “algorithms,” “decision making,” “healthcare” and “data collection”;

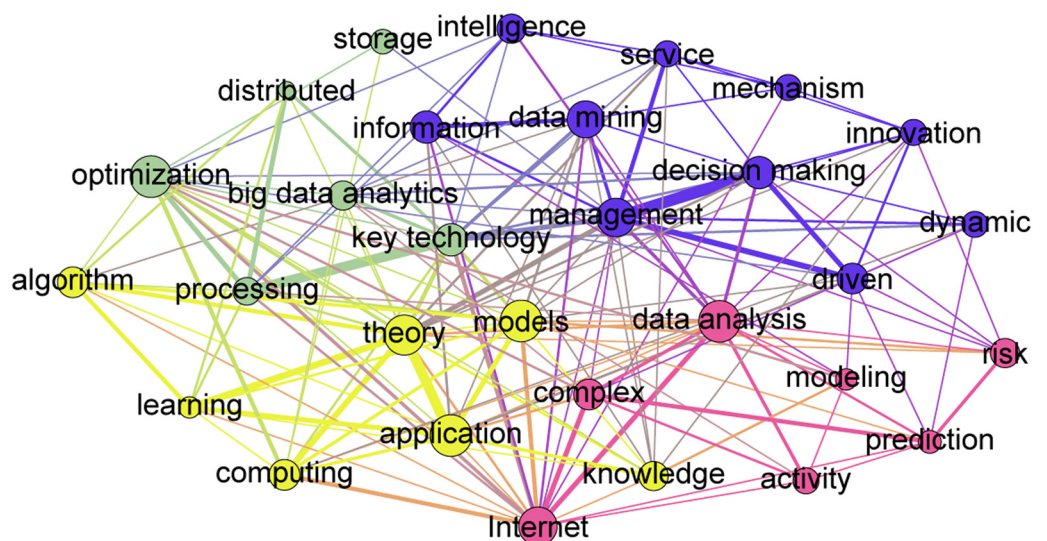


Fig 5. Semantic Network of 30 Most Frequently Occurring Terms in NSFC Big Data Proposals.

doi:10.1371/journal.pone.0154509.g005

- Blue group presents the problems and potential solutions in bioinformatics; key terms include: “high-dimensional data,” “bioinformatics,” “genomics,” “big data problems” and “big data analysis”;
- Orange group displays public management factors in the context of Big Data through keywords such as “big data management,” “complex systems,” “Internet,” “data analytics,” “analytics,” “government” and “data science”.

Following the above analytic logic, the 30 most frequently occurring terms in NSFC granted proposals were also clustered into four clusters (Fig 5):

- Orange cluster indicates the main theory and techniques in Big Data fields; the main phrases in this cluster are: “algorithm,” “machine learning,” “computing,” “models,” “theory,” “knowledge” and “application”;
- Blue cluster layouts highlight technologies involved with Big Data analysis; related phrases include: “optimization,” “processing,” “big data analytics,” “key technology,” “storage” and “distributed”;
- Green cluster presents the role of Big Data in complex environments through terms such as “Internet,” “complex,” “activity,” “data analysis,” “modeling,” “prediction” and “risk”;
- Purple cluster illustrates the use of Big in management practice, through key terms such as “information,” “data mining,” “management,” “intelligence,” “driven,” “decision making,” “knowledge,” “service,” “dynamic,” “mechanism” and “innovation”.

These results indicate that both NSF’s and NSFC’s proposals have considerable common research interests and foci on topics such as “algorithms,” “Internet,” “data analysis,” “data mining,” “decision making” and “machine learning. At the same time, there are distinctions in that the US concentrates more on engineering and specialized applications while China shows more emphasis on Big Data theory and concepts in management-related domains.

Based on the above proposal concentration analysis, we can see that the two countries share similar research topics for building up the Big Data field. However, since the two countries’ strategic goals and scientific development models may differ, the detailed concepts, techniques, and applications to which funding agencies pay attention are somewhat different.

Proposal Interdisciplinarity Analysis

IDR has become a significant thrust of NSF and other scientific agencies in recent years. Most bibliometric literatures on measuring IDR focus on examining the outputs of science—publications [16]. Following this traditional evolution, we extracted the grant numbers from NSF’s and NSFC’s Big Data awards and retrieved the papers acknowledging them from the Web of Science. The results show that 428 publications were supported by NSF Big Data awards and 644 publications by NSFC (Retrieved on Jan 3, 2016). As an initial way to understand the disciplinary orientation of these publications, we analyzed the Web of Science Categories (“WCs”), which are assigned to journals based on a combination of cross-citation patterns and editorial judgment. WCs offer a standard in bibliometric analysis for representing disciplinary or field participation [4]. Table 6 shows the Top 10 WCs for Big Data publications supported by NSF and NSFC. Not surprisingly, most of the funded papers are in the Computer Science and Engineering, Electrical & Electronic areas. Multidisciplinary sciences and interdisciplinary studies are other foci in the funded papers. NSFC funded papers also pay attention to some management issues such as Operations Research & Management Science.

