Analysis: Academic Betadiversity and the Future of Biological Interdisciplinary Research

## Abstract

Interdisciplinary work is a critical component of solving biological problems. Cross-field collaboration has led to theoretical and methodological breakthroughs and continues to drive scientific progress. Despite the fundamental nature of interdisciplinary work, we lack a basic understanding of the connections between biological fields. By tracking the contributions of hundreds of thousands of unique authors, I show that biological disciplines are part of a complex modular network connected by key bridges among fields. This analysis highlights the contribution of Environmental Sciences, Biotechnology, and Biodiversity and Conservation Biology in stitching together the biological sciences. Changes in connections among fields over time show the birth and growth of new disciplines within biology. Future growth among biological fields will require integration to forge new bridges and connect currently disparate topics.

## Introduction

Innovative discoveries come from collaboration across academic fields. The National Science Foundation and the National Institutes of Health highlight cross-disciplinary research as an integral ingredient for sparking scientific progress. From the mixing of quantitative genetics and evolutionary theory to create the modern synthesis () to biochemistry underlying Polymerase Chain Reaction(), the melding of biological fields has produced theoretical and methodological breakthroughs. Interdisciplinary work continues to drive innovation with the current “omics age” fueled by a tremendous expansion of bioinformatics. An ever-growing list of challenges, from the evolution of antibiotic resistance, to producing robust climate models, requires scientists to embrace a more expansive view of the life sciences.

Despite the critical importance of interdisciplinary work, it is difficult to assess the relative specialization and connections among biological fields. While studies have focused on citation indexes as a proxy of academic collaboration, individual citations may be a poor indicator of interdisciplinary work and the relevance of a particular citation to the aim of the research may be limited(). In contrast, tracking the publication careers of individual authors provides a new view into the depth and diversity of interdisciplinary research1. By studying the emergent connections of various related fields, we garner a new view of biology as an interconnected and complex system.

Networks provide a powerful way to visualize interdisciplinary research by using graph theory to represent connections among group members2. Graph theory is a branch of discrete mathematics that quantifies interactions among members of a set, called nodes, by measuring connections, called links, based on an interaction currency. To visualize the network of biological fields, we constructed an adjacency matrix wherein the dissimilarity among academic fields is measured by the abundance of articles published by individual authors in field-specific journals. The goal of this analysis was to determine 1) Which fields act as bridges to connect other fields within biology 2) Which fields have become more insular, and which more interdisciplinary, over time and 3) Where is there potential for increasing connectedness between fields?

To measure dissimilarity among life science fields, I extracted metadata for all articles published from 1995 to 2014 in 578 biological journals. Each of these journals was classified into one of thirty-nine biological fields (Table S1). Journals that did not belong to any field were discarded. After removing authors with less than four publications, I tallied 1,325,937 publications by 327,175 unique authors. After computing dissimilarity among fields, I calculated centrality and importance of these fields using the betweenness and degree network measures. Fields with high betweenness are central members of the network and connect other members. Fields with high degree directly interact with many other fields. Finally, fields with high eigen value centrality are connected to many well-connected members. Given that we expect centrality to increase with the total number of articles in a field, I propose a simple measure of insularity by dividing a field’s degree by the total number of publications. These four measures provide complimentary ways of evaluating centrality and connectedness within the biological sciences. In addition to these measures, I calculated the change in link strength between disciplines over time to quantify the shifting connections and affinities for biological interdisciplinary research.

## Results

The life sciences network is highly modular with distinct compartments each consisting of four to six biological fields. While there are some strong connections, such as between Ecology and Biodiversity and Conservation Biology, the majority of links are relatively weak. After removing extremely weak connections, the average author contributes to X fields. Environmental sciences and Biotechnology act as strong bridges between compartments and are the most well-connected fields within the life sciences. In general, fields with more articles were more well-connected. After correcting for this effect, Ornithology and Animal Behavior were examples of small fields with outsized contribution to connectivity, and Biochemistry and Microbiology were more insular fields, with many publications, but weak levels of interdisciplinary research.

Temporal patterns among connected fields were fairly static. The increases in connectivity between Environmental Sciences and Biotechnology highlight the importance of applied research to solve anthropogenic challenges (cite Nature paper). As fields gain popularity, they change the interactions among connected members. For example, the number of papers published in Molecular Biology journals has increased tremendously since 1995, and the current strong ties between Molecular Biology and Cell Biology may account for the decreased connection among Biochemistry and Cell Biology, which has decreased in the last two decades.

