Analysis: Academic Betadiversity and Future of Biological Collaboration

## Abstract

Interdisciplinary work is a critical component of solving biological problems. The mixing of fields 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 network. This network is highly modular and connected by several key fields which act as bridges between compartments of authors. Fields with high centrality include Environmental Sciences, Biotechnology, Biodiversity and Conservation Biology, highlighting these fields as key grounds for interdisciplinary work. Future growth among biological fields will require integration to forge new bridges to connect currently disparate topics.

## Introduction

Innovative discoveries come from the mixing of academic fields. The National Science Foundation and the National Institutes of Health both highlight cross-disciplinary as a central ingredient for sparking scientific breakthroughs. The melding of disparate fields of study has produced theoretical breakthroughs like the mixing of quantitative genetics and evolutionary theory to form the modern synthesis (), as well as methodological advances such as the biochemistry underlying Polymerase Chain Reactions. 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 wider view of life sciences.

Despite the critical importance of interdisciplinary work, it is difficult to assess the relative specialization and connections of biological fields. While studies have focused on citation indexes as a proxy of academic collaboration, individual citations may be a poor indicator of the level of interdisciplinary work and the relevance of a particular citation to the aim of the research may be limited(). In contrast, tracking individual authors as they published in research journals provides a new view into the depth and diversity of interdisciplinary research1. Here I use the contribution of individual authors to capture the interdisciplinary nature of each biological field. By studying the emergent connections of these fields, we garner a new view of biology as an interconnected and complex system.

Networks provides powerful way for visualizing 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 where each academic field is compared to every other field. This matrix consisted of the degree of dissimilarity between fields 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 life sciences? 2) Which fields have become more insular, and which more interdisciplinary over time? 3) Where is there potential growth among connections between fields?

To measure dissimilarity among life science fields, I extracted article metadata from all papers published from 1995 to 2014 for 578 biologicals journals. Each of these journals were 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 created a network to calculate the betweenness, degree, closeness and eigenvalue centrality of each field. These metrics all measure the centrality in graph networks. Fields with high betweenness are central members of the network and connect other members. Fields with high degree directly interact with many other fields. Fields with high closeness lie in the center of the overall network. Finally, fields with high eigen value centrality are connected to many well-connected members. 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 collaboration.

The life sciences network is highly modular, with distinct compartments 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. Environmental sciences and Biotechnology act as strong bridges between compartments and are the most well-connected fields within the life sciences. Authors that publish in environmental sciences also publish in Environmental Engineering, Atmospheric Sciences, Marines Sciences and Sustainability.

Temporal patterns among connected fields were fairly static. The largest increases in connectivity was between Environmental Sciences and Biotechnology, highlighting the importance of applied research to solve anthropogenic challenges (cite Nature paper). As fields gain popularity, they change the interactions among connected mebmers. 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.

Visualizing life-sciences as a complex network gives rises to many questions and spurs the development of interdisciplinary connections. While, the strong connections between Biochemisty and Molecular Biology stem from the joint goal of understand the mechanistic underpinnings of cellular life, these fields are on the opposite side of biological network from Paleontology. The emergence of ancient DNA approaches, 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.

