Research Policy 53 (2024) 105040

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Hedgehogs, foxes, and global science ecosystems: Decoding universities' research profiles across fields with nested ecological networks

Charles J. Gomez [a,*], Dahlia Lieberman [b], Elina I. Mäkinen [c]

[a] The University of Arizona, School of Sociology, United States of America
[b] City University of New York, Queens College, Department of Sociology, United States of America
[c] Tampere University, Faculty of Management and Business, Finland

ARTICLE INFO

Keywords:
Global science
Ecology
Ecosystem
Networks
Citations

ABSTRACT

Modern scientific research evokes ecological imagery and metaphors, given that it is global, interdependent, and diverse. Ecological network structures—like matrices of species inhabiting islands across an archipelago—can be reordered to form nested patterns. These patterns describe the overall health of ecosystems, place species on a spectrum between being described as generalists (foxes) or specialists (hedgehogs), and which of these interactions might appear or disappear. Using the number of citations universities receive for work published in a particular subfield taken from over 66 million scientific publications in OpenAlex, we construct and analyze yearly nested ecological networks of a dozen academic fields between 1990 and 2017. We find increasingly nested structures across fields infer future acknowledgment in different subfields. We argue that this framework can inform policy on scientific research and university funding and evaluation.

1. Introduction

Modern scientific fields evoke an image of a global, interdependent, and diverse “knowledge ecosystems” that reflects the complexities of scientific knowledge production (Jones et al., 2008; Leydesdorff et al., 2012; Nielsen and Andersen, 2021; Wagner et al., 2019; Wuestman et al., 2019). Just as natural ecosystems are described as interactive and diverse networks of flora, fauna, and biota, researchers in universities around the world produce and disseminate new knowledge through increasingly international collaborations that contribute to global discourses on the latest research trends (Graf and Kalthaus, 2018; Kwiek, 2020; Wagner et al., 2017). At the same time, universities tend to develop varying capabilities to conduct research. They often specialize and build a reputation in different research areas in a field and are thusly recognized for it by other researchers. Many governments achieve this by internationalizing just a handful of their top university research programs to make them “world-class” or by imposing research quality rubrics on their faculty (Mok and Chan, 2008; Pardo-Guerra, 2022).

The adage of *the fox and the hedgehog* (Berlin, 1953)—where “the fox knows many things, but the hedgehog knows one big thing”—is apt. Universities exist on a spectrum between the “fox” and the “hedgehog.” On the “fox-like” end, globally prominent universities are recognized in many different areas of a field (e.g., Stanford University or the University of Cambridge), while on the “hedgehog-like” end, universities are recognized in only few (e.g., University of Anbar in Iraq and Universidad de Los Andes in Colombia, which are both ranked in the 1001–1200 category in the 2024 QS World University Rankings[1]). However, most universities fall somewhere in between on a spectrum of generalists (fox-like) to specialists (hedgehog-like).

If we extend beyond the scientific fields-as-ecosystems metaphor and model the global production of scientific knowledge with the same dynamics as ecosystems, we could develop a new comprehensive understanding of how universities are recognized and acknowledged across fields. We argue that the specific research areas in the field that universities become recognized in by their peers may follow broader patterns explained by ecological models. Of particular importance are the emergent nested patterns of interactions (like species of birds inhabiting islands spread across an archipelago) that form networks. This is because ecosystems are dynamic systems of interactions between living creatures and their environment that achieve and sustain equilibrium. And ecosystems exhibit nested network structures (birds inhabiting chains of islands) when they achieve and sustain equilibrium. Ecosystems in equilibrium not only exhibit these nested structures, but are robust and thriving. Canonical measures applied to study these networks

---

* Corresponding author at: The University of Arizona, Social Sciences, Room 426, 1145 E. South Campus Drive, Tucson, AZ 85721, United States of America.
E-mail address: cjgomez@arizona.edu (C.J. Gomez).

[1] Source: https://www.topuniversities.com/university-rankings/world-university-rankings/2024

https://doi.org/10.1016/j.respol.2024.105040
Received 13 April 2022; Received in revised form 16 May 2024; Accepted 30 May 2024

like the “Nestedness based on paired Overlap and Decreasing Fill” (or NODF) measure (Almeida-Neto et al., 2008) and Shannon's information entropy (Shannon, 1948) could also describe scientific knowledge production of universities around the world and across the research areas in a field as a vibrant, thriving, and robust ecosystem.

We, therefore, ask what can nested ecological networks and their measures tell us about features of global science that may reveal their robustness, vibrancy, and future trends. There are many successful non-ecological applications in inter-organizational and international trade research (Bustos et al., 2012; Hidalgo and Hausmann, 2009; Ren et al., 2020; Saavedra et al., 2009; Saracco et al., 2015; Scholl et al., 2021; Tacchella et al., 2012; Uzzi, 1997). The network measures of nestedness applied to these examples are indicators of not only the overall vitality and robustness of these markets, but they also predict the emergence of trade relationships and comparative trade advantages (cf. Ulrich et al., 2009). By extension, these models have the potential to uncover the future acknowledgment and recognition of universities across subfields and the growth of global science overall. Successful ecological models of global science are vitally important to expand our understanding of how scientific fields systemically operate at an international level and could inform national research policies on collaborations, mobility, and resource allocation. Yet, given the potential for such a mapping, little work has sought to formally model scientific knowledge in different fields globally as natural ecosystems using the same frameworks and tools.

Using scientific publication metadata from OpenAlex (Priem et al., 2022), we analyze over 66 million papers from a dozen academic fields, 4794 universities in 176 countries, to test whether scientific knowledge production mirrors ecological dynamics. We count the number of citations researchers at different universities receive across subfields each year. By analyzing over 25 years of data from 1990 to 2017, we describe the global nested patterns of international research across fields and identify the span of universities' research presence across subfields. We also measure the extent to which these global nested patterns forecast the subfields which universities become acknowledged in over time.

We describe and forecast ecological patterns in the citations that universities receive within different academic fields over time. Not only is there a clear, emerging nested structure to global scientific production across fields, but citation acknowledgments are more distributed and proportional around the world as well. While the most acknowledged “fox-like” universities are among the most prominent in the world, even more “hedgehog-like” universities are acknowledged across more subfields over time. We then forecast future research acknowledgment across subfields based on university-to-subfield network structures and find that future recognition in a subfield is on average more likely to occur than not. Finally, we argue that this framework will help different universities identify successful research profiles based on their own demonstrated potential and strategize how to best advance their reputation to compete for attention and recognition globally.

2. Motivation

2.1. Ecological metaphors and global scientific fields as ecosystems

The attraction and theoretical success of ecological metaphors are embodied by the qualities that describe the structure and dynamics of ecosystems: scalability, interdependence, and diversity. Ecologies are scalable in that they are hierarchical and modular, comprising of local interactions that can have unintended system-level consequences, and vice-versa. Ecosystems are interdependent in that the various biota and environs interact towards the flourishing, survival, or death of one another. Ecosystems are diverse in that the various biota that interact with one another do so because of their complimentary or supplemental characteristics (e.g., species cohabitating across different locales, etc.). So, like studying different species of birds inhabiting islands of an archipelago, these qualities of ecosystems are applicable across diverse biospheres and species. However, they are also appealing to the study of other non-biologically centered settings.

Ecological metaphors are commonly employed in the study of social life. Generally, the metaphor describes various social units that—-whether people or organizations—locally interact with one another and their environs (broadly defined) which results in systemic (and sometimes unintended) outcomes. In U.S. sociology, the Chicago School (also referred to as the Ecology School) situated social interactions in terms of the urban setting that structured and influenced these interactions (for its history see, Cavan, 1983). Organizational ecology and its various theoretical forms employed population metrics to model the birth, life, and death of organizations in industries, where niche determined survivability as organizations balanced specialization with generalist pressures (Baum and Singh, 1994; Carroll and Swaminathan, 1992). The rise in self-organizing complex adaptive systems theorizing in the late twentieth century evoked ecological imagery and metaphors to model diverse problems like predicting the stock market, economic growth, and modeling urban growth (Grimm et al., 2005). Theories of network ecologies explore how tie formation and the emergent network structure are influenced by organizational environments, like friendship networks in classrooms (McFarland et al., 2014).

Ecological metaphors are rife in studies of innovation and universities. For instance, measures of entrepreneurial ecosystems like a composite indicator (CI) focus on how economic complexity and geographic diversity shape its structure and impact (Lafuente et al., 2021). Such ecological metaphors underscore the integration, differentiation, and proximity of research universities, venture capital, and highly-skilled labor that all drive technology start-ups and innovations in technological research clusters like Silicon Valley (Andrews et al., 2022; Leendertse et al., 2021; Leydesdorff, 2000). In fact, many case analyses of less successful technology clusters (e.g., Route 128 near Boston, etc.) employ the ecological metaphor to point to the siloed and less integrated nature of the various technology companies as compared to their more successful counterparts (e.g., California Bay Area) (Kapturkiewicz, 2021; Saxenian, 1996, 2007). The difficulty in replicating the success of Silicon Valley elsewhere is in large part due to how unique the ecological dynamics that gave rise to it in the first place are. Other studies on U.S. universities have creatively applied ecologically-inspired theories of organizations from sociology and management studies to understand the market forces behind their affordability (Eaton et al., 2019).

Surprisingly, little attention is paid to expanding the ecological metaphor to the global scientific enterprise that similar studies of universities and innovation have been so successful with. Global science as the ensemble of researchers in universities around the world producing and being recognized for knowledge exhibits many of the same characteristics as an ecosystem: global and scalable (Nielsen and Andersen, 2021; Wagner and Leydesdorff, 2005; Wagner et al., 2017), interdependent (Jones et al., 2008), and diverse (Collins, 2004; Hofstra et al., 2020). It is scalable in that nations are the most important boundary in science, highly influencing universities through public funds and researchers' careers. It is interdependent as science is a complex enterprise that requires collaboration and coordination across universities and countries. It is diverse in that academic fields are sprawling intellectual landscapes of subfields, schools of thought, and burgeoning areas of new inquiry. Researchers in universities collaborate and coordinate with one another across these local intellectual sites, publishing work and being recognized for it by their peers. This is typically embodied by citations received from published work: the currency of scientific knowledge transfer, diffusion, and recognition between researchers and universities. While citations as a reflection of this are wrought with issues (Glänzel and Thijs, 2004; Petersen et al., 2019; Tahamtan and Bornmann, 2019), they remain a fairly reliable indicator of recognition and acknowledgment in a field. Citations embody both recognition for work being done, as well as the diffusion of knowledge across fields, countries, organizations, and even individual researchers. Thus, the structure of citations received by researchers at universities around the world could be used to model ecological structures and dynamics beyond conceptual metaphors.

2.2. Nestedness as an informative ecological structure

Some studies go beyond the metaphorical applications of ecosystems and have employed the same empirical measures applied to real ecosystems to non-biological ecosystems instead, particularly economic systems and markets. This includes the supplier relationships in the New York garment industry (Saavedra et al., 2011, 2009) and predicting the presence and absence of industries across countries and national competitiveness in different industries (Bustos et al., 2012; Hidalgo and Hausmann, 2009; Scholl et al., 2021; Tacchella et al., 2012). This involves constructing bipartite networks—the structured relationship between one type of object to a different kind of object—and applying ecological measures to them. For instance, ecologists construct bipartite networks of different interactions like birds living across islands in an archipelago, as seen in Fig. 1. Instead, economists use bipartite networks to describe international trade as an ecosystem consisting of countries (one type of object) and the comparative advantages (the structured relationship) they have across different industries (the second type of object). These bipartite structures take the form of a binary absence-presence matrix of whether a relationship exists. If a relationship exists in the matrix, a value of 1 is assigned, otherwise 0 is assigned. This is conceptually shown in Fig. 1. In ecology, a presence occurs if a particular type of bird occupies an island in an archipelago. In economics, a presence occurs if a country has a revealed comparative advantage with some product in international trade.