Table 6. Top 10 WCs for Big Data Publications Supported by NSF and NSFC.

WCs	Records Supported by NSF	WCs	Records Supported by NSFC
Engineering, Electrical & Electronic	118	Computer Science, Artificial Intelligence	174
Computer Science, Information Systems	65	Engineering, Electrical & Electronic	147
Statistics & Probability	50	Computer Science, Information Systems	130
Computer Science, Software Engineering	49	Telecommunications	71
Computer Science, Artificial Intelligence	46	Computer Science, Software Engineering	53
Computer Science, Hardware & Architecture	34	Computer Science, Interdisciplinary Applications	51
Computer Science, Theory & Methods	34	Computer Science, Theory & Methods	48
Telecommunications	31	Engineering, Multidisciplinary	41
Multidisciplinary Sciences	26	Mathematics, Interdisciplinary Applications	39
Mathematical & Computational Biology	25	Operations Research & Management Science	39

doi:10.1371/journal.pone.0154509.t006

We are particularly interested in cross-field research knowledge transfer and overall discipline distribution. For such purposes the granularity of the WCs is effective—i.e., some 224 WCs differentiate sub-fields, so we have applied science overlay mapping to visualize these differences [56] in Fig 6. and Fig 7. Again, not surprisingly, Big Data papers from NSF or NSFC funding are dominated by Computer Science, followed by Math Methods. But the pattern of widespread engagement is remarkable, suggesting that Big Data research is not bottled up in a

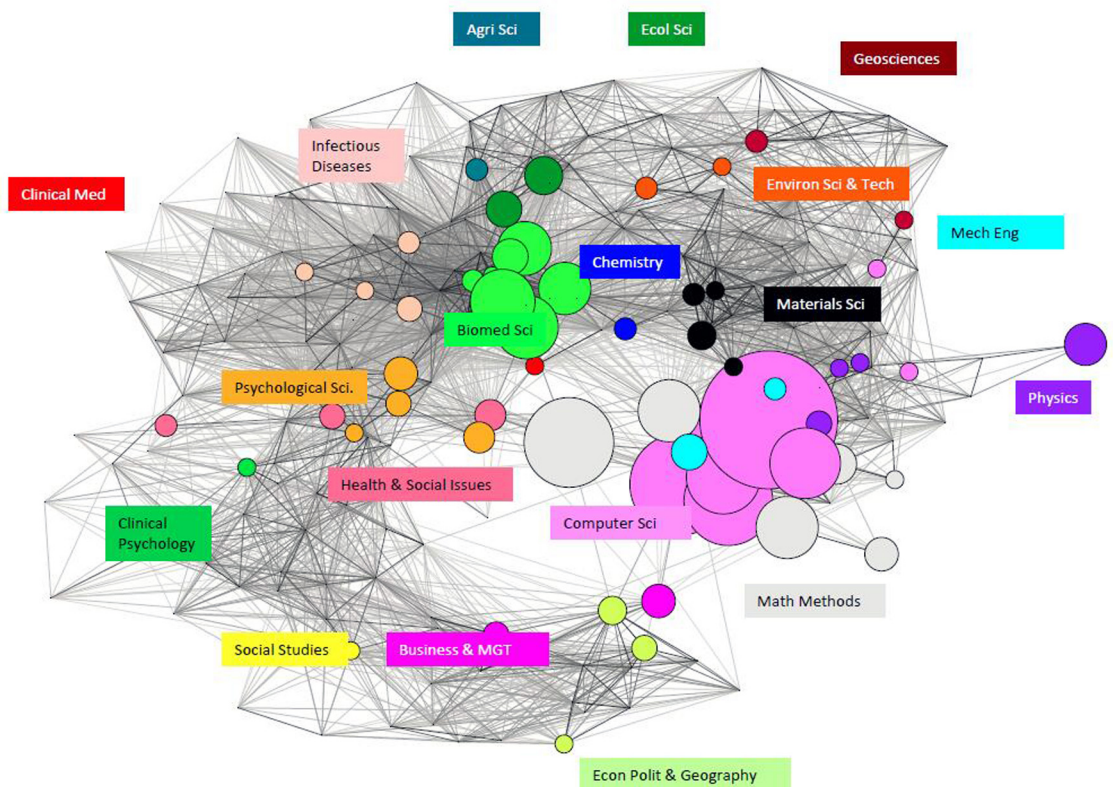


Fig 6. Big Data Research across the Disciplines of Publications Supported by NSF.

