OXFORD UNIVERSITY PRESS | BioScience

Collaboration across time and space in LTER Network

Journal:	*BioScience*
Manuscript ID	BIOS-19-0308.R1
Manuscript Type:	Professional Biologist
Date Submitted by the Author:	n/a
Complete List of Authors:	Huang, Tian-Yuan; Fudan University, School of Life Science Downs, Martha; University of California Santa Barbara, Long Term Ecological Research Network Office, National Center for Ecological Analysis and Synthesis Ma, Jun; Fudan University, School of Life Science Zhao, Bin; Fudan University, School of Life Science
Key words:	LTER, data sharing, collaboration, social network analysis, Monte Carlo method
Abstract:	The scale of ecological research is getting larger and larger. At such scale, collaboration is indispensable, yet there is little consensus on what factors enable collaboration. In the present article, we had investigated the temporal and spatial pattern of institutional collaboration within the US LTER Network based on the bibliographic database. Social network analysis and Monte Carlo method were applied to identify the characteristics of papers published by LTER researchers within a baseline of papers from 158 leading ecological journals. It was found that long-term and long-distance collaboration are more frequent in the LTER Network, and the underlying mechanisms were investigated and discussed. We suggest that the maturing infrastructure and environment for collaboration within the LTER Network could encourage the scientists to make large-scale hypothesis and ask big questions in ecology, breaking the boundaries of time and space.

Collaboration across time and space in LTER Network

Tian-Yuan Huang[1], Martha R. Downs[2], Jun Ma[1], Bin Zhao[1*]

[1]Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, Coastal Ecosystems Research Station of the Yangtze River Estuary, and Shanghai Institute of EcoChongming (SIEC), Fudan University, Shanghai 200433, China
[2]Long Term Ecological Research Network Office, National Center for Ecological Analysis and Synthesis, University of California - Santa Barbara, USA

Authors & emails:
Tian-Yuan Huang; huang.tian-yuan@qq.com
Martha R. Downs; downs@nceas.ucsb.edu
Jun Ma; ma_jun@fudan.edu.cn
*Bin Zhao (corresponding author)
Email: zhaobin@fudan.edu.cn
Address: School of Life Sciences, Fudan University, 2005 Songhu Road, Shanghai, 200438, China
Phone: 021-51630688

Data available at:

Huang, T., B. Zhao, and M.R. Downs. 2020. Institutional Collaboration in the US LTER Network based on bibliometric information (1981-2018) ver 1. Environmental Data Initiative.

https://doi.org/10.6073/pasta/1e1402b9ef8680e14d1c392c4e6a5f4a. Accessed 2020-03-04.

Abstract

The scale of ecological research is getting larger and larger. At such scale, collaboration is indispensable, yet there is little consensus on what factors enable collaboration. In the present article, we had investigated the temporal and spatial pattern of institutional collaboration within the US LTER Network based on the bibliographic database. Social network analysis and Monte Carlo method were applied to identify the characteristics of papers published by LTER researchers within a baseline of papers from 158 leading ecological journals. It was found that long-term and long-distance collaboration are more frequent in the LTER Network, and the underlying mechanisms were investigated and discussed. We suggest that the maturing infrastructure and environment for collaboration within the LTER Network could encourage the scientists to make large-scale hypothesis and ask big questions in ecology, breaking the boundaries of time and space.

Key words: collaboration, data sharing, LTER, Monte Carlo method, social network analysis

Introduction

Answering fundamental scientific questions in ecology now requires experiments and observations at greater temporal and spatial scales than ever before, making scientific collaboration indispensable. Early ecologists typically worked on specific problems as individuals or small teams. The Long Term Ecological Research (LTER) Program was established by the National Science Foundation (NSF) in 1980 — with long term collaboration among scientific teams as one of its goals. For nearly 40 years, the LTER Program has supported long term ecological research at a wide variety of sites and made fundamental contributions to general theory in community and ecosystem ecology (Hobbie 2003, Kominoski et al. 2018), as well as revealing key mechanisms of evolutionary biology and social sciences (Brodersen and Seehausen 2014, Gragson and Grove 2006, Redman et al. 2004). The International Long Term Ecological Research (ILTER) Network, of which the US LTER Network is now a member, was founded in 1993 during the US LTER All Scientists Meeting at Estes Park, Colorado. ILTER unites regional networks and local sites all around the world, enabling sharing, comparison, and synthesis of site-based long-term ecological observations to capture changes in ecological process at the global scale (Haase et al. 2018, Kim 2006, Trajanov et al. 2019, Vanderbilt and Gaiser 2017).

The concept of “big ecology” — the data-intensive science of ecological systems — has become prevalent in recent decades (Coleman 2010, Soranno and Schimel 2014). But big ecology requires long-term research, and long-term research cannot function without collaboration. The US LTER program emphasized collaboration at an early stage. In designing the program, special consideration was given to defining five core areas that were seen as foci for collaboration: primary production, dynamics of population, transport of organic matter, movements of inorganic matter, and disturbance (Callahan 1984). In addition, LTER researchers have consistently worked to ensure a supportive environment for cooperation among ecologists. For instance, the US LTER Network leadership developed a strategic plan in 2002, which defined research priorities and developed new goals for the coming decade, known as the "Decade of Synthesis" (https://lternet.edu/wp-content/uploads/2010/12/lter_2010.pdf). The EcoTrends project compiled common LTER data in a consistent format to enable the use and synthesis of long-term data (Peters et al. 2013). Other efforts include developing standards for meteorological measurements at LTER sites, holding triennial All-Scientists' Meetings and annual Science Council meetings, developing a network-wide data vocabulary, and multiple other improvements to the data management of the LTER Network (Greenland 1986, Müller et al. 2010, O'Brien et al. 2016, Servilla et al. 2016).

In the new era of big data, computer science and information technology are changing the modes of cooperation and innovation, bringing both opportunities and challenges to large-scale collaboration in ecological science (Hampton et al. 2013, Soranno and Schimel 2016, Farley et al. 2018). Using social network analysis, Johnson et al. (2010) found that individual US LTER sites evolved from isolated individuals to multiple connected groups and finally formed a large collaborative network. In our study, we explores whether participation in LTER research accelerates cross-institutional collaboration and compares LTER collaboration to that found in the broader ecological community. Here, the term "institution" means research units including universities and research institutes. We focus on multi-institutional collaboration because of its potential to benefit the involved institutions and accelerate dissemination of knowledge and ideas (Ye et al. 2012).

To establish a basis for comparison, we built a bibliometric database covering research articles published in 158 leading ecological journals according to ISI Journal Citation Reports and distinguished LTER papers published from 1981 to 2018. We compared LTER with the general corresponding ecological research and focused on three questions in our investigation: (1) What are the general patterns of collaboration among institutions within the US LTER Network? (2) Did the LTER program improve long-term collaboration among institutions and why? (3) Did the LTER program improve long-distance collaboration among institutions and why?

Establishing the databases

To build a comprehensive database of ecological literature, we referred to ISI Journal Citation Reports and retrieved the journal list under the "ecology" category. According to the journal list, we extracted all the available literature information from the Scopus database (https://www.scopus.com). We removed the incomplete entries and merged duplicate publications to produce at a bibliometric database of all research articles from 158 leading ecology journals (other document types such as review or meeting paper are excluded). A full list of selected journals is available in Table S1. This database serves as the source of LTER articles as well as the baseline for comparison (EBD, Figure 1).

A crucial part of database establishment was identifying LTER publications within the overall database. We narrowed the scope to peer-reviewed articles, and used title matching to extract the target papers. The LTER bibliography information was provided by LTER Network Office (LNO), and can be viewed at the Zotero library (https://www.zotero.org/groups/2055673/lter_network/items). It should be noted that we were extracting LTER publications from the ecological literature database, therefore LTER articles published in journals outside of our selected journal list (Table S1) are excluded from our investigation. After data cleaning, we retrieved the literature information of 4028 LTER articles published between 1981 and 2018. (The LTER program was established in 1980 and 2018 was the last complete year of data available for this analysis.) A full list of the selected LTER articles with basic information can be found in Table S2. Although the selected articles represent only about 25% of the articles published with LTER support in the same time period, we believe that it could well reflect the overall pattern of scientific activities, including collaboration, within the academic community of the US LTER Network. The databases used in this study include Ecological Bibliometric Database (EBD), LTER Bibliometric Database (LBD) and LTER Ecological Bibliometric Database (LEBD). Relationships of these bibliometric databases are displayed in Figure 1. All data processing procedures were conducted in R 3.6.1 (R Core Team 2019), using the "tidyverse" package (Wickham 2016) as a comprehensive tool for data arrangement.