Visualizing the life sciences as a complex network gives rises to many questions. While the strong connections between Biochemisty and Molecular Biology stem from the joint goal of understanding the mechanistic underpinnings of cellular life, these fields are on the opposite side of biological network from Paleontology. The emergence of ancient DNA analysis, as well as increasing interest in comparative phylogenetic methods, may see a greater niche overlap among authors publishing in these journals in the future. The current challenge of producing robust climate models highlights the observed connection between Oceanography, Atmospheric Sciences and Environmental Sciences. The unexplored combinations of biological fields, such as Forestry and Evolutionary Biology, or Biophysics and Animal Behavior, may yield new and unexplored fertile grounds of scientific development.

## Conclusion

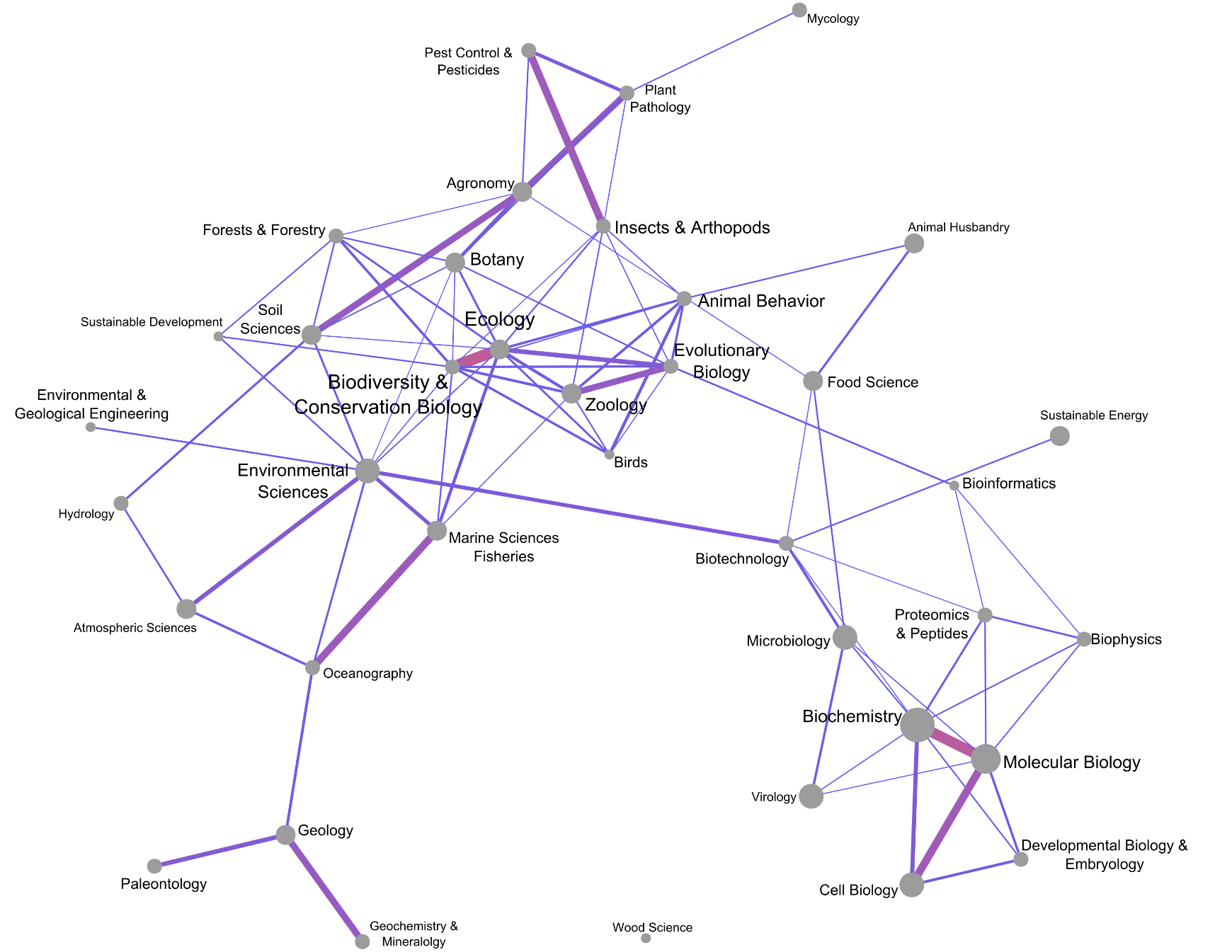
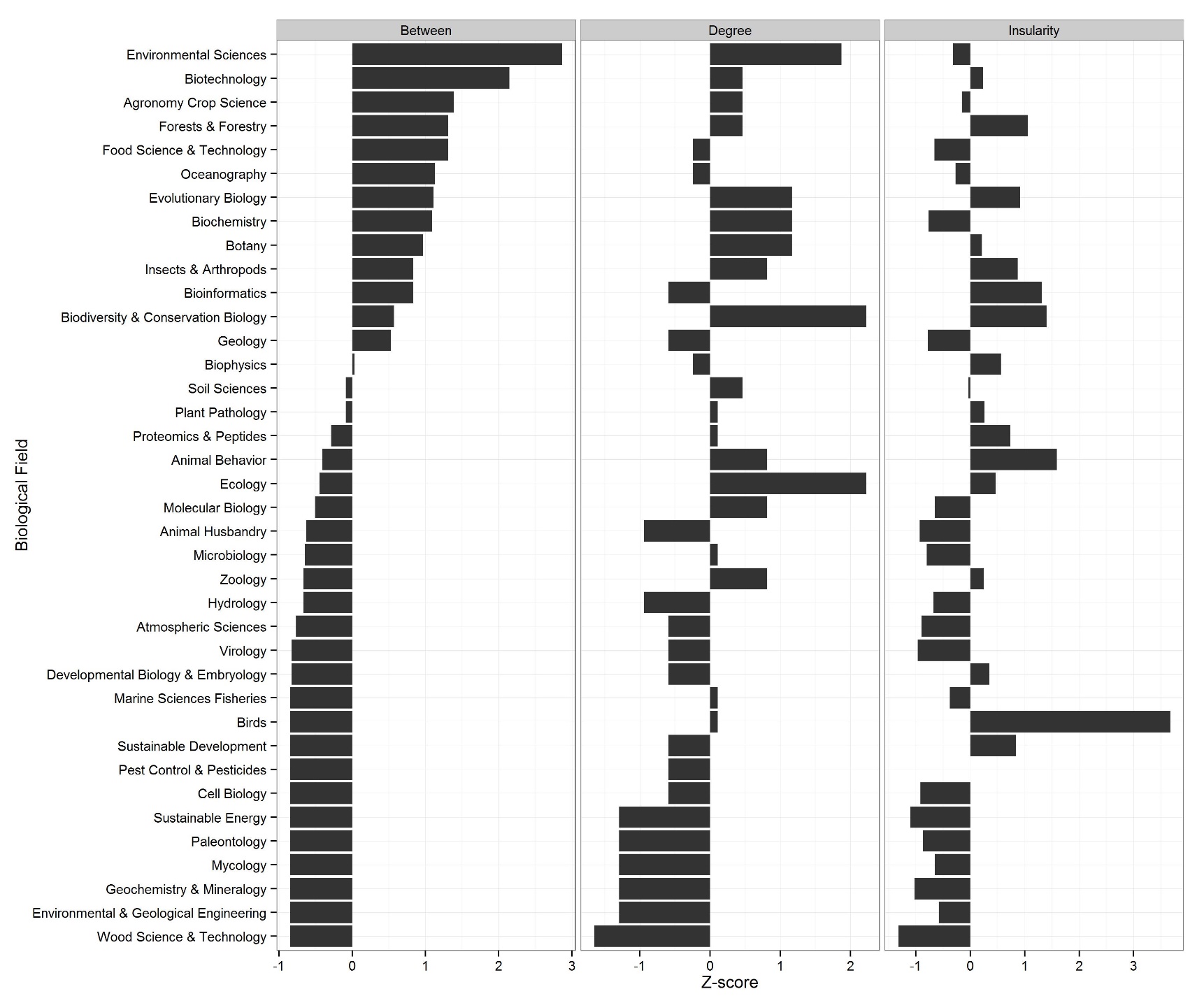
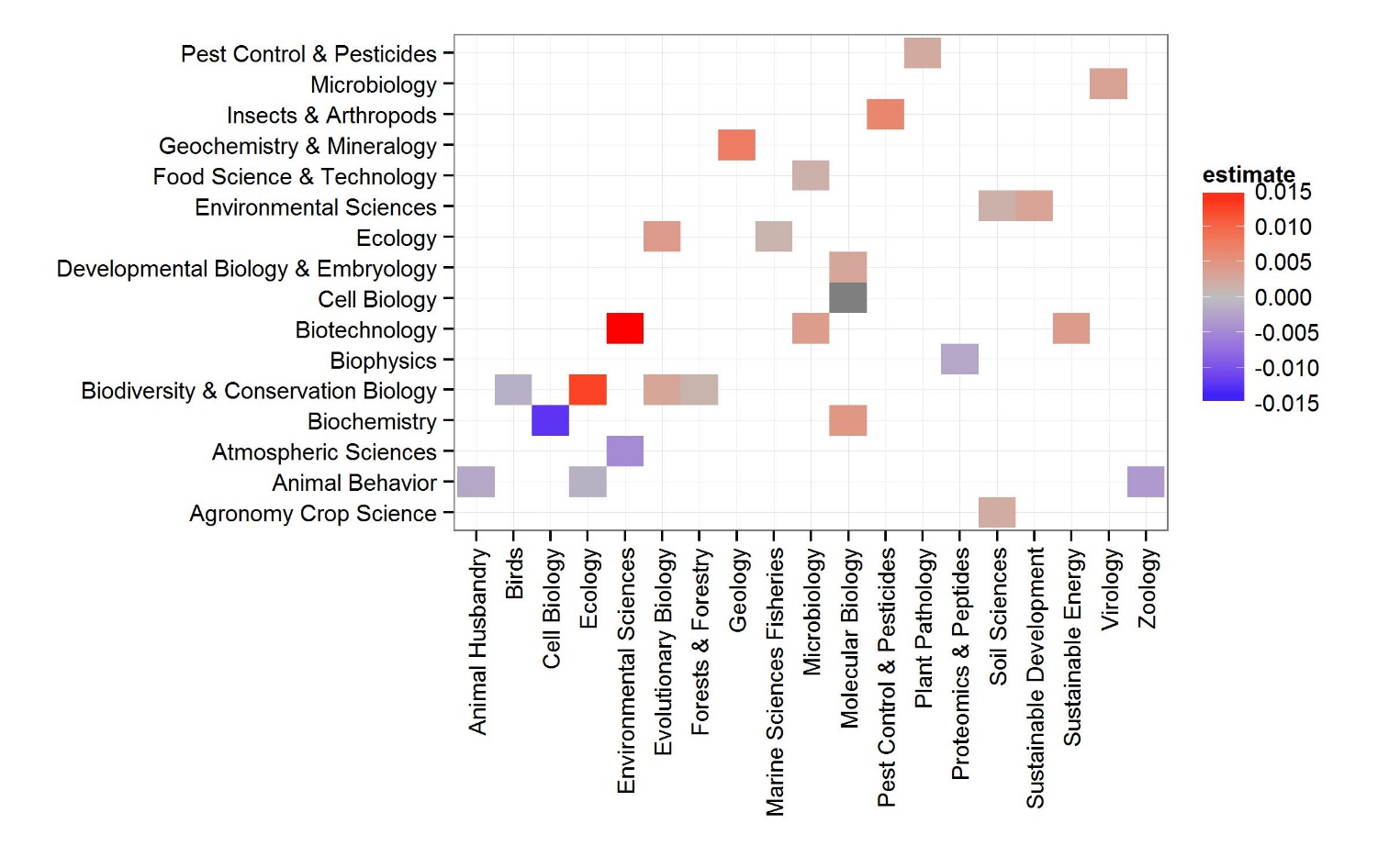