Nestedness is a canonical ecological feature of these “ecosystem”-inspired absence-presence interaction matrices. Nestedness implies that these systems are composed of a core set of interactions to which the rest of the network is attached. Nestedness also implies that specialist species interact primarily with their more generalist counterparts. Since generalists are largely stable when it comes to what they specialize in and their relative comparative advantages, nestedness can help enhance the

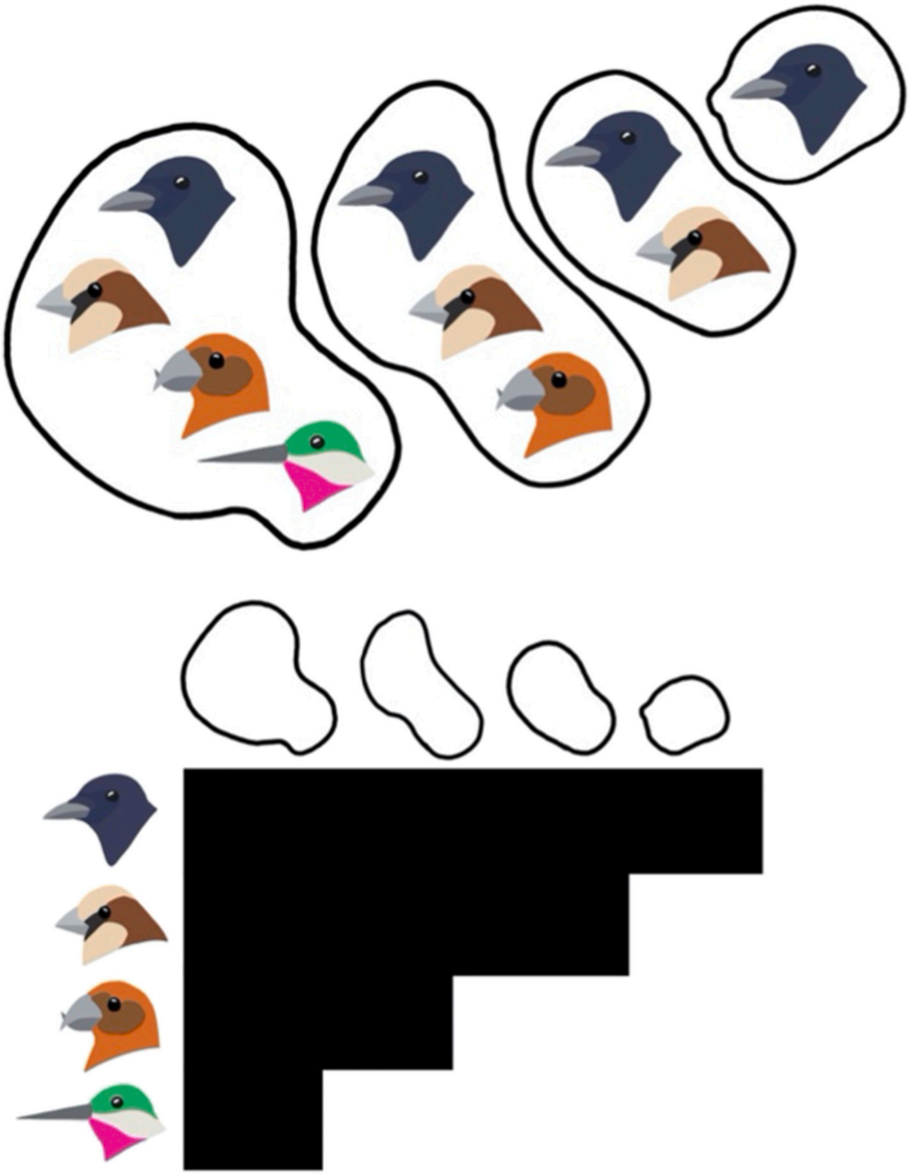


Fig. 1. Ecological example of a bipartite network structure and nestedness (Original Image Replicated From González et al., 2020). Bipartite interactions between different species of birds across islands in an archipelago in the top row. The bipartite structure at the bottom represents the presence of a bird inhabiting a particular island. Rearranging the structure of these presences reveals a nested structure.

survival of rare species. Nestedness can enhance systemic stability, and therefore, it is considered an important structural property of these bipartite networks.

In ecology, the high degrees of measured nestedness in interactive species matrices are telltale features of ecosystems in equilibria that are healthy and thriving. Ecologists have used nestedness to predict the biota (i.e., animals) in specific ecological sites, like birds inhabiting different islands. Nestedness means that rare species inhabit predominantly diverse patches, while ubiquitous species tend to inhabit both diverse and non-diverse locations. In the case of scientific fields as ecosystems, universities are like birds in that they inhabit different islands in a field (i.e., subfields), where “inhabit” means that they have received a majority of citations in their field, recognizing them for their work in the subfield (i.e., island).

There are fundamental ways, however, with which ecological systems of birds across an archipelago differ from scientific fields and where the metaphor is limited. For instance, universities are not like birds distributed across an archipelago because they do not have inherent circumstantial advantages (i.e., particular types of beaks that allow them to be more successful in eating prey in different islands); they are instead better resourced to hire more and prominent faculty. Universities have the agency to strategically decide to develop (or not develop) capacities in subfields. In contrast, a species of bird with a beak that is specialized for eating a particular type of food native to one island cannot choose to relocate to a different one and inhabit a different ecological niche.

In other ways, however, the metaphor highlights similarities. Again consider the case of the bird with the specialized beak limited to one island. Many universities are set up with specialized research and instructional foci, such as the natural science and engineering (e.g., MIT, Caltech, the Indian Institutes of Technology, Imperial College London, etc.) or business and economics (e.g., London School of Economics, Stockholm School of Economics, etc.). While many of these specialized universities excel in areas beyond their setup (e.g., MIT being a leader in economics and political science), others are limited and cannot develop a fox-like capacity in new areas, despite potential interest in doing so. For instance, the London School of Economics is a social science specialist university effectively restricted from developing capacities in the natural sciences without an enormous investment in resources. Furthermore, much like the bird with a beak bespoke to a particular island, other universities are also limited to “inhabiting” only certain subfields because of their setup. For instance, the University of Hawai'i is a world leader in volcanology because of its proximity to active volcanos but is not as recognized in other areas of geology because of its distinct advantage in the study of volcanos.

Nevertheless, the success of other applications of ecological metaphors and measures to areas of social life are indicative of its success in the case of global scientific fields-as-ecosystems. For instance, the limitations of the metaphor applied to universities are just as applicable to countries and firms engaged in trade, yet the conceptual and even predictive insights gleaned from these cases still hold. Despite the many differences, what links both ecologies with these socially-inspired examples, and what allows the metaphor to be as successfully applied as it is, are their shared features as systems: their scalability, interdependence, and diversity.

To this end, universities can be profiled “ecologically” by the breadth and depth of their research activities across fields and over time. Here, as we described in the introduction, the fable of the *hedgehog and the fox* is helpful: Fox-like universities could be known for many things, but hedgehog-like universities could be known for fewer things. In organizational ecology, firms that specialize thrive while those that are generalists risk extinction. By contrast, nations that have comparative advantages in many industries tend to have much stronger economies than those that specialize in just a few. However, there is evidence that universities recognize the value of being fox-like (a generalist), as we see university mergers motivated by becoming broader and bigger and doing better in university rankings (Docampo et al., 2015; Eastman and Lang, 2001). Describing the research profiles of universities and countries as a spectrum between fox-like and hedgehog-like is more of a reflection of the capacity to do research across multiple areas and be recognized for it.

For instance, large academic departments tend to have the resources and faculty to explore a broad array of specialized areas within their field. This breadth allows them to push the boundaries of research and knowledge in various sub-disciplines. This diversity is possible due to a department's size and access to extensive resources. However, smaller academic departments often need to focus on core areas that define their field to maintain depth and quality in their programs. These core areas are typically fundamental aspects of the discipline that ensure their academic identity is maintained and that graduate students receive a comprehensive education, even if the department cannot cover as many specialized topics. This approach allows smaller departments to provide a solid foundation for training researchers, ensuring they have the necessary skills and knowledge to succeed in their field.

The concept of nestedness was initially conceived by ecologists to quantify patterns using the concept of *entropy*, commonly known as a measure of order or disorder in a system. More precisely, entropy is a measure of the different ways in which a system could possibly be rearranged and re-ordered. If the structure of an ecosystem's interactions is highly internally ordered, then this is a strong indication that the ecosystem will not collapse and allow for better forecasting where future presences may occur. Higher entropy systems (think higher temperatures) are more “disordered,” reflecting more ways that the system could be internally arranged and reordered. Lower entropy systems (think lower temperatures) are more “ordered,” reflecting fewer ways. In the process of developing nestedness, ecologists have re-engaged with entropy over time (Sherwin et al., 2019).

Ecologists use a version of entropy from information science called Shannon entropy or information entropy as one among many measures of nestedness used to characterize the fox-like or hedgehog-like qualities of species. Information entropy (or simply entropy) is a measure extracted from features of absence-presence matrices that characterize species along a generalist-specialist (or fox-to-hedgehog) spectrum. The benefit is that entropy is a bounded measure, where 0 refers to no disorder in the system and 1 refers to complete disorder. This renders the measure comparable across different settings. As such, entropy can be used to characterize universities along the fox-to-hedgehog spectrum. For instance, birds that inhabit multiple islands in an archipelago would be associated with higher entropy while birds that inhabit only a handful of islands would be associated with lower entropy. High entropy universities are acknowledged and recognized across many subfields, while low entropy universities in far fewer. We offer our first proposition:

Proposition 1. The most well-known and prominent universities (fox-like) globally are associated with higher entropy scores (acknowledged across more subfields) than less-known and less prominent universities (hedgehog-like).

Nestedness is used in ecology to forecast the future appearance of particular interactions over time in ecosystems. When comparing a nested ecological matrix structure from observed data to what a hypothetical perfectly nested matrix structure might look like, ecologists can identify unexpected absences or presences and argue that these sites are the most likely to appear or disappear, respectively. The hedgehog-like or fox-like qualities of universities might also indicate which research areas they become recognized in over time.

The connection between the nestedness structure and identifying future presences and absences is well documented in applications of ecological measures to social data. Bustos et al. (2012) show how the absences of industries, both internationally and nationally, are predictable using the nested features of trade data, while Bustos and Yıldırım (2022) use several indices, including nestedness, to predict the likelihood of the presence of absences of industries. Ren et al. (2020)

argue that nestedness could also be used to predict the future economic growth of countries, specifically focusing on the United States and China. By analyzing the nested structure of universities-to-subfields as bipartite networks (or networks between two different types of nodes, such as universities and subfields), we can similarly infer which subfields become a recognized research area for different universities.

How does this work? An ecosystem is characterized by the nested structure of an absence-presence matrix. The more nested the structure becomes, the more internally ordered the absences and presences become too. We then can identify places in the matrix where we would expect a presence but instead find none, or an unexpected absence. If we identify an unexpected absence where we should expect a presence in a nested matrix, then the subsequent increasingly nested structure over time would suggest that the current unexpected absence could turn into a presence at some point in the future. So, by extension, an increasingly nested structure between universities and subfields could imply the future recognition of some universities by their peers in a given subfield. This may offer a useful indicator for areas of future growth across fields. To this end, we offer our second proposition and now turn to outline how we empirically model and test this.

Proposition 2. Nestedness can be used to forecast the future subfields that universities will become recognized in by their peers.

3. Data

We analyze over 66 million papers across 25 years from a dozen academic fields and 4794 universities in 176 countries. This work uses data primarily from one of the largest and most expansive repositories on academic publications, OpenAlex (Priem et al., 2022). From published papers, OpenAlex curates abstracts, titles, authors and their institutional affiliation, and citations made by the papers. OpenAlex classifies journals into various fields, which provides a reliable reflection of disciplinary boundaries and allows for selection across a wide variety of fields. As such, fields are identified and defined for our purposes as their field IDs in OpenAlex. These IDs are grouped into a six-tier hierarchy, where level zero (the highest level) is the broadest classification, comprising nearly 20 different classifications, including biology, computer science, and physics. Level one (the second highest level) parses these fields into slightly more granular divisions, such as astrophysics or solid-state physics instead of physics. These fields are: (1) biology, (2) business, (3) chemistry, (4) computer science, (5) economics, (6) environmental science, (7) geology, (8) materials science, (9) mathematics, (10) medicine, (11) physics, and (12) psychology. We use the level one division to categorize the division of knowledge in each of these fields. In the Supplemental Materials Table 1, we itemize and discuss the subfields for each of these twelve fields.