Construction of institutional collaboration network

Social network analysis (SNA) has long been used to quantify and visualize scientific collaboration in academia (Newman 2001, 2004). While research collaboration can take multiple forms (Katz and Martin 1997), by specifying different definitions of node and edge in the network, SNA can be applied to explore patterns and mechanisms of scientific collaboration under various scenarios (Gazni et al. 2012). The basic unit of scientific collaboration is the individual researcher; however, institutional policies and cultures also facilitate or constrain this behavior (Koseoglu 2016, Melin and Persson 1996). By focusing on institutions, we aim to reveal collaboration patterns at a larger scale and provide guidance for LTER and other collaborative program management from a different perspective. The whole network construction process could be largely divided into three steps: (1) Recognize the affiliations of each article; (2) Establish an edge list linking every two co-occurring affiliations in the sample articles; (3) Construct the undirected graph based on the established edge list. In our study, data retrieved from Scopus database API (https://dev.elsevier.com/sc_apis.html) via r-package 'rcopus' (Muschelli 2018) have attached IDs for each author and affiliation, which improves positive identification of authors and institutions. If two or more institutions co-occur in the same paper as author affiliations, these institutions are regarded as cooperating, therefore forming a collaboration network with nodes representing institutions joined by edges representing publications. We conducted our network analysis in R 3.6.1 (R Core Team 2019) using the "igraph" (Csardi and Nepusz 2006), "tidygraph"(Pedersen 2018) and "multinets"(Crepalde 2019) package.

Identification of collaboration patterns in LTER ecological studies among the general corresponding ecological studies (baseline) is a central focus of our work. To make an objective comparison, factors including sample size, publication year, research focus and author nationality should all be taken into consideration. For instance, LTER research might take greater interest in community and ecosystem ecology, so that simply using the whole EBD for a baseline might lead to a comparison of macroecology studies with non-macroecology, rather than focusing on the effects of LTER program. As different journals have specific aims and scopes, here we have restricted the journal selection to avoid this bias. Likewise, authors from the same period and nation might be more likely to cooperate. To avoid this complication, we made additional restrictions on the publication year and first author nationality. Interestingly, though US LTER mainly serves American scientists, we found many foreign scientists utilizing this platform and publishing papers as first authors, including examples from Canada (54), Australia (50), UK (48), China (37), Germany (35), etc. Moreover, the sample size may introduce another type of bias. Therefore, here we adopted a Monte Carlo method to form the baseline by making repeated simulations. For each simulation, we chose the same number of articles from EBD, using the same journals and the same specific years. That is to say, if we found 5 articles with first authors from the US published by journal J in 2001 in LEBD, then in every simulation we would randomly select 5 articles with an American as the first author, from journal J, published in 2001. Since the scale of computation is large, 99 simulations were performed to balance the test performance and computational burden. The LTER group was merged with the 99 simulations to form the final baseline group with 100 samples with the same distribution of journals, publication years, and first author nationality. For any calculated statistics, if the LTER group was much higher than the mean of baseline, then there was strong support that the statistical indicator of the LTER group was higher than that of the general baseline. This method tests whether LTER has had a significant effect on the state of collaboration.

Evolution of institutional collaboration in LTER Network

In order to get a general view of collaboration in LTER Network, we used the common bibliometric indicators, average author number and average institution number. These indicators reveal the extent of cooperation in scientific research. First, we calculated the average number of authors per paper and average number of unique institutions per paper at the annual scale for LEBD, then we did the same calculation on EBD within the same period (1981-2018) to form the baseline group. We found that the LTER group had more authors per paper than the baseline group (Figure 2a). This suggests that long-term ecological research demands more collaborating scientists than do general ecological studies. Surprisingly, not until the second decade did the LTER Network exhibit greater collaboration among institutions (Figure 2b). Collaboration in academic communities has increased recently, likely because collaboration is deemed to enhance research productivity, increase scientific impact, and help solve complicated multidisciplinary problems (Hara et al. 2014, Lariviere et al. 2014, Lee and Bozeman 2005). Our findings show that the extent of cooperation in ecology is increasing over the study period, and researchers who participate in long-term ecological studies are becoming even more cooperative (the gap between the LTER group and the baseline group increases over time in both Figure 2a and Figure 2b).

When the LTER network launched in 1980, surface mail and telephone were the primary forms of communication, restricting the potential for inter-institutional collaboration. As email, online journals, videoconferencing, and other tools for collaborative analysis have developed, these barriers have eased. At the same time, institutional norms have also changed, raising expectations that researchers will address large-scale ecological problems (Müller et al. 2010). Ecologists involved in the LTER program may have been on the leading edge of this trend both because long-term study demands collaboration to break the boundaries of time and space and because the program was structured to support it. For instance, LTER All Scientists' Meetings (ASM), initiated at Cedar Creek LTER site in 1985, were held only twice during the 1990s. But since 2000, the LTER ASM has become a routine gathering for scientists to exchange their perspectives and experience every three years. (https://lternet.edu/lter-all-scientists-meetings/).

To understand the evolution of cooperation structure among institutions in the LTER Network we apply social network analysis to visualize and quantify changes in collaboration patterns over time. First, we constructed the institutional collaboration network. Only LEBD was used in this process. As every LTER article contains information on the researchers' institutions, we could discern which institutions had contributed to each article. Two distinct institutions constitute a pair of cooperation relations. For instance, if institutions A, B, and C had cooperated to conduct the research and publish an article, we would record their cooperation relationships as "A--B; B--C; A--C". Using the published year of these articles, we constructed the institutional collaboration networks year by year from 1981 to 2018. Site information is also important for better understanding collaboration patterns, thus we have also displayed them in the visualization. If researchers from institution A has published an article with support of site X, then institution A would be linked to site X. In the visualization (Figure 3) based on multilevel network analysis (Lazega and Snijders

2015), institutions and sites are displayed as nodes in different forms, while cooperation relationships between institutions, relationships between sites and site-institution relationships are displayed as edges. Inspired by the previous research on network analysis of LTER collaboration (Johnson et al. 2010), we used two indicators from network science to measure the cohesion of collaboration networks: proportion of nodes in giant component and average degree of giant component. In network theory, a large group of nodes that are all connected to one another by paths of intermediate nodes is described as the giant component (Newman 2001). In such cases, the second-largest group of connected nodes is far smaller than the largest one. When it comes to our study, we wanted to know how many institutions were connected directly or indirectly to the largest group, which provides a measure of how completely the institutional network is connected. On the other hand, the degree of a node is the number of edges connected to it. While the proportion of nodes in the giant component tells us what proportion of institutions were in the largest group, the average degree of institutional nodes in the giant component indicates how intensely those institutions are connected to each other.

In figure 3, we display institutional collaboration networks at 10-year intervals, namely in 1985, 1995, 2005 and 2015. Clearly, more institutions can be tracked over time because more sites had joined the network. According to the history of LTER (https://lternet.edu/network-organization/lter-a-history/), only 6 sites had been established by 1981. This number rises to 11 in 1985 and reaches 18 in 1995, and by the year of 2015 there are 26 active sites. More LTER sites draw more scientists from various institutions to carry out research together, frequently on the same LTER sites. We also found that the overall density of the institutional collaboration network increases with time (Figure 3). In 1985, only 3 institutions were connected to each other, while in 1995 most institutions form small clusters to carry out their research. By 2005, few institutions were working on their own, and a condensed giant cluster had formed by 2015 (Figure 4). One factor contributing to this phenomenon may be single institutions managing more than one site. For instance, Coweeta (CWT) and Georgia Coastal Ecosystems (GCE) LTER sites were both headquartered at the University of Georgia, whereas Santa Barbara Coastal (SBC) and Moorea Coral Reef (MCR) LTERs are both run by the University of California, Santa Barbara.