doi:10.1371/journal.pone.0154509.g006

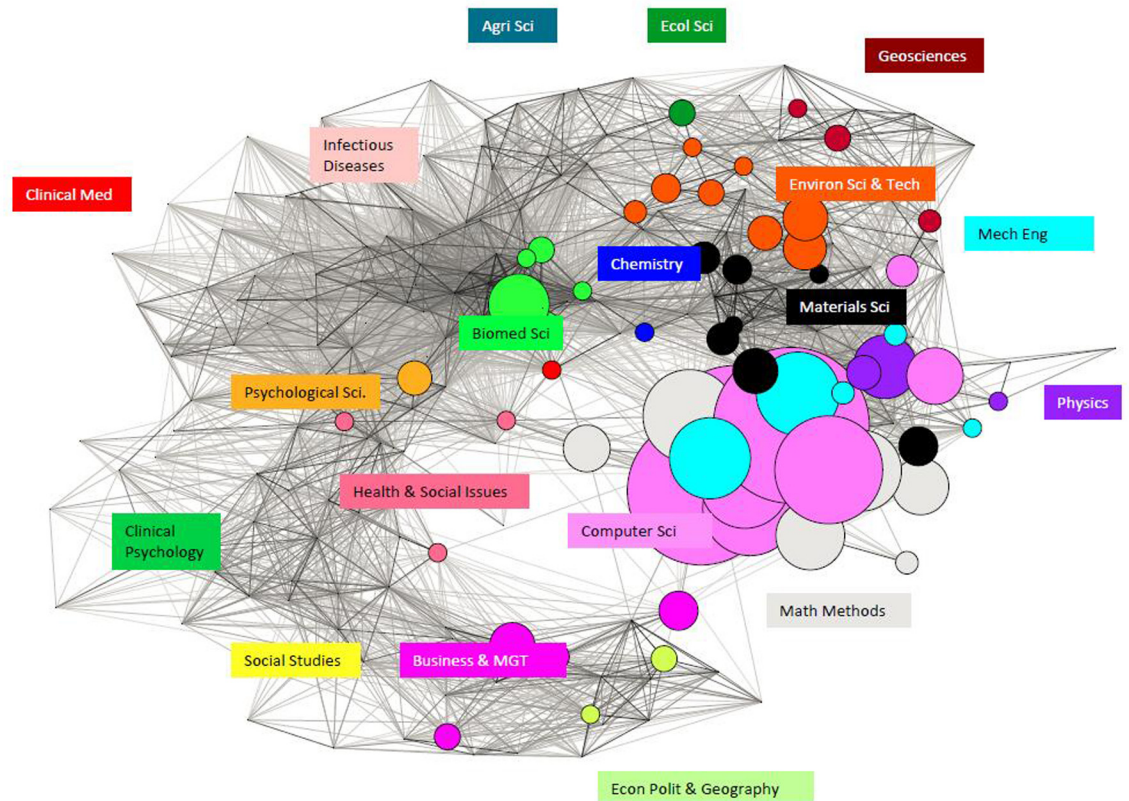


Fig 7. Big Data Research across the Disciplines of Publications Supported by NSFC.

doi:10.1371/journal.pone.0154509.g007

silos [38]. There are plenty of Big Data papers from sponsored research related to Biomedicine Science in the case of NSF and Environment Science & Technology in the case of NSFC.

Next, we examine interdisciplinarity from the view of funding information. For NSFC applications, each proposer is required to provide a main DAC and a secondary DAC. The main DAC is used to select peer reviewers, while the second DAC is used to provide alternatives in case the main DAC creates complications in the expert selection process. Since the DAC, to some extent, can greatly affect the final result, the applicants tend to use a single but familiar field rather than multiple fields in order to reduce the risk of being assigned to irrelevant experts outside their field [57]. The same situation occurs in the Big Data field and makes it very difficult to directly measure interdisciplinary attributes without sufficient information or indicators for cross-disciplinary engagement. We conducted a few personal interviews with the officials who worked in the NSFC, and were told that even though NSFC sensed that Big Data should be encouraged to be carried on through interdisciplinary or multidisciplinary research, currently, the Big Data projects granted by NSFC still tended to be more specialized in few particular disciplines, rather than being denoted as interdisciplinary or multidisciplinary studies.

Unlike NSFC, NSF provides one or more program tags to categorize the proposal. Additionally, NSF uses Program Element Codes (PEC) and Program Reference Codes (PRC) to track which NSF programs funded specific awards [58]. This approach to classifying proposals better encompasses the “messiness” of scientific inquiry and may provide a better unit of assessment for measuring interdisciplinary interaction and integration [17]. In Table 7, we find that 77.42% of proposals only offer one main PEC, which shows that most of the Big Data research

Table 7. Summary of Program Distribution of Big Data Projects in NSF.