4. Analytic approach

4.1. Creating the university-to-subfield citation matrix to measure nestedness

For each year and field, $Field_t$, we create a matrix of universities and subfields as described earlier. We take all the papers in $Field_t$ and look at all of the university affiliations on all of these papers. These affiliations are determined by GRID, the global research identifying database, and are how OpenAlex curates author affiliations. Most universities in our data are universities, but they also include government, non-profit and corporate scientific research institutes. Along with organization metadata from GRID curated on each paper, each paper is also affixed with a subfield classification based on the journal it was published in.

We then take all papers published in $Field_t$ and using papers' affiliated universities and subfield designations, we aggregate the total number of citations received by each university in each subfield within a five-year time window. This effectively creates a university-to-subfield matrix for $Field_t$, defined as ψ_{Field_t}. Here, the rows are the universities, the columns are the subfields of the field, and the entries are the total number of citations received by a given university in a given subfield in year t using the five-year window threshold. Given that different fields garner citations at different rates, we set a standard five-year window across fields for all the citations that a paper published in year t accumulates in the five-year period. We test several other time thresholds and find our results qualitatively hold.[2]

In addition, for each university, we remove all citing papers that have any authors from the same university. That way, recognition for work in each subfield is not mispresented by self-recognition from researchers at the same university. For example, for all the citations received by researchers from the University of California, Berkeley in the field of astrophysics for papers published in the year 2000, we exclude the citations received from papers authored by researchers at Berkeley in the subsequent five-year time window.

Our focus is global research across countries, however the number of recognized universities housed in countries is highly skewed. Countries like the United States, Canada, and countries in Western Europe house a disproportionate number of the most active and recognized universities. As such, a concern is that a fully comprehensive and uncensored matrix would only reflect the vibrancy of universities in certain countries, rather than the entire “ecosystem” of a scientific field's global research worldwide. For the results we present here, we censor our data to the universities in the top 20th percentile in each country and year that are the most recognized vis-à-vis citation using our measure. The way we define “top” university is based on the row sums in Ψ_{Field_t}, the discretized form of the university-to-subfield citation matrix ψ_{Field_t}, that we describe later. Countries are only included if they have at least 1 university present each year in the span of our data from 1990 to 2015. We also tested an all-inclusive matrix where all universities were included irrespective of country representation which yielded similar results.

While there will be variation within subfields, the grouping of the citation data per subfield in the field matrices already compares subfields with mostly similar practices. However, since citation practices vary widely across fields, we then take the natural log of the citation counts in the matrix (i.e., the total number of citations received in a five-year window) to account for extreme values and to render citation counts comparable across fields, a common practice not only in bibliometric research but also in ecological network measures that accounts for a handful of extreme values (cf. Liu and Gao, 2022; Lundberg, 2007; Puuska et al., 2014; Yuan and Zhao, 2022). In addition, citations are a reliable indicator of scientific recognition and acknowledgment at the level of analysis we use in this paper but are wrought with issues. One issue is comparing citations over time, given the exponential growth in scientific papers that may lead to an inflation in citation counts. We also rerun our results “deflating” citation sums using Petersen et al.'s (2019) measure. In each of these cases, our general findings hold.

4.2. Discretizing the university-to-subfield citation matrix and measuring nestedness with RCA, NODF, and entropy

Most successful ecological nested models, in particular those applied to international trade networks, use a revealed comparative advantage (RCA) to discretize their matrix data. RCA is defined as the fraction of a country's exports in some product x divided by the proportion of world exports of x. If this ratio is >1, a comparative advantage is revealed for this product and the entry for this country and product is set to be 1;

---

[2] Since we use citations as a proxy for recognition, we needed to determine how and when to censor that recognition. For our data, we limited the citations received to a window of 5 years after the paper was published in year t, between year $t+1$ and $t+5$. We tested and compared two other three citation windows of 2 years and 10 years after the publication year with our main results. We find that our overall trends remain consistent across all windows.

otherwise, it is set to 0. Exports and citations share a similar data structure in that there is a receiver, a sender, and some weighted exchange among them. Here, we are interested in what constitutes being acknowledged in a subfield by peer universities, so we adopt the RCA as our discretization approach, consistent with what is used in organizational and trade ecology papers that measure nestedness. For each $Field_t$, the revealed comparative advantage of a university *org* in subfield *i*, $RCA_{org,i}$, is defined as:

$$\mathrm{RCA}_{\mathrm{org,i}} = \frac{\frac{\mathrm{X_{org,i}}}{\sum_{\mathrm{j \in Field}} \mathrm{X_{org,j}}}}{\frac{\mathrm{X_i}}{\sum_{\mathrm{j \in Field}} \mathrm{X_j}}} \geq 1$$

where *Field* is the set of all the subfields in the field, where subfield $i \in Field$; $X_{org,i}$ are the citations received by university *org* in subfield *i*; X_i are the citations received within subfield *i*; $\sum_{j \in Field} X_{org,j}$ is the sum of all the citations received by university *org* across all subfields *j* in the field; and $\sum_{j \in Field} X_j$ is the total of all the citations received across all subfields *j* in *Field*. So, a university with an RCA ≥ 1 is recognized in that area more than other universities at or below the field average. Otherwise, the organization is at or below the field average and is not recognized in the subfield.

While a detailed exposition can be found in Almeida-Neto et al. (2008), we recapitulate their approach here. The "**N**estedness based on paired **O**verlap and **D**ecreasing **F**ill" (NODF) measure is based on two features of an absence-presence matrix like Ψ_{Field_t}: (1) decreasing fill (or **DF**) and (2) paired overlap (or **PO**), as shown and described in Fig. 2. A given absence-presence matrix like the discretized university-to-subfield citation matrix Ψ_{Field_t} contains *n* rows and *m* columns, where row *i* is located in an upper position in the matrix from row *j* and where column *k* is located to the left of column *l*.

MT is defined as the marginal total (i.e., the sum of 1 s) of any column or row in the matrix. DF, decreasing fill, assumes only one of two values for any pair of rows or columns based on the MT of these rows or columns: 0 or 100. For instance, consider row *i* and row *j* in the matrix, where row *i* is higher in the matrix than row *j*. If $\mathrm{MT_i} > \mathrm{MT_j}$, then $\mathrm{DF_{ij}}$ is assigned a value of 100. This means that the row sum for row *i* is greater than the row sum for row *j*. However, if this is not the case, $\mathrm{DF_{ij}}$ is assigned a value of 0.

With this, $\mathrm{PO_{ij}}$ is the percentage of 1 s in row *j* that are located at identical column positions to the 1 s observed in row *i*. In other words, for all the presences (1 s) in rows *i* and *j*, what percentage of the *n* columns do both rows have presences in them? By extension, paired overlap (PO_{kl}) is the percentage of 1 s in column *l* that are located at identical row positions in column *k*. In other words, for all the presences (1 s) in columns *k* and *l*, what percentage of the *n* rows do both columns have presences in them? For each left-to-right column pair and each up-to-down row pair, there exists some degree of paired nestedness ($\mathrm{N_{paired}}$) where:

$$\text{if } \mathrm{DF_{paired}} = 0, \text{then } \mathrm{N_{paired}} = 0$$

$$\text{if } \mathrm{DF_{paired}} = 100, \text{then } \mathrm{N_{paired}} = \mathrm{PO}$$

First, we calculated the paired nested degree for each pair of columns and for each pair of rows. Then, the total nestedness among columns was quantified as the average values for all pairs of columns. The total nestedness among rows follows the same procedure applied to columns. Finally, the degree of nestedness for the whole matrix is calculated as the sum of all values of paired nestedness divided by the total number of pairs (i.e., the sum of all combinations of pairs of columns *m* and pairs of *n* rows). We then calculate nestedness for all columns ($\mathrm{N_{col}}$) and all rows ($\mathrm{N_{row}}$) by summing all paired values of columns and rows across the *n* columns and *m* rows, given as $\sum N_{paired}$, and then averaging this sum by the *n*(*n*-1)/2 and *m*(*m*-1)/2 paired degrees of nestedness. As such, the measure of nestedness for a bipartite matrix with *n* rows and *m* columns is given by:

$$\mathrm{NODF} = \frac{\sum \mathrm{N_{paired}}}{\frac{\mathrm{n(n-1)}}{2} + \frac{\mathrm{m(m-1)}}{2}}$$

To render NODF scores comparable across matrices with different universities (rows) or subfields (columns), we normalize scores as a fraction of the maximum possible NODF score possible for the matrix as proposed by Song et al. (2017). The advantage of this is that standard NODF scores can be correlated to the geometry of the matrix (i.e., the number of rows and columns). This was done in **R** using the maxnodf package made by Hoeppke and Simmons (2021).

As an example, the unstructured and nested forms of Ψ_{Field_t} are shown in Fig. 3 for mathematics, biology, economics, and computer science in the year 2010. Fig. 3A is the discretized matrix, $\Psi_{Field_{2010}}$, where the cells in black are presences and the cells in whites are absences, as determined using the row and column *z*-score procedure. These data form universities (rows) to subfield (column) absence-presence matrices for each $Field_t$, where the cells in black are presences and the cells in whites are absences as determined by the RCA measure. Next the NODF rearranges rows and columns to form as nested of a structure as possible, where both re-arranged rows and columns progress in increasing degree order, as shown in Fig. 3B. Fig. 3C and D provides insights into the sampled field, with 3C revealing the subfields with the most, average, and least presences, and Fig. 3D showing the universities with the 10 most, average and median, and fewest number of presences.

We apply a scaled measure of diversity in the form of Shannon entropy (or entropy) to Ψ_{Field_t} to characterize the universities in a similar manner. Entropy describes the distribution of universities (as the rows) across subfields (as the columns) in Ψ_{Field_t}. Across each row in Ψ_{Field_t}, the research presences are converted from presences (or 1 s) to a proportion based on the total presences in the row (i.e., row sums), defined as p_{org}. Thus, for a university ($org \in Field_t$) in Ψ_{Field_t}, entropy is defined as the sum of these p_{org} across subfields ($subfield \in Field_t$):

$$Entropy_{\substack{org \\ \in Field_t}} = \sum_{\substack{subfield \\ \in Field_t}} -p_{org} log_2 p_{org}$$