In the meantime, the average degree of the institutional collaboration network is also increasing (Figure 4), indicating increasing collaboration cohesion over time. The proportion of nodes in the giant component rose after a slight decrease in the early 90's, and has remained at a high level ever since. In the early stages of LTER, only a few sites were selected and supported by NSF. Once some collaborations were formed, the giant component made up a large share of the total number of institutions. As more sites joined, they effectively diluted the collaboration intensity until they became integrated and began publishing together. Once the number of sites achieved stability, the level of collaboration also stabilized. By the year 2005, nearly all institutions are connected to the giant component of the network (Figure 4).

Another important trend for the LTER Network is that many researchers from various backgrounds have participated in the program, making LTER research more diversified (Figure 5). To explore such pattern, we extracted journal articles from LBD and investigated the subdisciplines of these papers according to the sub-subject area of journals from the Scopus source list (https://www.scopus.com/sources, retrieved at August 24, 2019). In Figure 5, we found that "Ecology" and "Ecology, Evolution, Behavior and Systematics" have taken the lead in all periods, but in recent decades LTER research has also integrated diverse subdisciplines such as "Aquatic Science", "Environmental Chemistry", "Water Science and Technology", "Soil Science", "Global and Planetary Change", "Plant Science", "Earth-Surface Processes" and "Forestry". While these fields have their own specific research focuses, their domain knowledge could all serve the ecological and environmental management.

Collaboration across time

In this section, we wanted to know: does participation in the LTER network increase the duration of collaborations? To answer it, we hypothesized that the LTER program boosts the number of institutional collaborations and confirmed the hypothesis using the Monte Carlo method. First, we needed a way to measure the temporal length of cooperation among institutions. In the previous analysis we recorded the cooperating institution pairs in each year from 1981 to 2018. Each year in which two institutions appeared on the same publication was recorded as a single year of collaboration — regardless of the number of authors from that institution, the order of authorship, or additional shared publications in the given year. We summarized the occurrence number of the institution pairs in this period and used it as an indicator of the temporal length of cooperation. For instance, if we found authors from institution A and institution B had cooperated and published 5 articles together in 2001, 2005 and 2010 in our database, then the cooperation time of institutions A and B was recorded as 3 years. It is true that the research may have started before the year of publication, and some types of cooperation might not be recognized in the author list with institutional information. Nevertheless, the overall frequency with which authors from given institutions publish together offers a reliable indicator of the collaboration intensity among institutions.

Next, we calculated the cooperation time for every institution pair using the data in LTBD database, so as to get the cooperation time of all recorded institution pairs involved in the LTER program. Then we counted how many pairs had cooperated for at least 3 years or 5 years between 1981 and 2018 and used those metrics to test the effect of LTER on length of collaboration. Details about the establishment of the baseline Monte Carlo simulations can be found above in the section introducing the construction of the institutional collaboration networks.

In our investigation, we found more frequent long-term collaborations in the LTER group than in 99 random selections of comparable baseline groups of the same size (Figure 6). Overall, for the LTER group there were 1081 pairwise institutional collaborations that were active for at least three years during the investigated 28 years, while in general only 250 collaborations would occur within an identically sized baseline group of articles published in the same period. Moreover, in the LTER group we tracked 367 collaborations lasting at least five years, while the corresponding average number in the baseline group is 40. The LTER program has clearly promoted frequent and long-lasting collaboration among ecological researchers from different institutions. One reason is that researchers from different institutions have worked together on the same LTER sites, which makes further cooperation convenient. On average, 66.7% of the publications from pairwise institutional long-term collaboration (pairwise institutional collaboration for at least 5 years) came from the same LTER site. But even accounting for this dynamic, collaborations of 5 or more years are 3 times as frequent among LTER publications as among the baseline publications.

Another important factor is the long-lasting alliances developed by active researchers. In LEBD we tracked a pair of institutions collaborating for at least 16 years with a total of 23 publications. One author had coauthored 20 articles, while the researcher with the second highest count of cross-institution publications had contributed to 17 of them. On inspection, we found that these two authors coming from two different institutions had maintained a long-term collaboration since 1999. The full catalog of institutional collaboration relationships in the LTER program can be found in Table S3. For collaborations of three or more years, the author with the most cross-institutional publications contributes to 54.6% of the total number of cooperated articles. We infer that these high frequency authors play an important role in long-term collaboration in LTER program. Further investigation showed that authors contributing to collaborations of longer than 9 years had worked at more institutions during their academic careers than the average collaborator (4.36 v. 1.23 institutions, respectively). While these are highly productive researchers who collaborate frequently, it appears that their professional mobility also plays a role in sustaining long-term collaborations.

Despite the increasing ease of videoconferencing and data sharing, we hypothesize that geographic distance remains a consideration for potential collaborators. Among long-term (>=5 years) institutional collaborations, 3.4% were between institutions in the same city. Among all other collaborations, those in the same city made up just 2.6%. Although geographic distance does not appear to play a defining role in initiation of collaboration, a shorter physical distance does appear to promote longer-lasting collaborations.

While many factors affect long-term collaboration, the core motivation is the desire to answer scientific questions that require long-term accumulation of knowledge and diverse expertise. The keywords associated with long-term collaborative publications may shed some light on the nature of those questions. For institutional collaborations of 5 or more years, the top 10 popular research topics are "nitrogen," "climate change," "stream," "global change," "carbon," "biodiversity," "metabolism," "disturbance," "stable isotope," and "species richness." Clearly, long-term collaborations have addressed the research hotspots of global climate change and biodiversity conservation in different ecosystems. This is not surprising, as both topics require long term observation and experimentation. But the increased frequency of certain other keywords in collaborative papers suggests the core thematic areas that were established at the beginning of the LTER program have, in fact, facilitated — or at least nucleated — cross-site collaboration. From its inception, the LTER program has encouraged researchers to focus long-term datasets around 5 core research areas: primary productivity, populations, organic matter, inorganic nutrients, and disturbance (Callahan 1984). Explicitly intended as a way to encourage collaboration, the continued utility of the core areas has been a subject of debate as scientific focus has shifted to new topics. The increased prevalence of core area-related keywords (nitrogen, carbon, biodiversity, metabolism, disturbance, species richness) among collaborative papers implies that designating core themes for the network has successfully facilitated collaborations around those themes — but cannot answer the question of whether equally important collaborations might have developed around other themes without that guidance. The popular keywords in publications of collaboration in LTER Network can be found in Table S3.

Collaboration across space

Generally, we would expect the likelihood of co-publication to decrease as geographic distance increases, however, in the context of globalization young ecologists may experience collaboration that breaks geographic and socioeconomic barriers (Parreira et al. 2017). To address general problems in ecology, LTER has emphasized ecological research based at multiple sites on a large spatial scale. Many sites have participated in the Nutrient Network, the Long-Term Intersite Decomposition Team (LIDET), and the International Tundra Experiment (Baker et al. 2000). More recently, LTER researchers have led and participated in extensive efforts such as the Global Lakes Ecological Observatory Network (GLEON) and research collaboration networks such as the Urban Sustainability (UREx) and the Permafrost Carbon research networks. In our study, we want to get a clear view of whether the LTER program has promoted long-distance collaboration over time. For each record representing an article in the database, we calculated the greatest distance between collaborators, based on the cities in which institutions were located. If the authors came from the same institution or institutions in the same city, the longest distance would be zero. Multiple institutions from different cities would be arranged to form distinct city pairs. According to the longitude and latitude of these cities, we calculated the geodesic distance for the city pairs and extracted the longest distance to represent the spatial span of the collaboration. We divided the whole period into 4 parts, namely 1981-1990, 1991-2000, 2001-2010 and 2011-2018. As above, we used the Monte Carlo method to establish baselines for each of the 4 time periods. We found that the average maximum distance between collaborating institutions rose rapidly over the last four decades, led by the LTER program (Figure 7). In all 4 periods, the LTER group was the top outlier among 100 samples, indicating that LTER has promoted cooperation in ecological research among institutions from distant areas. This trend became even more clear in the recent periods, as the gap between the LTER group and the baseline group has grown over time.