Number of PECs or PRCs	Number of Awards with that # of PECs	% of NSF Awards	Number of Awards with that # of PRCs	% of NSF Awards
0	0	0.00	119	12.92
1	713	77.42	203	22.04
2	145	15.74	300	32.57
3	44	4.78	183	19.87
4	11	1.19	65	7.06
5	6	0.65	27	2.93
6	0	0.00	14	1.52
7	1	0.11	2	0.22
8	1	0.11	2	0.22
9	0	0.00	4	0.43
10	0	0.00	1	0.11
13	0	0.00	1	0.11

doi:10.1371/journal.pone.0154509.t007

belongs to a single research area, albeit the other 22.58% are associated with two or more programs. In addition, 65.04% of the proposals are associated with more than one PRC.

In order to measure the degree of interdisciplinarity for a single granted project, we introduce the S-IDR indicator to calculate the degree of interdisciplinarity based on these codes. Borrowing the main ideas from our previous work to calculate the specialization score that encompassed distinct components (Variety, Balance and Disparity), we built an improved diversity indicator inspired by the Hill-type indicator proposed by Zhang et. al. [15] that gave more weight to variety. Higher S-IDR scores represent less interdisciplinarity. The equation of S-IDR is:

$$S - IDR = \frac{1}{1 - S} \tag{2}$$

$$S = \frac{\sum (f_i \times f_j \times \cos(PEC_i - PEC_j))}{\sum (f_i \times f_j)} \tag{3}$$

In eqs (2) and (3), i and j represent the number of PECs, f_i and f_j are frequencies of PEC_i and PEC_j , $\cos(PEC_i - PEC_j)$ is the cosine measure of association between the two PECs to measure the similarity between two PECs, based on a co-occurrence frequency of pairs of PECs in historical NSF proposals. That is, if two PECs frequent appear together in the proposals, then the cosine measure will be relatively high.

This method of measuring IDR offers a possible way to capture the extent or magnitude of IDR in a given funding portfolio. To determine the interdisciplinarity degree for a certain year or research area, we take the average S-IDR values for the total number of projects belonging to corresponding year or research area. The average S-IDR in the past few years is shown in Fig 8. The figure indicates that the degree of interdisciplinarity has increased from 2011 to 2015. In 2011, the S-IDR value peaked, possibly because investigators localized their studies in the confined scope. During the period of 2014–2015, the curve showed the trend was becoming calm and steady, since there was a time lag between the issued date and the publishing date of proposals.

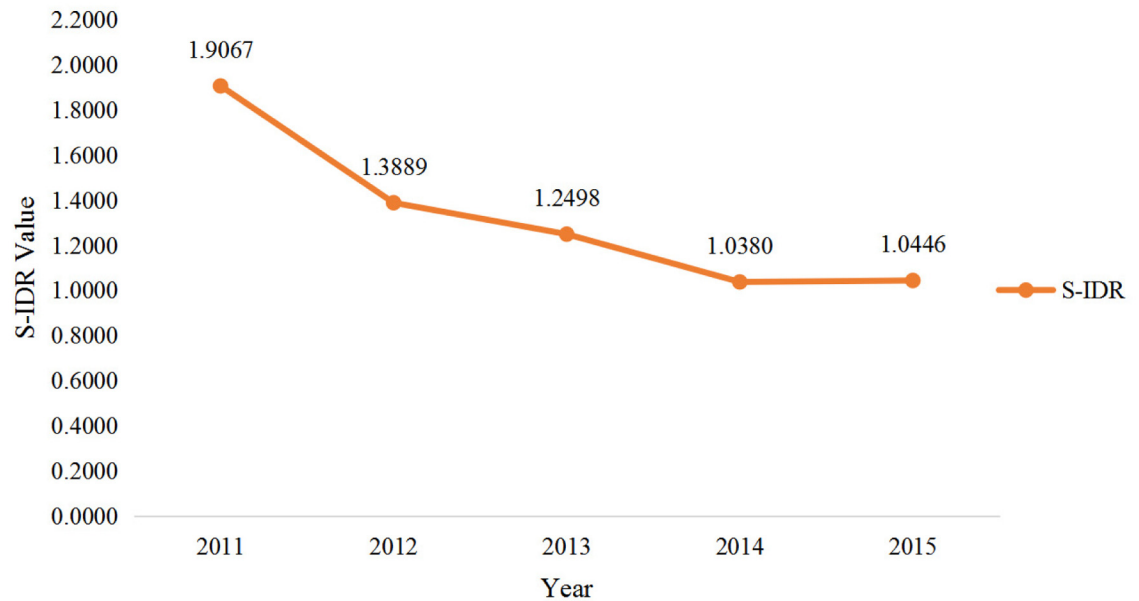


Fig 8. S-IDR Diversity Portfolio during the Period of 2011–2015.

doi:10.1371/journal.pone.0154509.g008

Conclusions and Discussion

This paper sought to understand how national scientific funding agencies support rapidly emerging interdisciplinary research. We focused on Big Data, a field that has extended out of the information technology area and grown at a remarkable pace from 2012 to 2015. We applied tech mining, interdisciplinarity evaluation, and network analysis techniques to retrieve and analyze information from funding proposals. In addition, we used a comparative perspective by contrasting the approach of NSF and NSFC to understand how these organizations have made adjustments in response to the rise of the Big Data area.