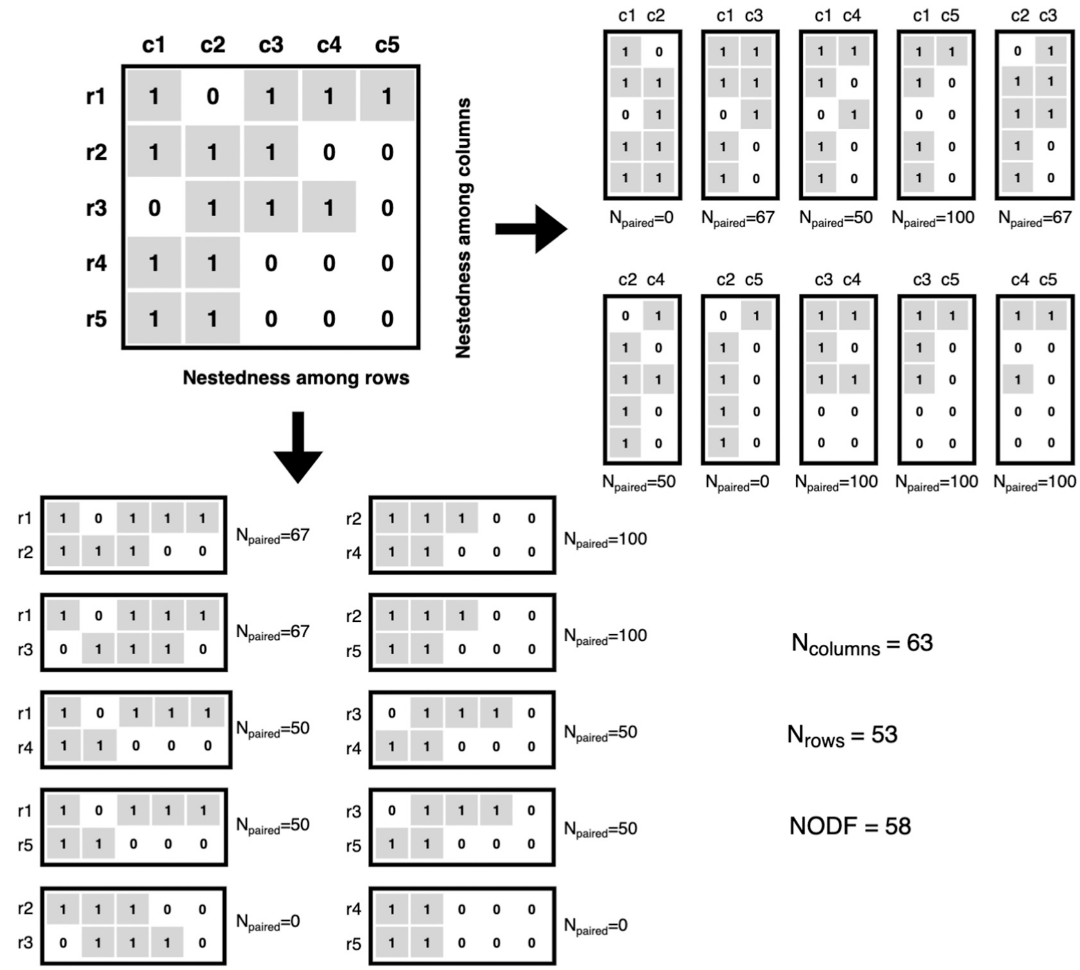
The lowest possible entropy is a university with a single presence across subfields in Ψ_{Field_t}, representing a p_{org} of 1 and an entropy of 0. With more presences across subfields, the higher the entropy score. So, universities with higher entropy scores are more fox-like relative to other universities, while those with lower entropy scores are relatively more hedgehog-like.

5. Results

We proposed that global scientific fields can be described as ecosystems using ecological measures like NODF and entropy. In Fig. 4, we first plot the NODF scores for each of our dozen fields over time between 1990 and 2012.[3] This is important when situating our results because higher nestedness over time could imply that global science is a robust and thriving ecosystem, in much the same way that higher nestedness implies this for biological ecosystems. We find that normalized NODF across fields significantly increases over time ($\mathrm{r}(276) = 0.52, p < 0.001$). Since NODF is a measure of order in ecological systems, it implies that not only is there a clearer structure to global scientific production across fields, but acknowledgments are more distributed and proportional across universities around the world. Since higher nestedness in ecological networks describes robust and vibrant ecologies that are both stable and expansive, this also could imply a global scientific research

---

[3] Since our data extend to 2017 and we use a five-year citation window, we stop our plots in 2012 that reflects all of the citations received between 2012 and 2017, inclusive.



	c1	c2	c3	c4	c5
r1	1	0	1	1	1
r2	1	1	1	0	0
r3	0	1	1	1	0
r4	1	1	0	0	0
r5	1	1	0	0	0

Fig. 2. How NODF is calculated (Image Source: Almeida-Neto et al., 2008). This illustration and explanation are taken from Almeida-Neto et al. (2008), which shows how nestedness is calculated according to the NODF approach. In this example, we calculate the paired nested degree for each pair of columns and for each pair of rows. Next, we measure the total nestedness among columns by taking the average values among all pairs of columns (right). Similarly, we follow the corresponding procedure among all pairs of rows (bottom). Finally, the matrix's nestedness is the sum of all combinations of pairs of columns and pairs of rows nestedness divided by the paired degrees of nestedness for a matrix with *n* rows and *m* columns (where both *n* and *m* in this example are 5).

enterprise that is intellectually diverse and growing.

5.1. Proposition 1: The most well-known and prominent universities (fox-like) globally are associated with higher entropy scores (acknowledged across more subfields) than less-known and less prominent universities (hedgehog-like)

Our first proposition argued that the universities that are more fox-like tended to be prominent in global science. We turn to plot trends among research institutes in Figs. 5 and 6. Fig. 5 plots the average entropy scores of different universities across fields for the years between 1990 and 2012. Fig. 5A plots the average entropy score of universities, where the green trend line represents the overall average of averages of all universities (the global average) and where we highlight several globally prominent universities: Harvard University, the University of Cambridge, and the University of Toronto. We also highlight universities with low entropy scores, as well. As shown in Fig. 5A, the average entropy is increasing ($r(4,289) = 0.53$, $p < 0.001$). And these prominent universities have the highest average entropy scores across fields and over time. Fig. 5B instead focuses on trends by country. Instead of looking at the average across countries, we take the maximum entropy score found in a country (based on the university within its borders) for a given $Field_t$ and take the national average of these maximum scores. The green trend line is the average of averages of these maximum country trends, reflecting an overall global average. Fig. 5B shows that the average maximum entropy score is rising across countries and for the world overall ($r(4,289) = 0.60$, $p < 0.001$). Fig. 5A and B shows that universities are generally becoming more fox-like than hedgehog-like.

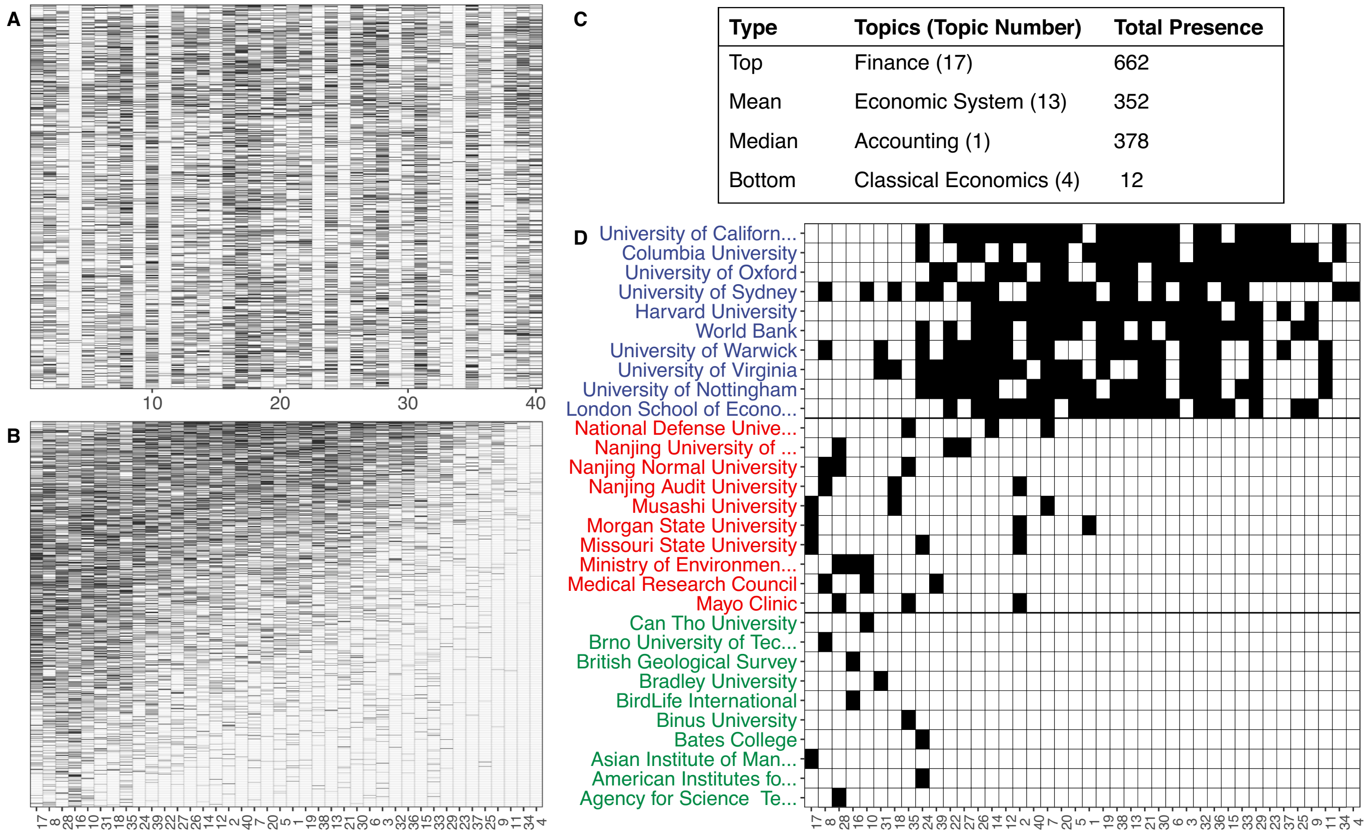
However, we also show how that these trends over time still exhibit differences among universities and their stature on the world's stage. To quantify the global prominence of universities, we use one of the oldest university rankings called the Academic Ranking of World Universities (AWRU) published by Shanghai Ranking Consultancy since 2009, and the accompanying GRID data scraped by (Selten et al., 2020). We relate the entropy scores of universities with their ranking in AWRU. For simplicity, we roughly parse the rankings into six categories[4]: 1–100, 101–200, 201–300, 301–400, 401–500, and unranked, where 1–100 is the highest ranked category, 101–200 is the second highest, and so forth.

Fig. 6A plots the average entropy scores using boxplots for universities with the ten highest average entropy scores on the left in blue text, the ten lowest average entropy scores on the right in red text, and the global average in the center in green text. Fig. 6B plots the maximum national entropy score, with the countries with the ten highest average

---

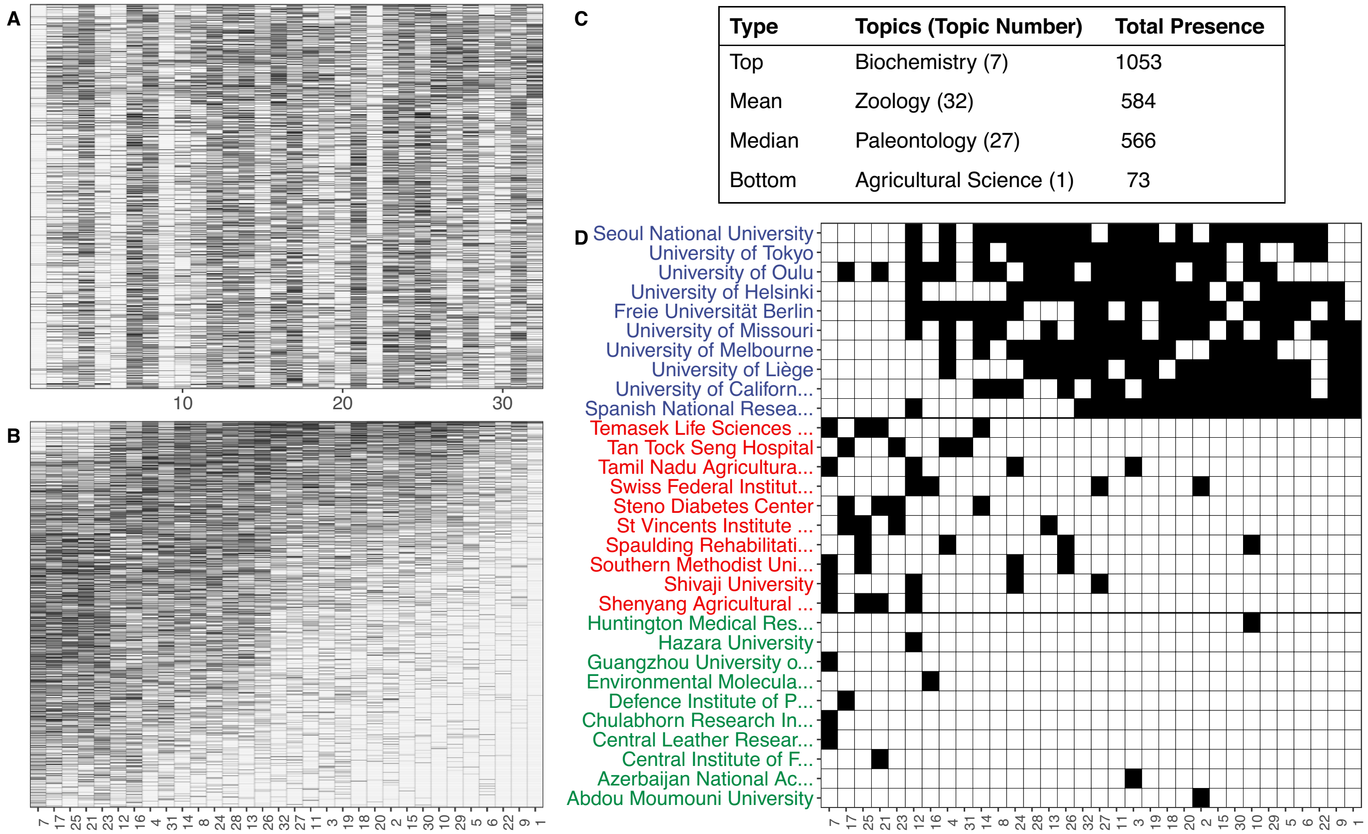
[4] Beyond the rank 200, the categories used by AWRU change over time. For instance, some years the ranking “201–300” might slightly change to be “201–300,” “201–302,” “203–300,” or “203–304” in different years. We take these other groupings to be mostly approximate to the ones we show here.