Next, we wanted to know *how* the LTER program improved long-distance collaboration. We sorted the records according to the maximum distance of pairwise institutional collaboration for every paper (S4 Table) to identify the collaborations bridging the greatest distances. Articles resulting from longer-distance collaborations usually involved more authors and institutions than average. Focusing on the top 15 papers (by distance), we found that data is the bond linking researchers from institutions across the world. Two of the studies with the longest collaboration distance are related to the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project (Hudson et al. 2014, Hudson et al. 2017), which aims to collect data from scientists worldwide to produce a global database of terrestrial species' responses to human pressures (https://www.predicts.org.uk/). In these two papers, hundreds of researchers who contributed to the database were all listed as authors. One the one hand, it is a wonder that scientists who might never meet or directly communicate with one another could be banded together in an effort to predict biodiversity changes worldwide; on the other hand, these papers raise new questions about the ambiguous definition of authorship and incentive mechanisms for collaborators. Other reviews of global projects include CTFS-ForestGEO (Anderson

Teixeira et al. 2015), monitoring the response of forests to global change, and BioTIME (Dornelas et al. 2018), establishing a database containing time series data of species abundance at a global scale. Other examples of global public databases include CoRRE (Langley et al. 2018), FRED (Freschet et al. 2017), ForestGEO (Langley et al. 2018) and TRY (Pierce et al. 2017). LTER and other long-term networks, such as AmeriFlux, ForestGEO, CZO, ILTER, and NEON often form the foundations of these global databases, with individual, un-networked studies adding site diversity, geographic extent, and filling key gaps. The formalized data structures, thorough documentation, and public availability of LTER and other network data make them especially useful in this context and help explain the large difference in geographic extent between LTER and baseline publications.

Lessons learned

At the advent of the Big Data era, fewer and fewer researchers work on their own on a specific focused problem (Woodward et al. 2014). Large-scale and multidisciplinary research programs have become the norm, which demands more and better collaboration among ecologists. It is encouraging to find that ecological researchers are becoming more collaborative than ever before (Figure 3). The LTER Network is one of the programs leading the collaboration trend, breaking barriers of time and space to answer fundamental ecological questions at ever larger scales. The rich experience of the LTER program can lead us to better collaborations both in the LTER Network and in other large collaborative scientific programs, such as FLUXNET and Natura 2000.

An academic institution, in some sense, is an organizational form for scientific collaboration. Facilitated by geographical closeness, common policy, and shared culture within the organization, academic activities are conducted and managed among researchers within the same institutions at a lower cost. Scientists make their choice of working institution in consideration of various factors including academic prestige, financial support, working environment and family life; however, most ecologists carry out scientific research for common reasons: to satisfy their curiosity for nature and meet the needs of society. If an institution is regarded as a collaboration of sorts — gathering talents to teaching and research careers, then large programs like LTER are fostering collaborations of collaboration to address challenging large-scale research questions from a high-level perspective. Our investigation found that LTER sites serve as focal points for researchers from multiple institutions to communicate and cooperate with each other. While in early years, many researchers from each institution typically worked at only one site (Figure 3), in recent decades the collaboration network of institutions in LTER has become more and more cohesive (Figure 3, Figure 4). Johnson et al (2010) found that if two sites were managed by the same institution, they were more likely to form cross-site collaborations. In our study, we conjecture that the cooperation of researchers from two institutions working on the same site might potentially lead to additional collaborations beyond that first pair of sites. An inspiring case is the collaboration between Colorado State University and University of New Mexico. According to our data records, these two universities jointly published a paper supported by the Sevilleta LTER (SEV) in 1989. Since then, these two institutions have 6 more joint publications based on SEV in the 1990s, including a cross-site paper in 1997. In the meantime, researchers from Colorado State University published 3 papers based on the Konza Prairie (KNZ) LTER site, whereas researchers from University of New Mexico had no publications based on KNZ before 2000. From 2004 to 2018, however, we tracked 14 co-publications by the two universities that were supported by both KNZ and SEV. We speculate that the long-term cooperation between these two universities at the SEV site may have facilitated an expanding cross-site collaboration afterward.

In a more general sense, data sharing plays a vital role in LTER collaborations. By sharing data, ecological observations and experiments can be conducted and integrated regardless of the restrictions of time and space. Early in its development, the LTER Network began encouraging scientists to share their research data, placing it among a small group of pioneers to do so (Michener 2015). By the end of 2018, there were over 6,000 data packages contributed by LTER sites available on the Environmental Data Initiative repository (https://portal.edirepository.org/), which is the successor to the LTER Network Data Portal. Nevertheless, this achievement was not arrived at overnight, but in a stepwise process (Porter 2010). Early in the 1980s, it was hard for researchers to even know what data had been collected in the LTER Network. At the 1990 LTER All-Scientists Meeting, the LTER Coordinating Committee developed site data management policy guidelines, and the concept of data-sharing took root in researchers' hearts. As online data sharing expanded in 1994, LTER sites even embraced a competition among sites to make more data available (Porter 2010). Today, the quality of data receives as much attention as the quantity and greatly facilitates the re-use of data packages contributed by the LTER Network.

Two main reasons that researchers have sometimes preferred not to share their data are: (1) Logistical barriers made data sharing inconvenient; (2) They wanted to use data in their subsequent work without competition (Parr and Cummings 2005, Michener 2015). Duke (2006) argued that “technology and infrastructure are not the ultimate limiting factors for data sharing — the individual scientist is.” Recent progress in information technology and platforms such as DataONE (https://www.dataone.org/), Ecological Data Wiki (https://ecologicaldata.org/home) or even Github (https://github.com/) have greatly reduced logistical barriers to sharing data and analyses. In some quarters, reluctance to share data persists. In some cases, this is based on authorship expectations and in others on the belief that users cannot effectively interpret data without input from the original source. In our research, we find that long-term collaboration between institutions relies heavily on a cadre of individual scientists who nucleate collaborations through data synthesis or movement among sites and institutions. Factors such as geographic distance have relatively minor effects. Based on our analysis, a few types of policies emerge as likely to support robust collaboration: (1) continued emphasis on publishing accessible, well-documented data; (2) increased emphasis on expanding the use of tools to combine and analyze publicly available data in a documented and reproducible manner (3) the creation and maintenance of stable platforms for interaction; only in the LTER program’s third decade did sites reach their collaboration potential; (4) mobility — between sites and institutions — played a surprisingly large role in enhancing collaboration. Clearly, the quest for collaboration shouldn’t dictate major life choices for individual researchers, but similar effects might be achieved through graduate student training, temporary exchanges of personnel, sabbatical fellowships, and synthesis projects.

A final consideration — call it the friendship factor — cannot be ignored just because it is difficult to measure. LTER Network principal investigators have been meeting annually for decades, as have information managers. A large portion of the network gathers every 3 years at All Scientists' Meetings. These meetings — as well as online interactions — greatly reduce the activation energy required to initiate a collaboration and the risk in pursuing a collaboration with an untested colleague. But they can also limit those benefits to researchers who are already a part of the network. Improving social relations within scientific communities can help reduce collaboration costs and foster more novel and creative synthetic research, *if* new partners are consistently invited into the circle (AlShebli et al. 2018, Bercovitz and Feldman 2011) and provided with opportunities to develop and apply leadership skills. The LTER Network not only provides researchers with invaluable long-term ecological data, but also helps them identify potential collaborators from different backgrounds all over the world. As the infrastructure and environment of collaboration continue to mature, researchers are emancipated from concerns of research feasibility and manageability, thus more likely to make large-scale hypothesis and ask big questions. In the era of collaborative science, only by bringing scientists together and integrating their efforts across times and spaces, can we effectively and efficiently address the most challenging scientific questions in ecology and beyond.

Acknowledgement

The underlying US LTER Bibliographic Database could not have been assembled without the help of US LTER information managers and administrators, past and present. John Porter offered great help on data gathering. Corinna Gries gave helpful suggestions on data arrangement. Sheng-Qi Dai gave great suggestions on this research. Here we thank them with our most sincere gratitude. This study was supported by grant from the National Key Research and Development Program of China (Grant No. 2018YFD0900806) and the National Natural Science Foundation of China (Grant No. 31170450). MD's participation was supported by the US LTER Network Office through the National Science Foundation (Grant No. DEB-1545288).

Reference

Adams HD, Zeppel MJ, Anderegg WR, Hartmann H, Landhäusser SM, Tissue DT, Huxman TE, Hudson PJ, Franz TE, Allen CD. 2017. A multi-species synthesis of physiological mechanisms in drought-induced tree mortality. Nature ecology & evolution 1: 1285.