The proposal activity analysis concluded that both the number and the funding amount of awards are increasing in the US and China. The principal investigators in most awards come from academic organizations, but corporations and individuals from the US have a greater likelihood of being funded by NSF, while research institutes, such as the Chinese Academy of Science, are more apt to be funded by NSFC.

The proposal concentration analysis suggested that Information Science is the core area for Big Data in both the US and China. Engineering and Mathematical & Physical Sciences are two other important disciplines involved with Big Data in the US, while in China, Management Sciences and Earth Sciences are the two other most common disciplines. From a keyword-standpoint perspective, scholars from both the US and China share common research interests in certain Big Data techniques, e.g. “algorithm,” “modeling,” and “machine learning.” However, US proposal titles are more oriented toward engineering and specialized applications while China emphasizes theory and management concepts.

Our results provide some evidences of the notion that different funding agencies use different approaches to ramp-up funding for emerging technologies. Big Data projects granted by NSFC are specialized in particular disciplines rather than being tagged as interdisciplinary or multidisciplinary, but for NSF, the degree of interdisciplinarity is more pronounced. The ability to leverage funding in multiple areas is reflected in NSF’s approach, while NSFC’s method indicates more direct efforts to target resources.

The main contributions of this paper are two-fold. On the one hand, we explored the interdisciplinary issue by using funding data (research proposals) rather than publication or patent data, and presented information relating to engagement of academic individuals, networks, and the evolutionary process of science. We also addressed two different national funding agencies and analyzed two funding datasets to perform this comparison. Such contrasting research methods add a layer of understanding to the diversity of approaches that countries can apply to address rapidly emerging ST&I areas.

In this paper, language distinctions and funding system differences complicated the possibility of a detailed, systematic comparison, and the lack of detailed funding information from NSFC served as a barrier to further comparative analysis with NSF. This imbalanced information should be kept in mind when pursuing such approaches. Building on the logic and methods in this paper, future research could pursue the question of how to further assess the relationships between research funding and research output, and, considerably more challenging, track the interactions between funding and innovation.

Supporting Information

S1 Table. The 87 Research Areas and Their Corresponding Codes of NSFC.

(DOCX)

S1 File. Raw Data of Big Data Proposals Funded by NSF and NSFC.

(XLSX)

Acknowledgments

We acknowledge support from the US National Science Foundation (Award NO. 1527370), Chinese National Programs for High Technology Research and Development (Grant NO. 2014AA015105), and the General Program of National Natural Science Foundation of China (Grant NO. 71373019). Besides, we are grateful for the scholarship provided by the China Scholarship Council (CSC Student ID 201406030005). The authors thank Zhinan Wang and anonymous reviewers for their very helpful comments and suggestions. The findings and observations contained in this paper are those of the authors and do not necessarily reflect the views of the supporters or the sponsors.

Author Contributions

Conceived and designed the experiments: YH AP. Performed the experiments: YH YZ. Analyzed the data: YH XW. Contributed reagents/materials/analysis tools: YH YZ JY AP XW. Wrote the paper: YH YZ JY.

References

1. Noor A. Emerging interdisciplinary fields in the coming intelligence/convergence era. *Central European Journal of Engineering*. 2012; 2(3):315–24. doi: [10.2478/s13531-012-0014-4](https://doi.org/10.2478/s13531-012-0014-4)
2. Alexander J, Bache K, Chase J, Freyman C, Roessner JD, Smyth P, editors. An exploratory study of interdisciplinarity and breakthrough ideas. *Portland International Conference on Management of Engineering & Technology (PICMET '13)*; 2013 July 28-August 1; San Jose, CA.
3. The National Academies. *Facilitating Interdisciplinary Research*. Washington, DC: The National Academies Press; 2004. 332 p. doi: [10.17226/11153](https://doi.org/10.17226/11153)
4. Porter AL, Roessner JD, Cohen AS, Perreault M. Interdisciplinary research: meaning, metrics and nurture. *Research Evaluation*. 2006; 15(3):187–95. doi: [10.3152/147154406781775841](https://doi.org/10.3152/147154406781775841) PubMed PMID: WOS:000244345300004.