Economics 2010



Type	Topics (Topic Number)	Total Presence
Top	Finance (17)	662
Mean	Economic System (13)	352
Median	Accounting (1)	378
Bottom	Classical Economics (4)	12

Biology 2010



Type	Topics (Topic Number)	Total Presence
Top	Biochemistry (7)	1053
Mean	Zoology (32)	584
Median	Paleontology (27)	566
Bottom	Agricultural Science (1)	73

Fig. 3. Example of the university-to-subfield matrix, $\boldsymbol{\Psi}_{\mathit{Field}_t}$, for economics, biology, mathematics, and computer science in the year 2010. (A) The matrix of university presences is discretized using the revealed comparative advantage (RCA) approach to produce the university-to-subfield matrix, Ψ_{Field_t}. Darkened cells indicate a presence, while white cells indicate an absence. (B) The matrix Ψ_{Field_t} is reordered into a nested matrix based on row and column presences and their overlap. (C) Example of subfields in each field. Subfield presences are summed. Subfields are distinguished between those with many presences and those with few presences. (D) Example of universities and research institutes in each field: the top 10 (blue), middle 10 (red), and bottom 10 (green) universities in the nested matrix, based on presences. The more presences a university has across subfields, the more fox-like it is, and the fewer presences, the more hedgehog-like it is. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

maximum entropy scores on the left in blue text, the countries with the ten lowest average maximum entropy scores on the right in red text, and the global average of the maximum national entropy scores in the center in green text. Fig. 6C[5] pools all university entropy scores for each AWRU grouping in each of the twelve disciplines and plots the average as data points along with their standard errors.

Higher entropy scores mean that the university is acknowledged (or is present) across more subfields. Well-known universities and resource-wealthy countries are associated with the highest entropy scores on average, as shown in Fig. 6A and B. Across disciplines, the higher the universities' rank in AWRU, the higher the entropy score. This is clearly shown by the declining trends in Fig. 6C. Fig. 6 supports Proposition 1 that fox-like generalists are associated with being more prominent in global science.

5.2. *Proposition 2: Nestedness can be used to forecast the future subfields that universities will become recognized in by their peers*

We test our second proposition namely whether nestedness can be used to forecast future presences across subfields in future years. We follow a similar setup to Bustos et al. (2012) but for parsimony, we focus on the appearance of future presences across subfields rather than disappearance since university entropy is uniformly on the rise and acknowledgment across subfields is more common (Figs. 4 and 5). This not only demonstrates whether ecological frameworks mirror patterns of global scientific research and its emergent nested structures, but also serves to validate this framework using a predictive model.

Fig. 7 conceptually visualizes our analytic setup with the example of predicting a presence. For each field and year, $Field_t$, we take the currently observed matrix, Ψ_{Field_t}, and generate its hypothetical perfectly nested matrix (i.e., a matrix with a NODF of 1) that we define as, $\widehat{\boldsymbol{\Psi}_{\mathit{Field}_t}}$. To generate $\widehat{\boldsymbol{\Psi}_{\mathit{Field}_t}}$, we use the maxnodf package in **R** (Hoeppke and Simmons, 2021) which calculates the maximum NODF that can be achieved in the actual matrix with a given number of rows (universities), columns (subfields), and values (0 s and 1 s), calculated using the actual discretized matrix for year *t*. We then regress the actual observed future matrix in year $t + T$, $\Psi_{Field_{t+T}}$ for $Field_{t+T}$, on the hypothetical perfectly nested matrix for $Field_t$ that we generated, $\widehat{\boldsymbol{\Psi}_{\mathit{Field}_t}}$. We use a cross-classified logistic regression model with random effects for both universities ($RE^{Uni,t+n}$) and subfields ($RE^{Sub,t+n}$) present in those matrices. (This was done in **R** using the lme4 package by Bates et al., 2015.) Since we are interested in the odds of a future presence in years $t + 1$ through $t + 5$, we use logistic regression. As these matrices are bipartite in structure, a cross-classified model is appropriate since subfield presences (columns) also exist across multiple universities (rows), and vice-versa.

More formally, we regress the actual future value in year $t + T$ for the n^{th} university (row) and the m^{th} subfield (column) in $\Psi_{Field_{t+T}}$ (defined as $Y_{(n,k)}^{Future,t+T}$) on the corresponding n^{th} university (row) and the m^{th} subfield (column) in $\widehat{\boldsymbol{\Psi}_{\mathit{Field}_t}}$ the hypothetical perfectly nested matrix, (defined as $X_{(n,m)}^{Perfect,t}$). This cross-classified logistic regression model is defined as:

$$Y_{(n,m)}^{Future,t+T} = \beta_1 X_{(n,m)}^{Perfect,t} + RE^{Org,t+T} + RE^{Sub,t+T} + \alpha_{(n,m)}^{t,t+T} + \epsilon_{(n,m)}^{t,t+T.}$$

where $\alpha_{t,t+T}$ and $\epsilon_{(n,m)}^{t,t+T}$ are the intercept and the error, respectively, for this specific model.

We collect the statistically significant beta coefficients for $X^{Perfect,t}$ (the hypothetical perfectly nested matrix in year *t*, $\widehat{\boldsymbol{\Psi}_{\mathit{Field}_t}}$) and take the exponent of them[6] to report the higher or lower odds of a presence or absence in Fig. 8. Here, the odds ratios are plotted by *T* years in the future and by the NODF of the matrix, respectively. In total, we run 1380 of these models (i.e., twelve disciplines from 1990 to 2012[7] with each year compared to a five-year $t + T$ window). Running and analyzing these logistic models also serves to validate our overall framework: if nestedness did imply where future presences might appear, then ecological measures may be a viable way to better understand and characterize global scientific knowledge production.

When pooled and averaged across all models (i.e., fields, $t + T$ time frames, and over years), the odds of a predicted presence are 2.11 times the odds to occur than not ($N = 1380$, Fig. 8A and B). Fig. 8A is a plot of the average statistically significant (two-tailed *t*-test at least at the 0.05 significance level) beta coefficient from the models converted into odds ratios for each field and in each of its five $t + T$ time frames. Starting at the bottom row in Fig. 8A on the right and moving left, chemistry, biology, business, and medicine have among the highest odds ratios of a future presence, while economics, materials science, and environmental science are the lowest. Nevertheless, the odds are all quite high even in these cases. For instance, in economics across $t + T$ time frames, a future presence is on average 1.5 times the odds than not. For medicine and biology, the odds increase the further out in the future the prediction goes, but for most fields, the prediction is stable, indicating that the future presence is not necessarily tied to how far in the future the prediction is.

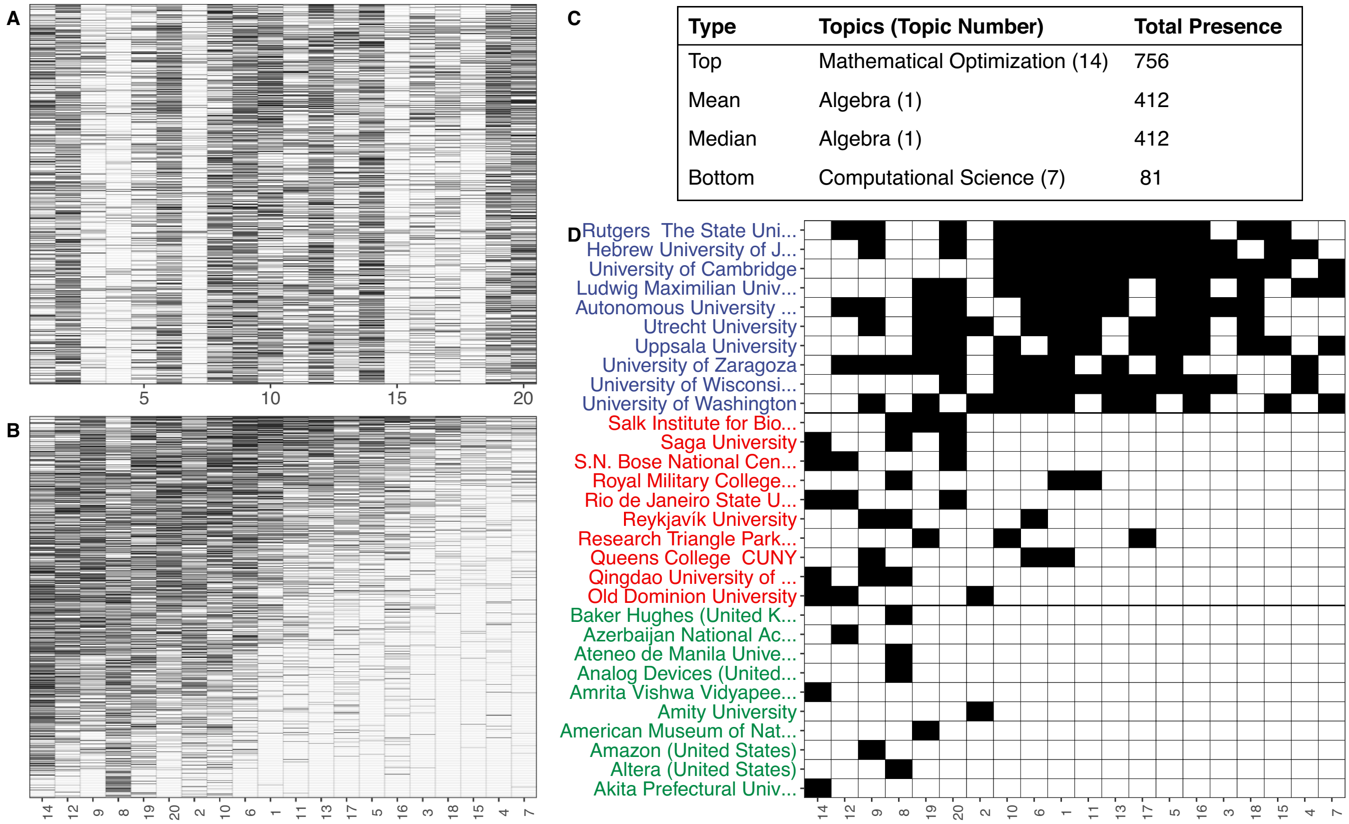
Finally, we generated the hypothetical matrix for each $Field_t$ based on its actual matrix. What is unclear is the relationship between nestedness of this actual matrix for $Field_t$ and whether it is associated with better odds ratios further in the future (e.g., $t + 5$ over $t + 1$). In other words, is nestedness related to better forecasting of future presences? In Fig. 9, we create scatterplots for each $t + T$ time frame of the odds ratios from each model on the *y*-axis by the normalized NODF score on the *x*-axis for each Ψ_{Field_t}. We further split these panes into two decades with data from the 1990s (upper row of panes) and data from 2000s (lower row of panes) to better identify any overall temporal trends that are distinct from future time frames (i.e., $t + 1$, $t + 2$, etc.). We find that higher degrees of nestedness are not associated with better forecasting the emergence of presences in future time windows. In both the 1990s and 2000s, there is a statistically significant trend between the odds ratios in early future years, $t + 1$ and $t + 2$, and field nestedness. However, the strength of this correlation is both weaker in magnitude and not statistically significant for later time frames, $t + 4$ and $t + 5$, in both the 1990s and 2000s. This implies that for later future time frames (i.e., $t + 3$, $t + 4$, and $t + 5$), higher odds ratios of forecasted presences are generally not associated with higher levels of nestedness. This also supports Proposition 2 that

---

[5] Fig. 6C covers data from 2005 to 2012, since AWRU rankings reliably start in 2005.