AlShebli, B.K., Rahwan, T. & Woon, W.L. 2018. The preeminence of ethnic diversity in scientific collaboration. Nature communications 9: 5163

Anderson Teixeira KJ, Davies SJ, Bennett AC, Gonzalez Akre EB, Muller Landau HC, Joseph Wright S, Abu Salim K, Almeyda Zambrano AM, Alonso A, Baltzer JL. 2015. CTFS-Forest GEO: a worldwide network monitoring forests in an era of global change. Global change biology 21: 528-549.

Baker KS, Benson BJ, Henshaw DL, Blodgett D, Porter JH, Stafford SG. 2000. Evolution of a multisite network information system: the LTER information management paradigm. BioScience 50: 963-978.

Bercovitz, J., and M. Feldman. 2011. The mechanisms of collaboration in inventive teams: Composition, social networks, and geography. Research Policy 40:81–93.

Brodersen J, Seehausen O. 2014. Why evolutionary biologists should get seriously involved in ecological monitoring and applied biodiversity assessment programs. Evolutionary Applications 7: 968-983.

Callahan JT. 1984. Long-Term Ecological Research. Bioscience 34: 363-367.

Coleman DC. 2010. Big ecology: the emergence of ecosystem science. Univ of California Press.

Crepalde N. 2019. multinets: Multilevel Networks Analysis. R package version 0.2.2.

Csardi G, Nepusz T. 2006. The igraph software package for complex network research. InterJournal, Complex Systems 1695: 1-9.

Dornelas M, Antao LH, Moyes F, Bates AE, Magurran AE, Adam D, Akhmetzhanova AA, Appeltans W, Arcos JM, Arnold H. 2018. BioTIME: A database of biodiversity time series for the Anthropocene. Global Ecology and Biogeography 27: 760-786.

Duke CS. 2006. Data: Share and Share Alike. Frontiers in Ecology & the Environment 4: 395.

Farley SS, Dawson A, Goring SJ, Williams JW. 2018. Situating Ecology as a Big-Data Science: Current Advances, Challenges, and Solutions. BioScience 68: 563-576.

Gazni A, Sugimoto CR, Didegah F. 2012. Mapping world scientific collaboration: Authors, institutions, and countries. Journal of the American Society for Information Science and Technology 63: 323-335.

Gragson T, Grove M. 2006. Social Science in the Context of the Long Term Ecological Research Program. Society & Natural Resources 19: 93-100.

Greenland D. 1986. Standardized Meteorological Measurements for Long-Term Ecological Research Sites. Bulletin of the Ecological Society of America 67: 275-277.

Haase P, et al. 2018. The next generation of site-based long-term ecological monitoring: Linking essential biodiversity variables and ecosystem integrity. Science of The Total Environment 613-614: 1376-1384.

Hampton SE, Strasser CA, Tewksbury JJ, Gram WK, Budden AE, Batcheller AL, Duke CS, Porter JH. 2013. Big data and the future of ecology. Frontiers in Ecology & the Environment 11: 156-162.

Hara N, Solomon P, Kim SL, Sonnenwald DH. 2014. An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration. Journal of the Association for Information Science & Technology 54: 952-965.

Hobbie JE. 2003. Scientific accomplishments of the long term ecological research program: An introduction. Bioscience 53: 17-20.

Hudson LN, Newbold T, Contu S, Hill SL, Lysenko I, De Palma A, Phillips HR, Senior RA, Bennett DJ, Booth H. 2014. The PREDICTS database: a global database of how local terrestrial biodiversity responds to human impacts. Ecology and evolution 4: 4701-4735.

Hudson LN, Newbold T, Contu S, Hill SL, Lysenko I, De Palma A, Phillips HR, Alhusseini TI, Bedford FE, Bennett DJ. 2017. The database of the Predicts (Projecting responses of ecological diversity in changing terrestrial systems) project. Ecology and Evolution 7: 145-188.

Johnson JC, Christian RR, Brunt JW, Hickman CR, Waide RB. 2010. Evolution of Collaboration within the US Long Term Ecological Research Network. Bioscience 60: 931-940.

Katz JS, Martin BR. 1997. What is research collaboration? Research policy 26: 1-18.

Kim E. 2006. Development, potentials, and challenges of the International Long-Term Ecological Research (ILTER) Network. Ecological Research 21: 788-793.

Kominoski JS, Gaiser EE, Baer SG. 2018. Advancing Theories of Ecosystem Development through Long-Term Ecological Research. BioScience 68: 554-562.

Koseoglu MA. 2016. Mapping the institutional collaboration network of strategic management research: 1980-2014. Scientometrics 109: 203-226.

Langley JA, Chapman SK, La Pierre KJ, Avolio M, Bowman WD, Johnson DS, Isbell F, Wilcox KR, Foster BL, Hovenden MJ. 2018. Ambient changes exceed treatment effects on plant species abundance in global change experiments. Global change biology 24: 5668-5679.

Lariviere V, Sugimoto C, Tsou A, Gingras Y. 2014. Team size matters: Collaboration and scientific impact since 1900. Journal of the Association for Information Science & Technology 66: 1323-1332.

Lazega E, Snijders TA. 2015. Multilevel network analysis for the social sciences: Theory, methods and applications. Springer.

Lee S, Bozeman B. 2005. The Impact of Research Collaboration on Scientific Productivity. Social Studies of Science 35: 673-702.

Letcher SG, Lasky JR, Chazdon RL, Norden N, Wright SJ, Meave JA, Pérez García EA, Muñoz R, Romero Pérez E, Andrade A. 2015. Environmental gradients and the evolution of successional habitat specialization: a test case with 14 Neotropical forest sites. Journal of Ecology 103: 1276-1290.

Li Y, Xu M, Zou X. 2006. Effects of nutrient additions on ecosystem carbon cycle in a Puerto Rican tropical wet forest. Global Change Biology 12: 284-293.

Lutz JA, Furniss TJ, Johnson DJ, Davies SJ, Allen D, Alonso A, Anderson Teixeira KJ, Andrade A, Baltzer J, Becker KM. 2018. Global importance of large-diameter trees. Global Ecology and Biogeography 27: 849-864.

Melin G, Persson O. 1996. Studying research collaboration using co-authorships. Scientometrics 36: 363-377.

Michener WK. 2015. Ecological data sharing. Ecological Informatics 29: 33-44.

Müller F, Baessler C, Frenzel M, Klotz S, Schubert H. 2010. Long-Term Ecosystem Research Between Theory and Application. Springer.

Muller Landau HC, Condit RS, Harms KE, Marks CO, Thomas SC, Bunyavejchewin S, Chuyong G, Co L, Davies S, Foster R. 2006. Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. Ecology Letters 9: 589-602.

Muschelli J. 2018. rscopus: Scopus Database 'API' Interface. R package version 0.6.6.

Newman ME. 2001. The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences of the United States of America 98: 404-409.

Newman ME. 2004. Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Science 101: 5200.

O'Brien M, Costa D, Servilla M. 2016. Ensuring the quality of data packages in the LTER network data management system. Ecological Informatics 36: 237-246.

Parr CS, Cummings MP. 2005. Data sharing in ecology and evolution: Trends in Ecology & Evolution. Trends in Ecology & Evolution 20: 362-363.

Parreira MR, Machado KB, Logares R, Diniz-Filho JAF, Nabout JC. 2017. The roles of geographic distance and socioeconomic factors on international collaboration among ecologists. Scientometrics 113: 1539-1550.

Pedersen TL. 2018. tidygraph: A Tidy API for Graph Manipulation. R Package Version 1.1.0.

Peters D, Laney CM, Lugo AE, Collins SL, Driscoll CT, Groffman PM, Grove JM, Knapp AK, Kratz T, Ohman M, Waide RB., Yao J. 2013. Long-Term Trends in Ecological Systems: A Basis for Understanding Responses to Global Change. U.S. Department of Agriculture, Agricultural Research Service, Technical Bulletin Number 1931: 378.

Pierce S, Negreiros D, Cerabolini BE, Kattge J, Díaz S, Kleyer M, Shipley B, Wright SJ, Soudzilovskaia NA, Onipchenko VG. 2017. A global method for calculating plant CSR ecological strategies applied across biomes world-wide. Functional Ecology 31: 444-457.