5. Stirling A. A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*. 2007; 4(15):707–19. doi: [10.1098/rsif.2007.0213](https://doi.org/10.1098/rsif.2007.0213) PubMed PMID: WOS:000247907100012.
6. Shannon CE, Weaver W. *The mathematical theory of communication*: University of Illinois Press; 1959.
7. Simpson EH. Measurement of diversity. *Nature*. 1949; 163:688. doi: [10.1038/163688a0](https://doi.org/10.1038/163688a0)
8. Rao CR. Diversity: Its measurement, decomposition, apportionment and analysis. *Sankya: The Indian Journal of Statistics*. 1982; 44(1):1–22.
9. Rafols I, Meyer M. Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics*. 2010; 82(2):263–87. doi: [10.1007/s11192-009-0041-y](https://doi.org/10.1007/s11192-009-0041-y) PubMed PMID: WOS:000274111600005.
10. Porter AL, Cohen AS, Roessner JD, Perreault M. Measuring researcher interdisciplinarity. *Scientometrics*. 2007; 72(1):117–47. doi: [10.1007/s11192-007-1700-5](https://doi.org/10.1007/s11192-007-1700-5) PubMed PMID: WOS:000247139300009.
11. Porter AL, Roessner JD, Heberger AE. How interdisciplinary is a given body of research? *Research Evaluation*. 2008; 17(4):273–82. doi: [10.3152/095820208x364553](https://doi.org/10.3152/095820208x364553) PubMed PMID: WOS:000262359500005.
12. Porter AL, Rafols I. Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*. 2009; 81(3):719–45. doi: [10.1007/s11192-008-2197-2](https://doi.org/10.1007/s11192-008-2197-2) PubMed PMID: WOS:000271989600009.
13. Carley S, Porter AL. A forward diversity index. *Scientometrics*. 2012; 90(2):407–27. doi: [10.1007/s11192-011-0528-1](https://doi.org/10.1007/s11192-011-0528-1) PubMed PMID: WOS:000299088900005.
14. Leinster T, Cobbold CA. Measuring diversity: the importance of species similarity. *Ecology*. 2012; 93(3):477–89. doi: [10.1890/10-2402.1](https://doi.org/10.1890/10-2402.1) PubMed PMID: WOS:000302957300006. PMID: [22624203](https://pubmed.ncbi.nlm.nih.gov/22624203/)
15. Zhang L, Rousseau R, Glänzel W. Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account. *Journal of the Association for Information Science and Technology*. 2016; 67(5):1257–65. doi: [10.1002/asi.23487](https://doi.org/10.1002/asi.23487)
16. Wagner CS, Roessner JD, Bobb K, Klein JT, Boyack KW, Keyton J, et al. Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*. 2011; 5(1):14–26. doi: [10.1016/j.joi.2010.06.004](https://doi.org/10.1016/j.joi.2010.06.004) PubMed PMID: WOS:000285626000002.
17. Nichols LG. A topic model approach to measuring interdisciplinarity at the National Science Foundation. *Scientometrics*. 2014; 100(3):741–54. doi: [10.1007/s11192-014-1319-2](https://doi.org/10.1007/s11192-014-1319-2) PubMed PMID: WOS:000340569800009.
18. Zhang Y, Chen H, Zhang G, Porter AL, Zhu D, Lu J. Topic analysis and forecasting for science, technology and innovation: methodology with a case study focusing on big data research. *Technological Forecasting and Social Change*. 2016; 105:179–91. doi: [10.1016/j.techfore.2016.01.015](https://doi.org/10.1016/j.techfore.2016.01.015)
19. Lok C. Science funding: Science for the masses. *Nature*. 2010; 465(7297):416–8. doi: [10.1038/465416a](https://doi.org/10.1038/465416a) PubMed PMID: WOS:000278043700011. PMID: [20505707](https://pubmed.ncbi.nlm.nih.gov/20505707/)
20. Fitchett JR, Head MG, Cooke MK, Wurie FB, Atun R. Funding infectious disease research: a systematic analysis of UK research investments by funders 1997–2010. *PLoS One*. 2014; 9(8):e105722. doi: [10.1371/journal.pone.0105722](https://doi.org/10.1371/journal.pone.0105722) PubMed PMID: WOS:000340880900046. PMID: [25162631](https://pubmed.ncbi.nlm.nih.gov/25162631/)
21. Roebber PJ, Schultz DM. Peer review, program officers and science funding. *Plos One*. 2011; 6(4). doi: [10.1371/journal.pone.0018680](https://doi.org/10.1371/journal.pone.0018680) PubMed PMID: WOS:000289404700026.
22. Wu J. Distributions of scientific funding across universities and research disciplines. *Journal of Informetrics*. 2015; 9(1):183–96. doi: [10.1016/j.joi.2014.12.007](https://doi.org/10.1016/j.joi.2014.12.007) PubMed PMID: WOS:000349770500019.
23. Lane J. Assessing the impact of science funding. *Science*. 2009; 324(5932):1273–5. PubMed PMID: WOS:000266635100026.
24. Wang J, Shapira P. Is there a relationship between research sponsorship and publication impact? An analysis of funding acknowledgments in nanotechnology papers. *Plos One*. 2015; 10(2). doi: [10.1371/journal.pone.0117727](https://doi.org/10.1371/journal.pone.0117727) PubMed PMID: WOS:000350062300029.
25. Auranen O, Nieminen M. University research funding and publication performance—An international comparison. *Research Policy*. 2010; 39(6):822–34. doi: [10.1016/j.respol.2010.03.003](https://doi.org/10.1016/j.respol.2010.03.003) PubMed PMID: WOS:000278943600011.
26. Fortin J-M, Currie DJ. Big Science vs. Little Science: How scientific impact scales with funding. *Plos One*. 2013; 8(6). doi: [10.1371/journal.pone.0065263](https://doi.org/10.1371/journal.pone.0065263) PubMed PMID: WOS:000322361200013.
27. Wan Q, Xu Y, Dong E. Diabetic nephropathy research in China: Data analysis and review from the National Natural Science Foundation of China. *Journal of Diabetes*. 2015; 7(3):307–14. doi: [10.1111/1753-0407.12265](https://doi.org/10.1111/1753-0407.12265) PubMed PMID: WOS:000353502500004. PMID: [25565189](https://pubmed.ncbi.nlm.nih.gov/25565189/)