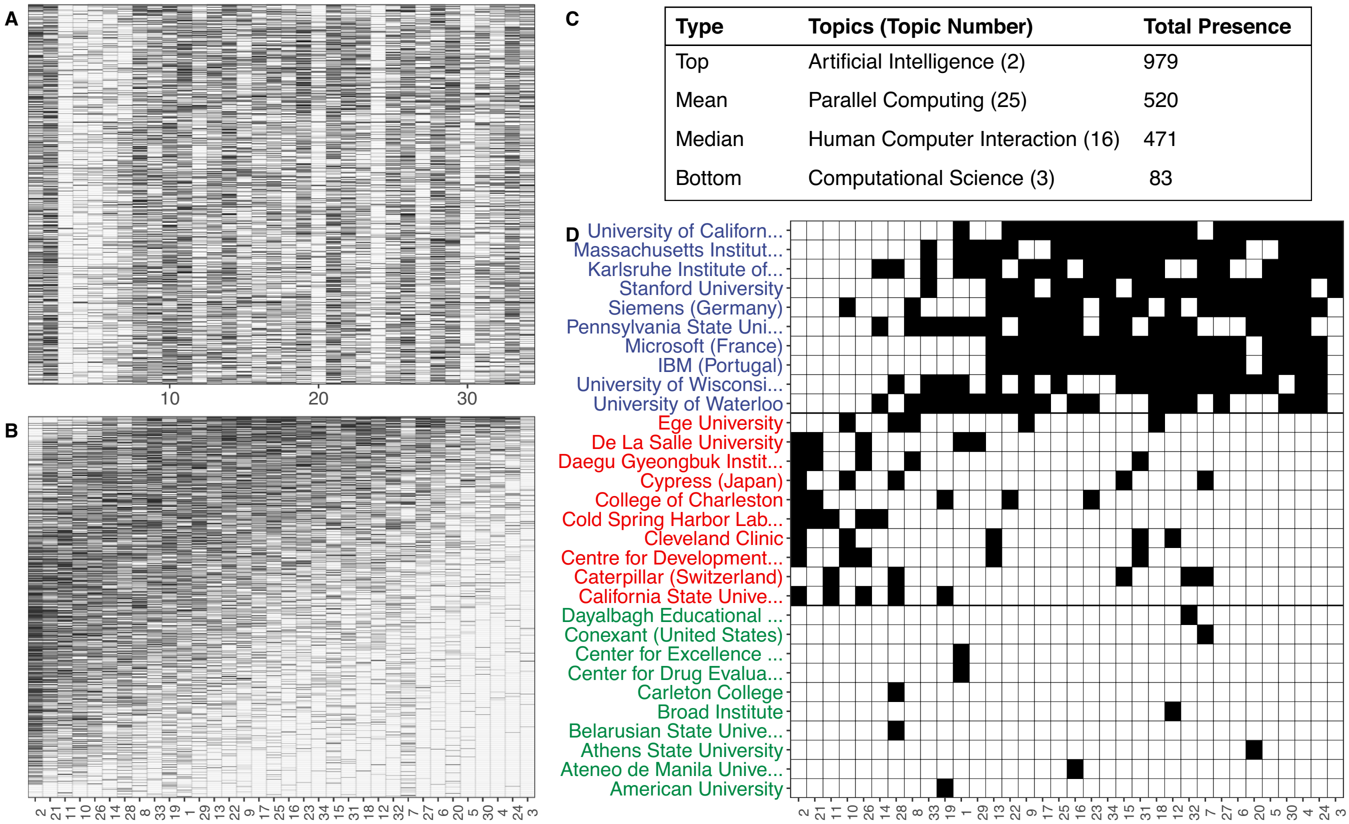
[6] We report e^{β_1}.

[7] As our analyses end in 2017, we end our analysis with 2012 since it is the last year with a reliable $T = 5$ citation time window.



Mathematics 2010

Type	Topics (Topic Number)	Total Presence
Top	Mathematical Optimization (14)	756
Mean	Algebra (1)	412
Median	Algebra (1)	412
Bottom	Computational Science (7)	81



Computer Science 2010

Type	Topics (Topic Number)	Total Presence
Top	Artificial Intelligence (2)	979
Mean	Parallel Computing (25)	520
Median	Human Computer Interaction (16)	471
Bottom	Computational Science (3)	83

Fig. 3. (*continued*).

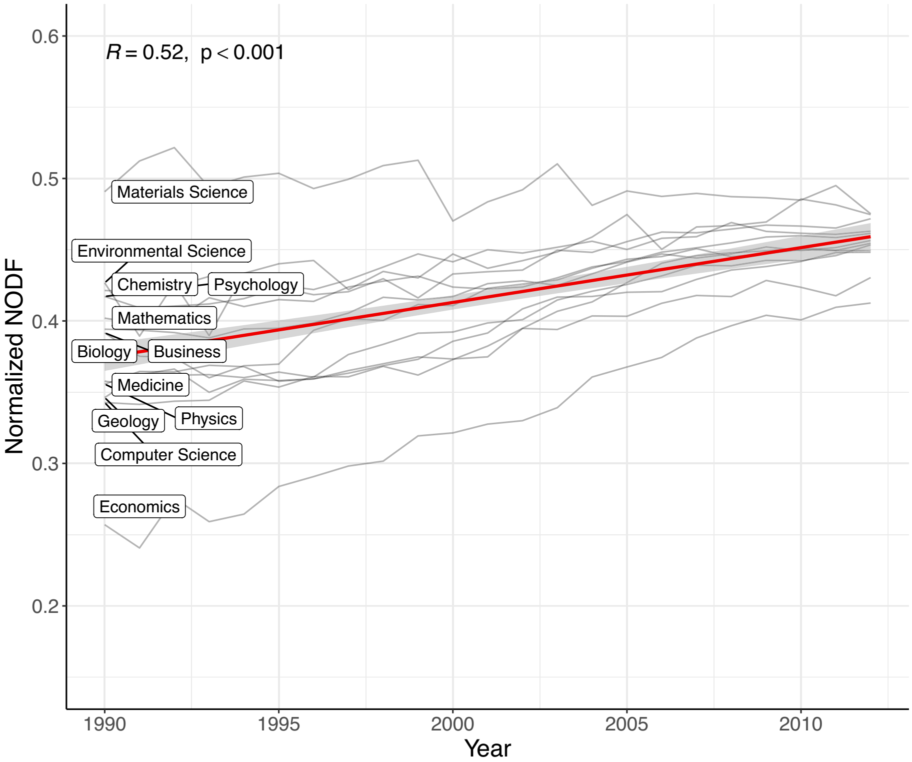


Fig. 4. Normalized NODF of Fields over Time. Normalized NODF scores over time across fields, including the overall trend, where a higher NODF score indicates greater nestedness in the field.

the inherent feature of nestedness can generally be used to forecast a university "inhabiting" a subfield in the future. In other words, neither lower nor higher levels of nestedness necessarily produce better forecasts. Instead, it is the mere feature of nestedness which can be used to forecast future presences.

6. Discussion

In this paper, we investigated what nested ecological networks can tell us about patterns in global science. We analyze the citations received by over 66 million papers in OpenAlex published between 1990 and 2017 across a dozen academic fields. Specifically, we identify which universities are recognized across subfields using revealed comparative advantages (RCAs) over time using the accumulated number of citations they receive for their published work (Figs. 4, 5, and 6). We find that nested ecological networks, as characterized by NODF and information entropy, applied to these data infer the future recognition of universities across subfields (Fig. 8).

First, we find that not only is there a clearer structure to global scientific production across fields, but acknowledgments are more distributed and proportional across universities around the world (Fig. 4). We describe universities across a spectrum between "hedgehogs" and "foxes" (Figs. 5 and 6). We find that higher-ranked universities tend to be more fox-like than lower-ranked universities, which tend to be more hedgehog-like. However, the details of the data selection and indicator measurements of these rankings are not transparent, but are likely based on citation counts. Hence, the correlations here could also be the result of the fact that the rankings and entropy actually measure the same dimension of performance. Nonetheless, rankings likely do not capture the spread of these citations across multiple subfields.

On average, the odds of a forecasted future acknowledgment are 2.11 times the odds to occur than not within a future five-year window (Fig. 8). Furthermore, more or less nestedness is not related to better future forecasts per se (Fig. 9); instead, nestedness itself is an inherent feature that can be used to forecast future presences. This implies that ecological frameworks do capture the dynamics of global scientific research in terms of acknowledgment among universities across fields.

Previous studies have successfully described the future structures of economic systems (e.g., international trade networks, supplier networks, etc.) using nested ecological network measures of nestedness and specialization (Brintrup et al., 2018; Bustos et al., 2012; Hidalgo and Hausmann, 2009; Ren et al., 2020; Saavedra et al., 2011, 2009; Scholl et al., 2021; Tacchella et al., 2012). While ecological metaphors are often invoked when describing the wider scientific enterprise (Gaziano, 1996; Leydesdorff, 2000; Papaioannou et al., 2009; Saxenian, 1996), we similarly extend and further develop these metaphors to those efforts that use bipartite nested ecological networks and their associated measures like NODF and entropy. Most of the metaphorically oriented work on ecologies focuses on firms and innovation. Our empirical focus is on the acknowledgment of knowledge production within the global scientific enterprise as opposed to regional innovation clusters or inter-firm market forces. Universities' scientific and research outputs are often taken as thermometers for national scientific prowess in the forms of various ranking schemes (Hazelkorn and Gibson, 2017; Moed et al., 1985; Shin et al., 2011). Surprisingly, few attempts have sought to bridge these disparate but complimentary strands, to expand beyond metaphors and use nested ecological networks towards global scientific knowledge production of universities and countries.

The potential of our analytic framework that applies nested ecological networks and its measures to universities' citation patterns across subfields is that it offers new ways to describe, model, and gain insights into the global scientific research enterprise. Our results here remain consistent despite our various modeling choices and assumptions. We argue that this is not the result of artifacts of our modeling choices or database effects unique to OpenAlex, but instead reflects latent ecologically oriented structures and dynamics within global scientific knowledge production.