Poorter L, Van Der Sande MT, Thompson J, Arets E, Alarcón A, Álvarez Sánchez J, Ascarrunz N, Balvanera P, Barajas Guzmán G, Boit A. 2015. Diversity enhances carbon storage in tropical forests. Global Ecology and Biogeography 24: 1314-1328.

Poorter L, Van DSMT, Arets EJMM, Ascarrunz N, Enquist B, Finegan B, Licona JC, Martínez-Ramos M, Mazzei L, Meave JA. 2017. Biodiversity and climate determine the functioning of Neotropical forests. Global Ecology and Biogeography 26: 1423-1434.

Porter JH. 2010. A Brief History of Data Sharing in the U.S. Long Term Ecological Research Network. Bulletin of the Ecological Society of America 91: 14-20.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Redman CL, Grove JM, Kuby LH. 2004. Integrating Social Science into the Long-Term Ecological Research (LTER) Network: Social Dimensions of Ecological Change and Ecological Dimensions of Social Change. Ecosystems 7: 161-171.

Servilla M, Brunt J, Costa D, Mcgann J, Waide R. 2016. The contribution and reuse of LTER data in the Provenance Aware Synthesis Tracking Architecture (PASTA) data repository. Ecological Informatics 36: 247-258.

Soranno PA, Schimel DS. 2014. Macrosystems ecology: big data, big ecology. Frontiers in Ecology & the Environment 12: 3.

Trajanov A, Spiegel H, Debeljak M, Sandén T. 2019. Using data mining techniques to model primary productivity from international long-term ecological research (ILTER) agricultural experiments in Austria. Regional Environmental Change 19: 325-337.

Vanderbilt K, Gaiser E. 2017. The International Long Term Ecological Research Network: a platform for collaboration. Ecosphere 8: e01697.

Wickham H. 2016. tidyverse: Easily install and load ‘tidyverse’packages. R package version 1.2.1.

Woodward G, Dumbrell AJ, Baird DJ, Hajibabaei M. 2014. Preface. ix-xiii*ix-xiii in Woodward G, Dumbrell AJ, Baird DJ, Hajibabaei M, eds. Advances in Ecological Research, vol. 51. Academic Press.

Ye Q, Song H, Li T. 2012. Cross-institutional collaboration networks in tourism and hospitality research. Tourism Management Perspectives 2: 55-64.

Figures

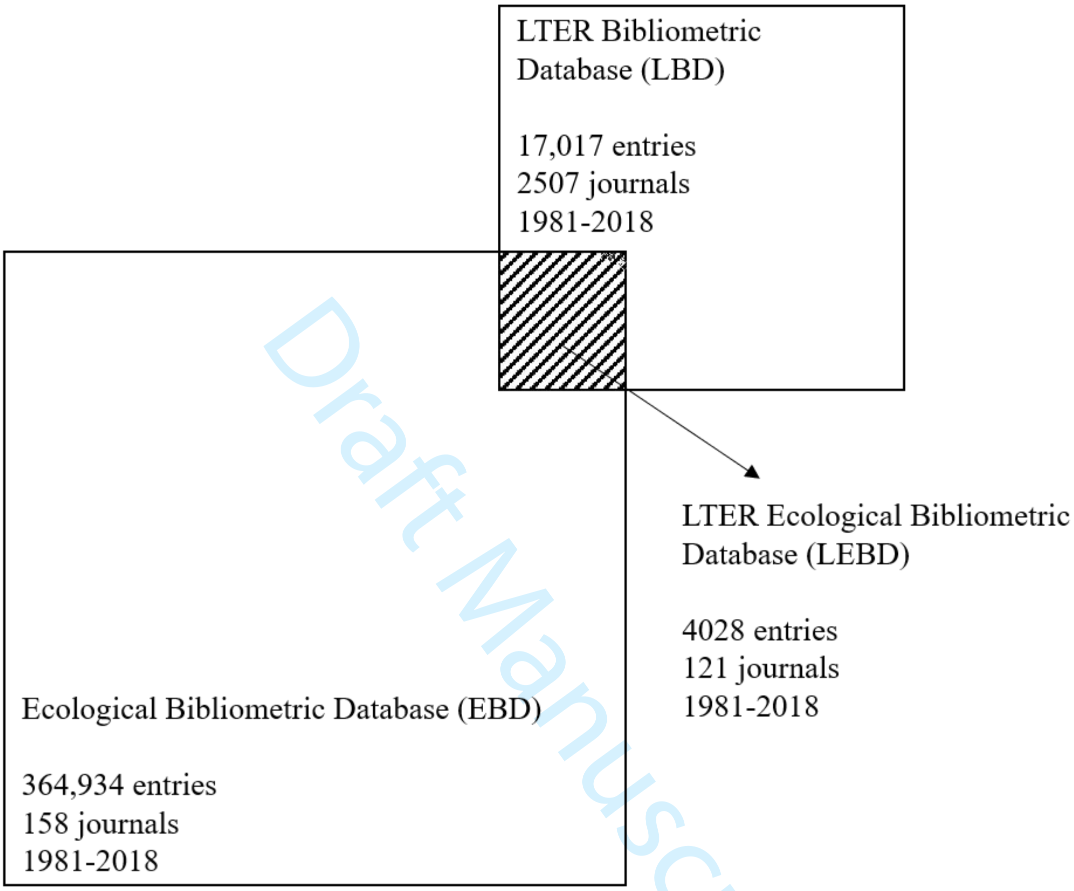

Figure 1. Relationship of bibliometric databases used in the study.

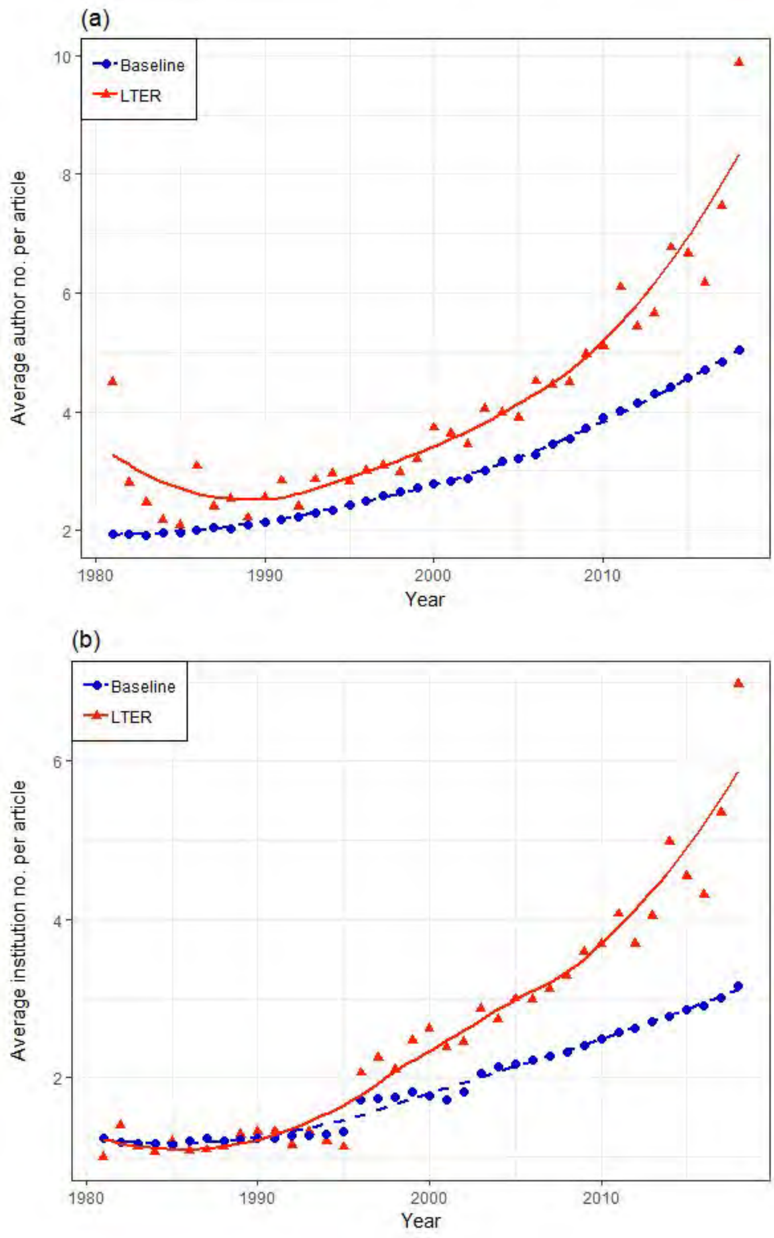

Figure 2. Comparison of change in collaboration of US LTER Network vs Baseline. (a) Comparison of average number of authors per paper. (b) Comparison of average number of participant institutions per paper. The Baseline group was constructed using the whole database of ecological literature.