28. National Science Foundation of China. Part X: Other work of NSFC 2013 [cited 2015 May 18]. Available from: <http://www.nsf.gov.cn/Portals/0/fj/english/fj/report/2013/pdf/010.pdf>.
29. Poo M-m, Wang L, Yang W. A new face at Natural Science Foundation of China-An interview with NSFC President Wei Yang. *National Science Review*. 2014; 1(1):157–60. doi: [10.1093/nsr/nwt031](https://doi.org/10.1093/nsr/nwt031) PubMed PMID: WOS:000356854800027.
30. Nadkarni NM, Stasch AE. How broad are our broader impacts? An analysis of the National Science Foundation's Ecosystem Studies Program and the broader Impacts requirement. *Frontiers in Ecology and the Environment*. 2013; 11(1):13–9. doi: [10.1890/110106](https://doi.org/10.1890/110106) PubMed PMID: WOS:000314483100015.
31. Kamenetzky JR. Opportunities for impact: Statistical analysis of the National Science Foundation's broader impacts criterion. *Science and Public Policy*. 2013; 40(1):72–84. doi: [10.1093/scipol/scs059](https://doi.org/10.1093/scipol/scs059) PubMed PMID: WOS:000315071900009.
32. National Science Foundation. Introduction to interdisciplinary research [cited 2015 May 15]. Available from: https://www.nsf.gov/od/oa/additional_resources/interdisciplinary_research/index.jsp.
33. Hogarth RM, Soyer E. Using simulated experience to make sense of Big Data. *Mit Sloan Management Review*. 2015; 56(2):49–54. PubMed PMID: WOS:000348635600010.
34. McAfee A, Brynjolfsson E. Big Data: The management revolution. *Harvard Business Review*. 2012; 90(10):60–6, 8, 128. PubMed PMID: WOS:000309093300026. PMID: [23074865](https://pubmed.ncbi.nlm.nih.gov/23074865/)
35. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, et al. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute; 2011.
36. Office of Science and Technology Policy. Obama administration unveils “Big Data” initiative: announces \$200 million in new R&D investments 2012 [cited 2015 May 18]. Available from: https://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf.
37. Liu Y, He J, Guo M, Yang Q, Zhang X. An overview of Big Data industry in China. *China Communications*. 2014; 11(12):1–10. PubMed PMID: WOS:000347669800001.
38. Porter AL, Huang Y, Schuehle J, Youtie J. Meta data: Big Data research evolving across disciplines, players, and topics. *Proceedings of the 2015 IEEE International Congress on Big Data*; New York: IEEE Computer Society; 2015. p. 262–7.
39. Porter AL, Cunningham SW. *Tech Mining: Exploiting new technologies for competitive advantage*: John Wiley & Sons, Inc.; 2005.
40. Porter AL, Kongthon A, Lui JC. Research profiling: Improving the literature review. *Scientometrics*. 2002; 53(3):351–70. doi: [10.1023/a:1014873029258](https://doi.org/10.1023/a:1014873029258) PubMed PMID: WOS:000175172000006.
41. Hicks D, Tomizawa H, Saitoh Y, Kobayashi S. Bibliometric techniques in the evaluation of federally funded research in the United States. *Research Evaluation*. 2004; 13(2):78–86. doi: [10.3152/147154404781776446](https://doi.org/10.3152/147154404781776446) PubMed PMID: WOS:000224644200001.
42. Guo Y, Huang L, Porter AL. The research profiling method applied to nano-enhanced, thin-film solar cells. *R & D Management*. 2010; 40(2):195–208. doi: [10.1111/j.1467-9310.2010.00600.x](https://doi.org/10.1111/j.1467-9310.2010.00600.x) PubMed PMID: WOS:000274389700007.
43. Porter AL, Schoeneck DJ, Roessner D, Garner J. Practical research proposal and publication profiling. *Research Evaluation*. 2010; 19(1):29–44. doi: [10.3152/095820210x492512](https://doi.org/10.3152/095820210x492512) PubMed PMID: WOS:000277986700003.
44. Tian W, Vogel D, Ma JA, Gu JB. IS-Supported managerial control for China's research community: An agency theory perspective. *Journal of Global Information Management*. 2010; 18(4):53–81. doi: [10.4018/jgim.2010100104](https://doi.org/10.4018/jgim.2010100104) PubMed PMID: WOS:000282228300004.
45. Editorial. Community cleverness required. *Nature*. 2008; 455(7209):1-. doi: [10.1038/455001a](https://doi.org/10.1038/455001a) PMID: [18769385](https://pubmed.ncbi.nlm.nih.gov/18769385/)
46. Huang Y, Schuehle J, Porter AL, Youtie J. A systematic method to create search strategies for emerging technologies based on the Web of Science: illustrated for 'Big Data'. *Scientometrics*. 2015; 105(3):2005–22. doi: [10.1007/s11192-015-1638-y](https://doi.org/10.1007/s11192-015-1638-y) PubMed PMID: WOS:000365130100034.
47. Porter AL, Guo Y, Chiavatta D. Tech mining: text mining and visualization tools, as applied to nano-enhanced solar cells. *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery*. 2011; 1(2):172–81. doi: [10.1002/widm.7](https://doi.org/10.1002/widm.7) PubMed PMID: WOS:000304257500007.
48. Guo Y, Ma T, Porter AL, Huang L. Text mining of information resources to inform Forecasting Innovation Pathways. *Technology Analysis & Strategic Management*. 2012; 24(8):843–61. doi: [10.1080/09537325.2012.715491](https://doi.org/10.1080/09537325.2012.715491) PubMed PMID: WOS:000308075800008.
49. National Science Foundation. Big Data research initiative [cited 2015 May 17]. Available from: <http://www.nsf.gov/cise/news/bigdata.jsp>.