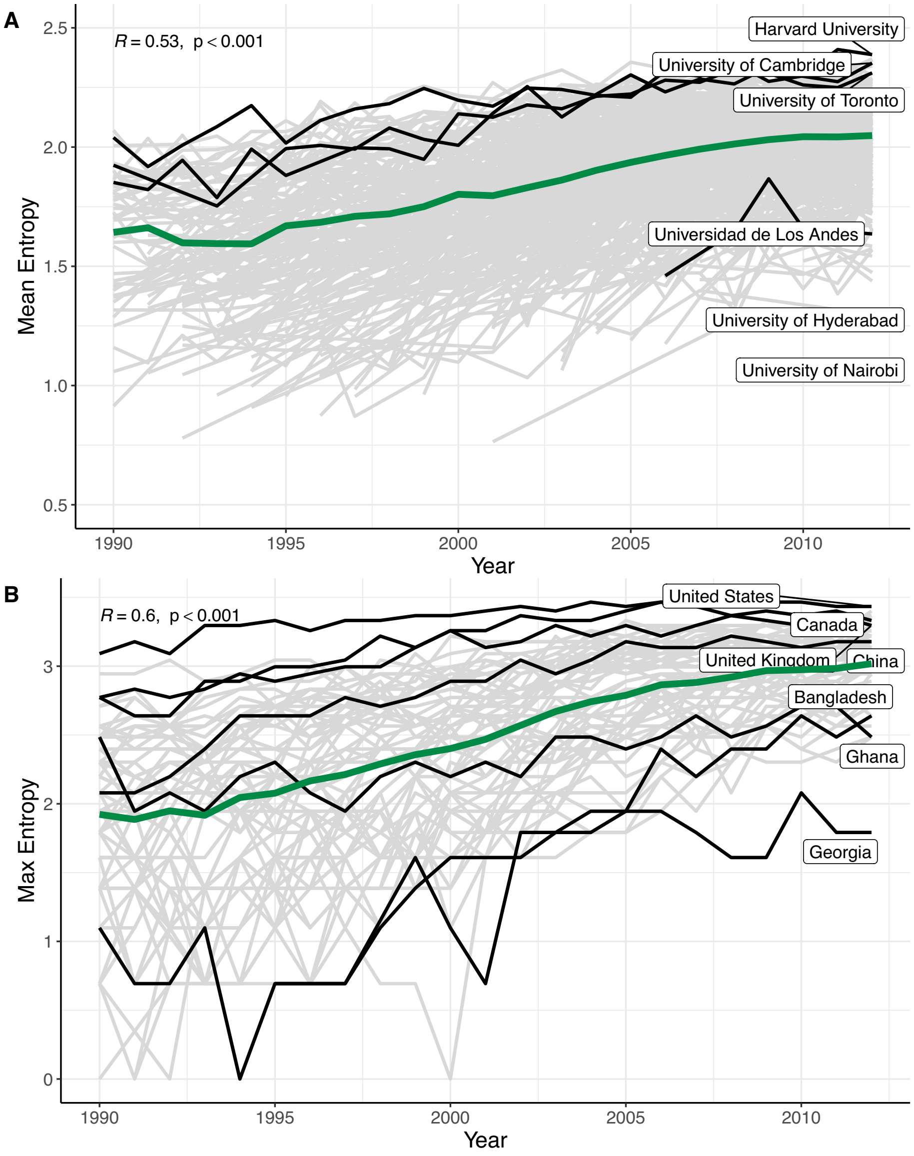


Fig. 5. Average entropy measures of universities and maximum entropy of countries. (A) Average entropy scores of universities over time. (B) The average maximum entropy scores of countries based on their universities over time. A higher entropy score is associated with acknowledgment in more subfields (more fox-like).

6.1. Policy implications

Applying nested ecological networks and their measures to global scientific knowledge production is significant to various efforts and concerns in the scientific research policy and innovation studies spaces. First, as universities no longer operate only within their national boundaries, frameworks for thinking about how they develop in transnational, interdependent contexts have become essential. Here, our findings can have practical implications. One outcome of science becoming global is increased competition for things like reputation and resources (Gomez et al., 2022), which global university rankings have made visible (e.g., Adler and Harzing, 2009; Hazelkorn, 2015). Relatedly, there has been a tendency among smaller universities to expand through university mergers to do better in global university rankings and possibly resemble those more fox-like universities (Docampo et al., 2015; Eastman and Lang, 2001). The concern is, however, that there are resource barriers to creating a highly recognized fox-like university. As such, smaller universities may miss chances to analyze their own focus areas and their strengths and weaknesses. New frameworks to situate and explore these questions are needed in higher education and scientific research policy circles.

Specifically, our findings on how universities are recognized across research areas and subfields have implications for higher education institutions that seek to identify their own profiling areas or potential areas of expertise. Understanding what kinds of unique research profiles universities could develop and where the most innovative connections exist may help higher education institutions make strategic decisions about how to compete for attention in the global higher education market. For example, in Finland, universities can apply for competitive university profiling funding from the primary research funder (Research Council of Finland) to strengthen and specialize in timely research areas and increase research quality (Pölönen and Auranen, 2021). Another example is universities in China, represented by their meteoric rise in international scientific rankings. Chinese researchers increasingly cite other researchers in China (Larivière et al., 2018) and rapidly increased the volume of publications (Wagner et al., 2022; Xie and Freeman, 2019). While our ecologically based framework still captures the rise in prominence of Chinese universities, they remain below where they should be given what is documented in other work on global science. This may be because we consider rates of citations from other

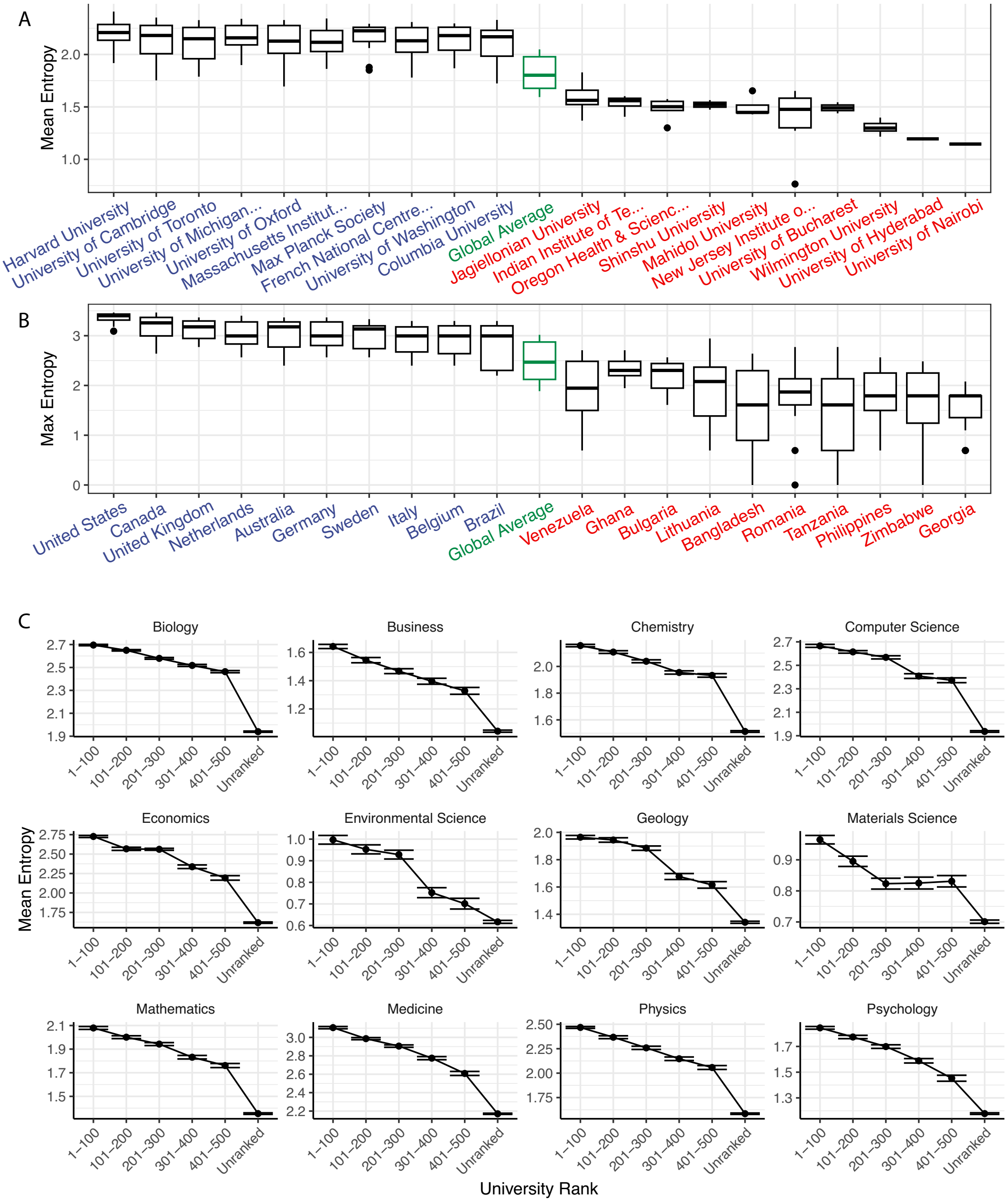


Fig. 6. Entropy of universities and countries with AWRU. (A) Boxplots of the average entropy scores of select universities. (B) Boxplots of the average maximum entropy scores of select countries. (C) Average entropy scores of universities by AWRU categories.

universities worldwide, which some claim that "clubbing effects" drive the meteoric rise in the prominence of Chinese universities as measured by citations (Tang et al., 2015). As such, our findings could help to better reveal the actual strengths and weaknesses of universities both domestically and internationally across fields and help reframe identification strategies for higher education and government science and funding agencies.

Finally, our findings may be used to think about what research areas may come into focus in the future—trying to determine the "next big thing" in science is extremely difficult and close to guesswork (Acuna et al., 2012; Penner et al., 2013). While our framework cannot be used to predict what research areas will come into focus next, it has the potential to shed light on what subfields might garner more attention than others. Such insights help universities develop successful research profiles based on their own demonstrated potential rather than predicting what other institutions will do.

6.2. Future work

The application of the ecological network approach we use here for global scientific research is intended to catalyze future projects on the dynamics and structure of scientific research. However, there are diverse and disparate avenues with which this framework can be applied beyond global science. For instance, organizational ecology researchers have yet to bridge its theoretical and empirical framework with social network analysis. The use of ecological bipartite measures offers such a promising methodological bridge. This framework is also applicable to other citation-based innovation data, like patent data to forecast the emergence or decline of intellectual and technological trends. Future work can also extend to more granular and specific nation or field-specific trends, rather than the wider, global approach we employ here. In addition, future work can relate the bipartite structure presented here (or its potential variants) to characteristics that define the

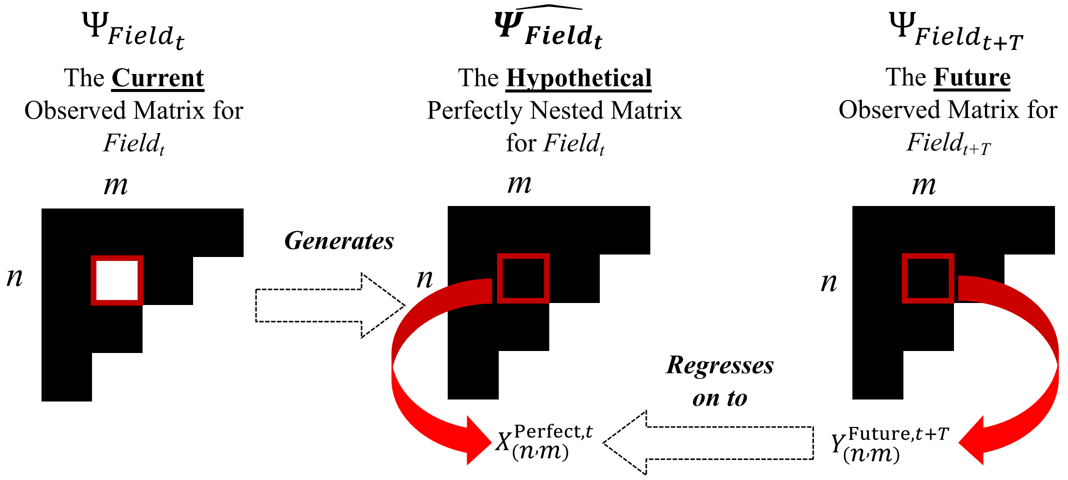


Fig. 7. Predicting future presences. We set up a cross-classified logistic regression by taking the actual observed matrix for $Field_t$ and generate its hypothetical perfect nested counterpart for $Field_t$. We then regress the values (n, m) in the future observed matrix in year $t + T$ on its corresponding values in the hypothetical matrix for year t.