Figure 1 Network visualization of biological fields. Stronger connections among fields are shown by thicker and redder links. The size of text is proportional to the number of strong links to other academic fields. The size of node is proportional to the number of publications tallied within that field. Weak links with dissimilarity less than 0.05 have been removed.

Figure 2 Centrality measures for each biological field. Fields with high betweenness are central connect other fields. Fields with high degree directly interact with many other fields. Fields with high insularity have few connections given the number of publications in that field. For comparison, all fields were standardized and ordered by betweeness.

Figure 3 Estimated change in link strength as a function of year for academic fields. Only fields with significantly positive (red) or significantly negative slopes (blue) are shown. Years were binned into two-year groups to account for differing publishing rates in each field.

## Methods Summary

Journals were classified following the google scholar journal classification for Life and Earth Sciences (<https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=bio>). Each classification consisted of twenty journals. While the boundaries of discrete classifications will always be difficult to define, the vast majority of journals can be placed into one category. High ranking cross-field journals such as *Nature* were removed. Using the source title for the remaining journal, I queried the Scopus API for the metadata on all articles between 1995-2014 (Appendix A). Academic betadiveristy was defined by constructing an author by field matrix with each of the ~ 300,000 unique authors as rows, with the number of publications in each discipline in each column and calculated using Horn’s distance. Unique authors followed the Scopus ID, which cross-references names with affiliations and disciplines to avoid merging authors with the same name. The network statistics were calculated using the igraph package in R, and a correlation tests showed only moderate connection between indices (Appendix B). To calculate change in connectivity over time, I divided the dataset into two year chunks and recalculated both the dissimilarity among fields and the network statistics for each period. Linear regression was used to estimate the change in connectivity over time. Only relationships which had a significantly (p < 0.05) positive or negative slope were shown. All source code, network statistics, and plots of each field over time is available in the supplementary materials.

Works Cited

1. Parker, J. N., Allesina, S. & Lortie, C. J. Characterizing a scientific elite (B): Publication and citation patterns of the most highly cited scientists in environmental science and ecology. *Scientometrics* **94,** 469–480 (2013).

2. Newman, M. E. J. Coauthorship networks and patterns of scientific collaboration. *Proc. Natl. Acad. Sci. U. S. A.* **101 Suppl ,** 5200–5205 (2004).