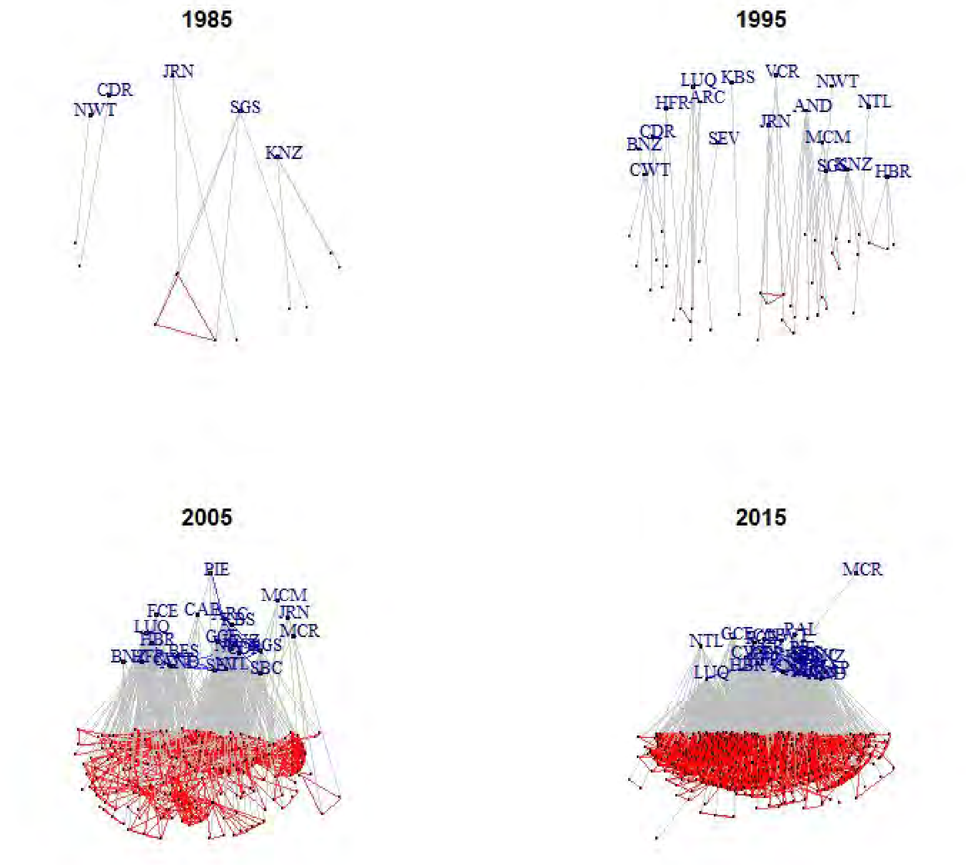

Figure 3. Evolution of institutional collaboration with site information in the US LTER Network. The networks of four years (1985, 1995, 2005 and 2015) are visualized and displayed. In the networks, the nodes at the top layer with labels represent the US LTER sites, while nodes at the bottom layer represent the institutes. Grey edge between institution and site indicates that institution has carried out research in a site (in grey), edge between two sites shows cross-site collaboration (in blue), while edge between two institutions shows institutional collaboration (in red).

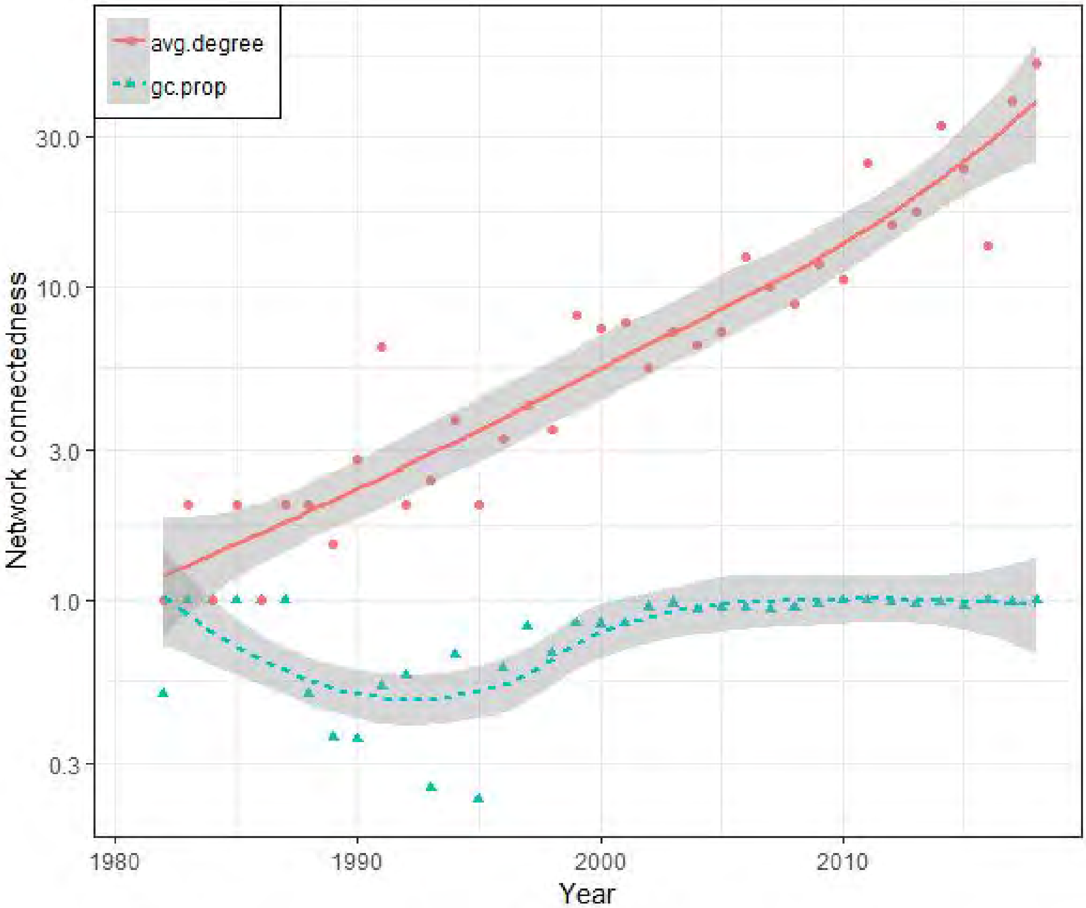

Figure 4. Temporal change of the LTER institutional collaboration network cohesion from 1981 to 2018 (avg.degree: average degree of nodes in giant component; gc.prop: proportion of nodes in giant component. Giant component is the largest cluster in the network). The grey areas represent 95% CI. The length of y-axis has been plotted on a log scale.

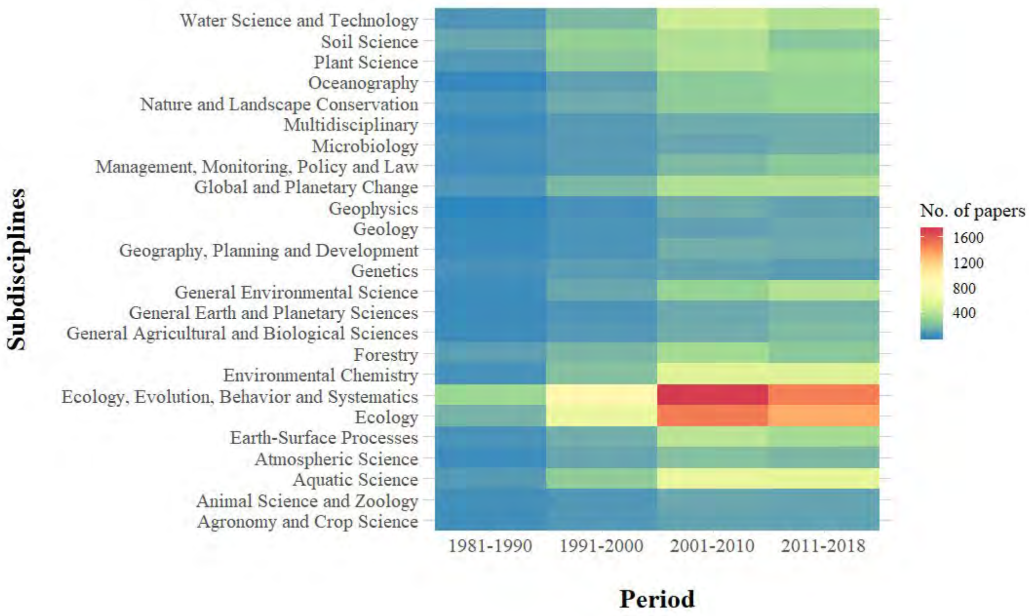

Figure 5. Temporal change in interdisciplinarity of the LTER research from 1981 to 2018. The data source came from LBD, and the top 25 ranked subdisciplines were selected.