50. Wang C, Chen W, Zhang J, Song S, Lu R. Challenging scientific problems for technologies and applications of Big Data. *Bulletin of National Natural Science Foundation of China*. 2014;(2:):92–8.
51. Gaughan M, Bozeman B. Using curriculum vitae to compare some impacts of NSF research grants with research Center funding. *Research Evaluation*. 2002; 11(1):17–26. doi: [10.3152/147154402781776952](https://doi.org/10.3152/147154402781776952) PubMed PMID: WOS:000179389900002.
52. Halffman W, Leydesdorff L. Is inequality among universities increasing? Gini coefficients and the elusive rise of elite universities. *Minerva*. 2010; 48(1):55–72. doi: [10.1007/s11024-010-9141-3](https://doi.org/10.1007/s11024-010-9141-3) PubMed PMID: WOS:000276435400004. PMID: [20401157](https://pubmed.ncbi.nlm.nih.gov/20401157/)
53. Park HW, Leydesdorff L. Decomposing social and semantic networks in emerging "big data" research. *Journal of Informetrics*. 2013; 7(3):756–65. doi: [10.1016/j.joi.2013.05.004](https://doi.org/10.1016/j.joi.2013.05.004) PubMed PMID: WOS:000323859700021.
54. Zhang Y, Porter AL, Hu Z, Guo Y, Newman NC. "Term clumping" for technical intelligence: A case study on dye-sensitized solar cells. *Technological Forecasting and Social Change*. 2014; 85:26–39. doi: [10.1016/j.techfore.2013.12.019](https://doi.org/10.1016/j.techfore.2013.12.019) PubMed PMID: WOS:000337553700004
55. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008; 2008(10):P10008. doi: [10.1088/1742-5468/2008/10/P10008](https://doi.org/10.1088/1742-5468/2008/10/P10008) PubMed PMID: WOS:000260529900010.
56. Rafols I, Porter AL, Leydesdorff L. Science overlay maps: A new tool for research policy and library management. *Journal of the American Society for Information Science and Technology*. 2010; 61(9):1871–87. doi: [10.1002/asi.21368](https://doi.org/10.1002/asi.21368) PubMed PMID: WOS:000281011400011.
57. Wu J, Jin M, Ding X-H. Diversity of individual research disciplines in scientific funding. *Scientometrics*. 2015; 103(2):669–86. doi: [10.1007/s11192-015-1549-y](https://doi.org/10.1007/s11192-015-1549-y) PubMed PMID: WOS:000352995000018.
58. National Science Foundation. Award search help [cited 2015 May 15]. Available from: <http://www.nsf.gov/awardsearch/advancedSearchHelp.jsp>.