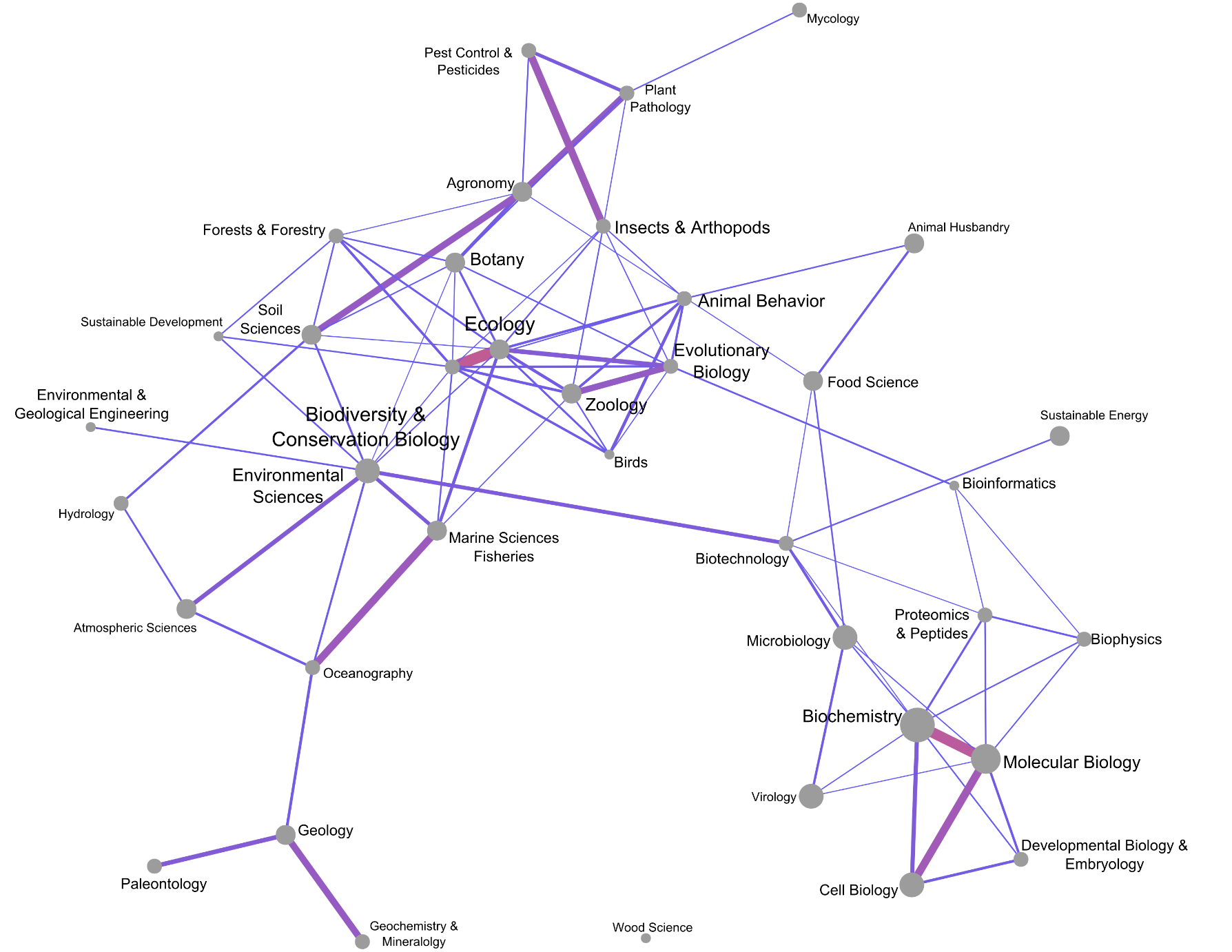
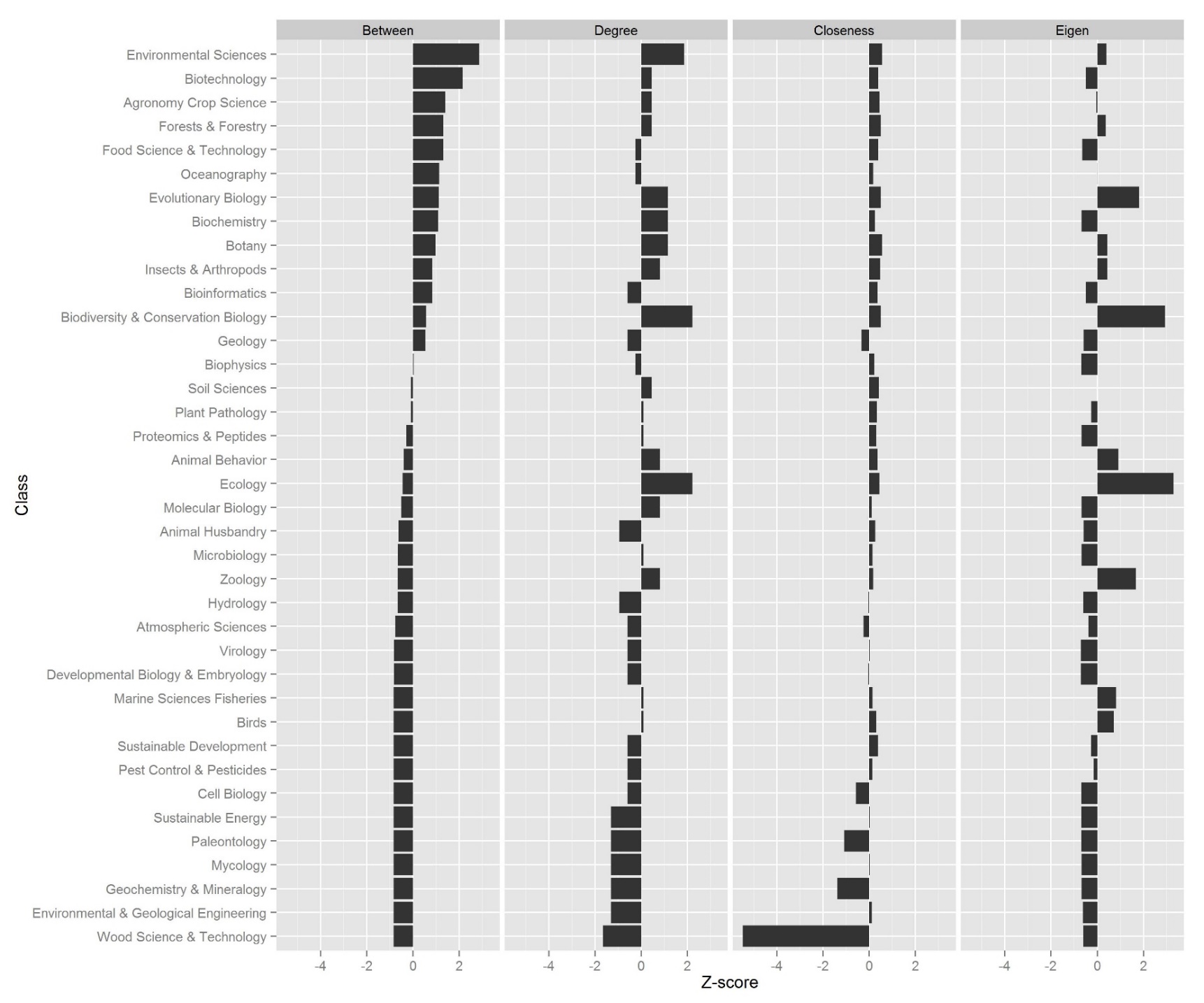
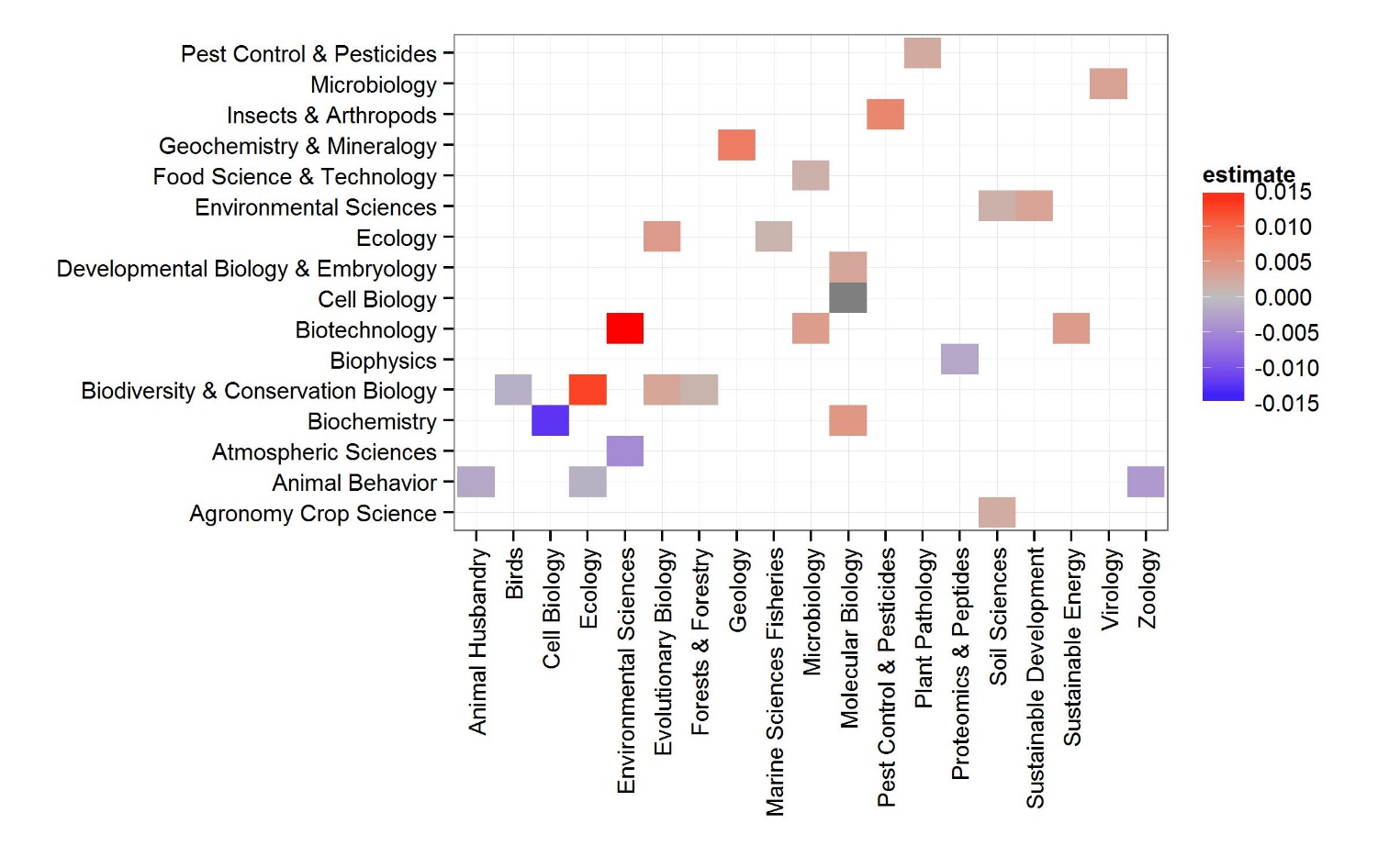


Figure 1 Network visualization of biological fields. Stronger connections among fields are shown by thicker and redder links. The size of text is proportional to the number of strong links to other academic fields. The size of node is proportional to the number of publications tallied within that field.

Figure 2 Centrality measures for each biological field. Fields with high betweenness are central members of the network and connect other members. Fields with high degree directly interact with many other fields. Fields with high closeness lie in the center of the overall network. Fields with high eigen value centrality are connected to many well connected members. For comparison, all fields were standardized and ordered by highest (most central) to lowest (least central).

Figure 3 Estimated change in link strength as a function of year for academic fields. Only fields with significantly positive or significantly negative estimates 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 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 between one or two categories. 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). Metrics each contributed information to the centrality of academic fields, and were only moderately correlated (Appendix B). Academic betadiveristy was defined by constructing an author by field matrix with each of the 300,000 unique authors as a row, with the number of publications in each discipline in each column. The network statistics were calculated using the igraph package in R, and a correlation tests showed only moderate connection between indices. 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).

**Appendix**