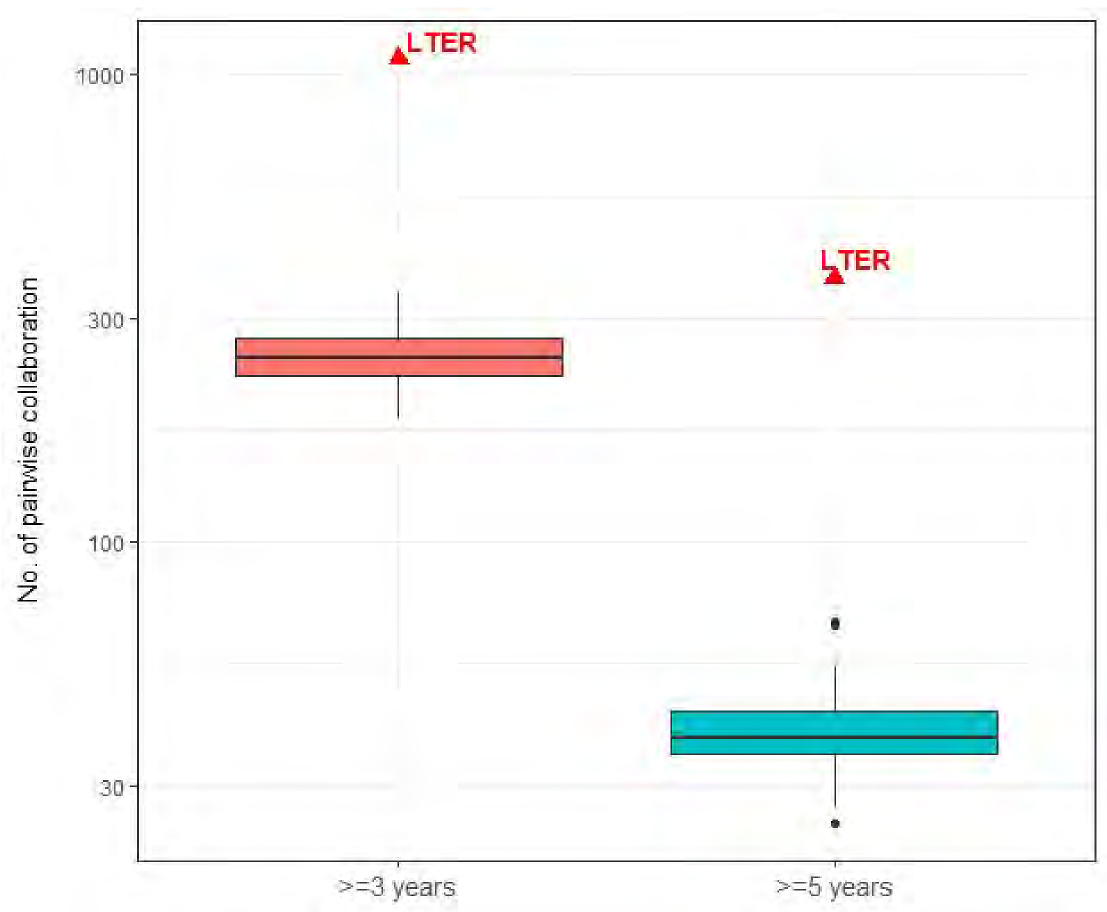

Figure 6. Exploration of whether the LTER program affects the duration of pairwise institutional collaborations. The figure compares the number of longer cooperation (≥ 3 years or ≥ 5 years) between the LTER group and baseline group. The boxplot depicts the number of collaborations lasting more than 3 years and more than 5 years for each of 99 Monte Carlo simulations drawn from the ecological literature. Simulation sample sizes were identical to the sample size of LTER publications. Length of actual LTER collaborations are displayed using a red triangle symbol with text label.

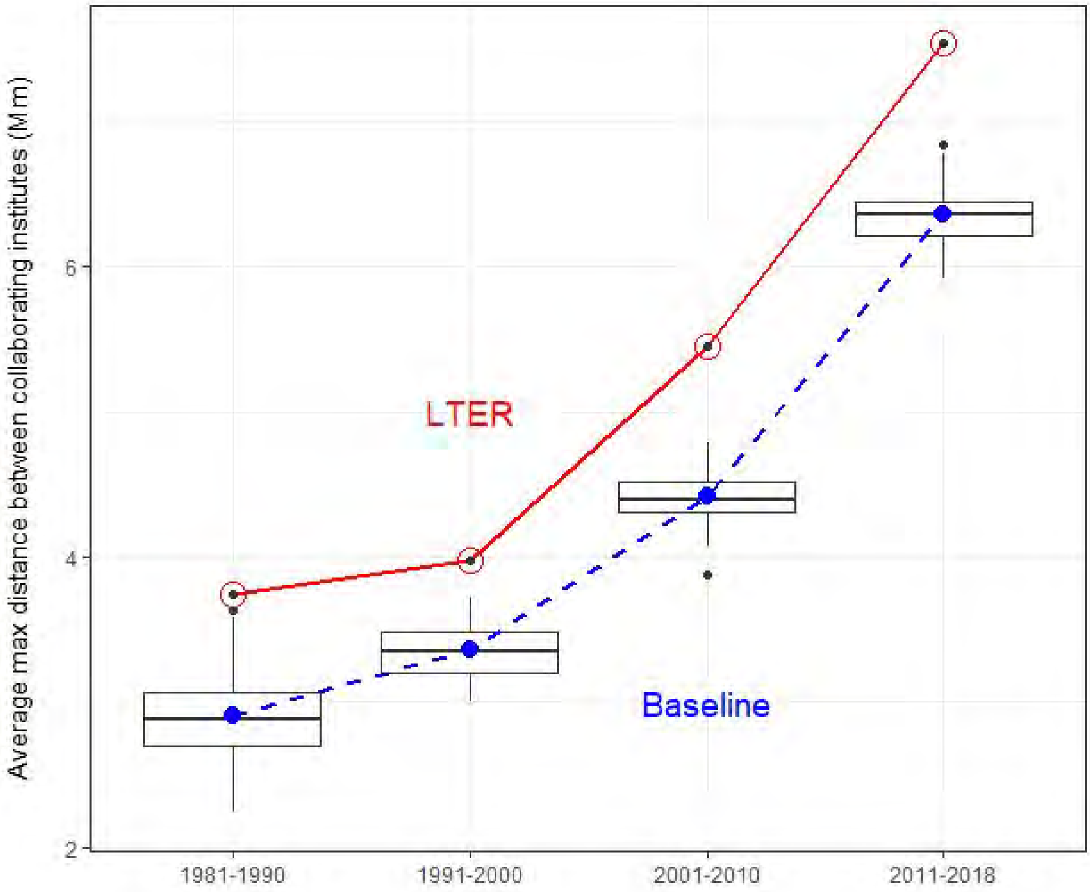

Figure 7. Exploration of whether the LTER program affects long-distance collaboration. The figure shows the change of average max distance between collaborating institutions in different periods. The boxplots show 100 samples of articles including 99 samples of simulation and 1 LTER sample. The overall trend (for all 100 samples) is depicted in dashed blue lines, while the trend of LTER is highlighted using solid red lines.