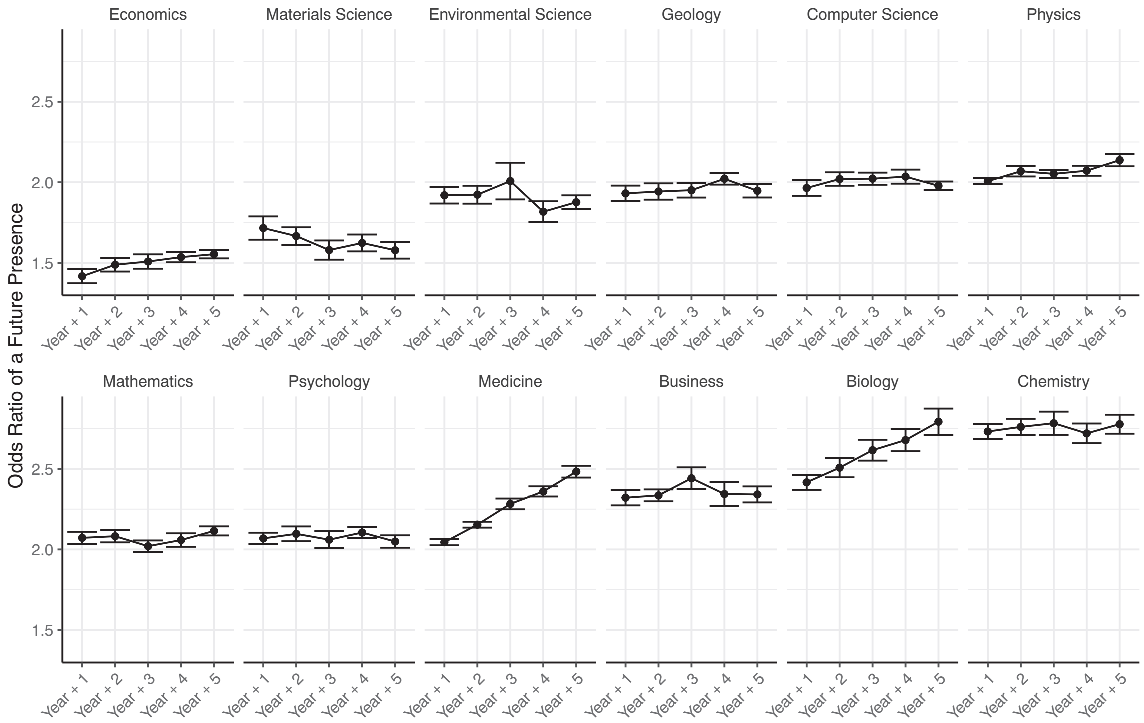


Fig. 8. Odds ratios of a future presence by discipline. For each discipline, the average odds ratio of a future presence from the cross-classified logistic regression models for $t + T$ years (e.g., current year t to $t + T$ future years, where T is 1, 2, 3, 4, or 5 years in the future).

intellectual vibrancy of the field, such as its rate of novelty, its intellectual core or research frontier (Chu and Evans, 2021; Cole and Cole, 1971; Evans et al., 2016), among other metrics. Finally, while we use citations as our focal unit of analysis, text metadata from scientific abstracts are increasingly available, as are the variety of topic models that are applied to these data. For instance, topic models applied to text data can uncover other intellectual or research trends latent in these data and employ the same ecological models used here to measure and track the structure and dynamics of scientific research at multiple scales.

7. Conclusion

We show that ecological metaphors of modern global scientific research reflect more meaningful and deeper structures and dynamics of research universities and recognized research areas across fields. Analyzing over 66 million papers from across a dozen academic fields published between 1990 and 2017, we use citation and subfield classification schemes to identify which subfields researchers from which universities and research institutes are distinctly recognized by using the number of citations they receive for work done in that particular area. Ecological patterns and trends are evident in global scientific knowledge production. We validate our application of an ecological framework by forecasting which research areas are eventually taken up by which universities and research institutes over time, running over a thousand logistic regression models. This holds implications for policies on the funding and evaluation of research and higher education institutes.

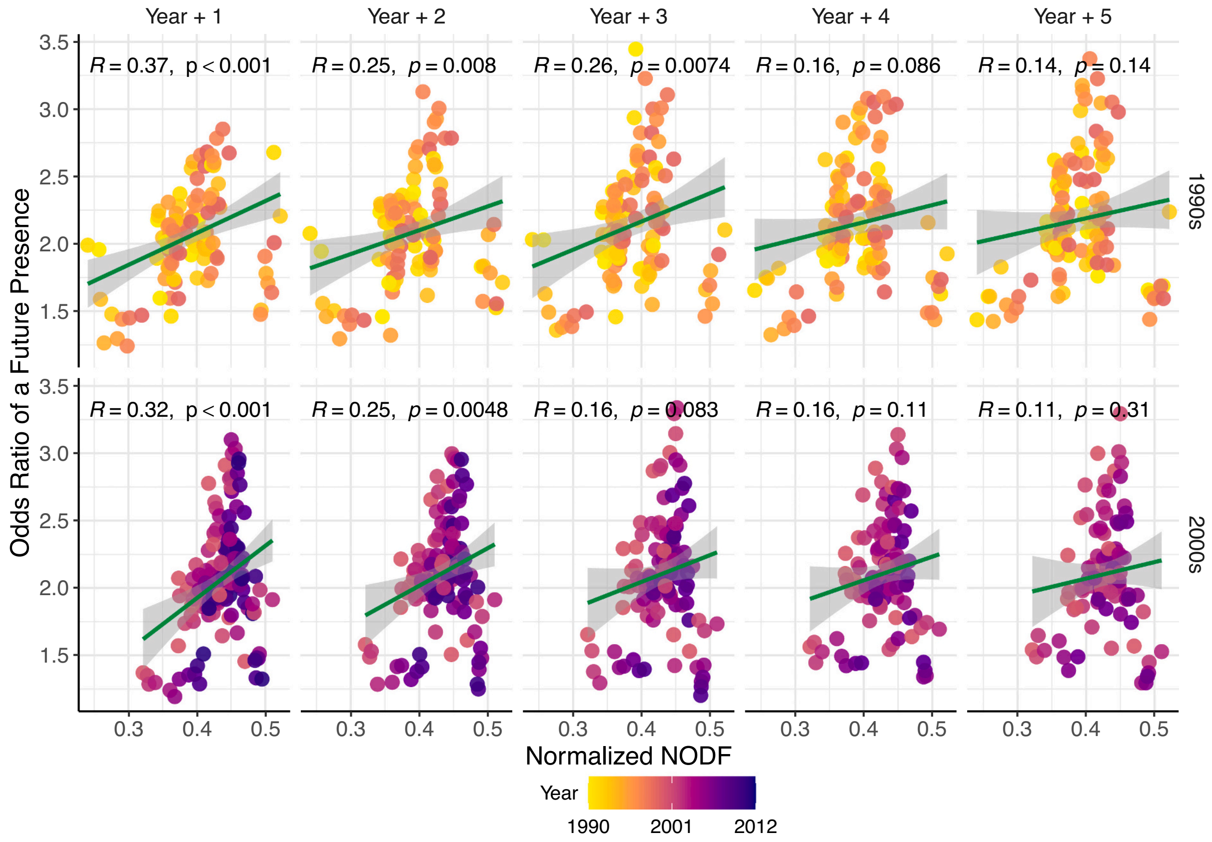


Fig. 9. Odds ratios of a future presence by NODF. For each future year *T*, the odds ratio of a future presence in $t+T$ future years on the *y*-axis by NODF on the *x*-axis, parsed by time frame and decade (1990s and 2000s), and color-coded by year. Odds ratios are how much more likely is a presence to occur than not in the future based on the hypothetical perfectly nested matrix in year *t*, $\widehat{\Psi_{Field_t}}$.

Funding

This project was funded by the National Science Foundation (NSF) grant #2022395.

CRediT authorship contribution statement

Charles J. Gomez: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Resources, Visualization, Writing – original draft, Writing – review & editing. **Dahlia Lieberman:** Formal analysis, Visualization, Writing – original draft. **Elina I. Mäkinen:** Investigation, Project administration, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The code and data are available on the Harvard Dataverse repository at the following link: https://doi.org/10.7910/DVN/2X0GHY.

Acknowledgments

We would like to thank Sebastian Muñoz-Najar Galvez, Jaime Sohn, the participants of the Queens College Department of Sociology's workshop series, and the participants of the University of Arizona School of Sociology's speaker symposium who all provided insightful and thoughtful comments on various versions of this article. All mistakes are our own.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.respol.2024.105040.

References

Acuna, Daniel E., Allesina, Stefano, Kording, Konrad P., 2012. Predicting scientific success. Nature 489 (7415), 201–202. https://doi.org/10.1038/489201a.

Adler, Nancy J., Harzing, Anne-Wil, 2009. When knowledge wins: transcending the sense and nonsense of academic rankings. Acad. Manag. Learn. Edu. 8 (1), 72–95.

Almeida-Neto, Mário, Guimarães, Paulo, Guimarães, Paulo R., Loyola, Rafael D., Ulrich, Werner, 2008. A consistent metric for Nestedness analysis in ecological systems: reconciling concept and measurement. Oikos 117 (8), 1227–1239. https://doi.org/10.1111/j.0030-1299.2008.16644.x.

Andrews, Raymond J., Fazio, Catherine, Guzman, Jorge, Liu, Yupeng, Stern, Scott, 2022. The startup cartography project: measuring and mapping entrepreneurial ecosystems. Research Policy 51 (2), 104437.

Bates, Douglas, Mächler, Martin, Bolker, Ben, Walker, Steve, 2015. Fitting linear mixed-effects models using Lme4. J. Stat. Softw. 67 (1) https://doi.org/10.18637/jss.v067.i01.

Baum, Joel A.C., Singh, Jitendra V., 1994. Organizational niches and the dynamics of organizational founding. Organ. Sci. 5 (4), 483–501.

Berlin, Isaiah, 1953. The Hedgehog and the Fox: An Essay on Tolstoy's View of History. Simon and Schuster.

Brintrup, Alexandra, Barros, Jose, Tiwari, Ashutosh, 2018. The nested structure of emergent supply networks. IEEE Syst. J. 12 (2), 1803–1812. https://doi.org/10.1109/JSYST.2015.2493345.

Bustos, Sebastian, Yıldırım, Muhammed A., 2022. Production ability and economic growth. Research Policy 51 (8), 104153.

Bustos, Sebastián, Gomez, Charles, Hausmann, Ricardo, Hidalgo, César A., 2012. The dynamics of Nestedness predicts the evolution of industrial ecosystems. PloS One 7 (11), e49393. https://doi.org/10.1371/journal.pone.0049393.

Carroll, Glenn R., Swaminathan, Anand, 1992. The organizational ecology of strategic groups in the American brewing industry from 1975 to 1990. Ind. Corp. Chang. 1 (1), 65–97.

Cavan, Ruth Shonle, 1983. The Chicago School of Sociology, 1918-1933. Urban Life 11 (4), 407–420.

Chu, Johan S.G., Evans, James A., 2021. Slowed canonical Progress in large fields of science. Proc. Natl. Acad. Sci. 118 (41) https://doi.org/10.1073/pnas.2021636118.

Cole, Jonathan, and Stephen Cole. 1971. "Measuring the Quality of Sociological Research: Problems in the Use of the 'Science Citation Index.'" Am. Sociol. 6(1): 23–29.

Collins, Harry, 2004. Gravity's Shadow: The Search for Gravitational Waves. University of Chicago Press.

Docampo, Domingo, Egret, Daniel, Cram, Lawrence, 2015. The effect of university mergers on the Shanghai ranking. Scientometrics 104, 175–191.

Eastman, Julia, Lang, Daniel, 2001. Mergers in Higher Education: Lessons from Theory and Experience. University of Toronto Press.

Eaton, Charlie, Sheisha Kulkarni, Robert Birgeneau, Henry Brady, and Michael Hout. 2019. "The organizational ecology of college affordability: research activity, state Grant aid policies, and student debt at U.S. Public Universities." *Socius* 5: 2378023119862409. doi:https://doi.org/10.1177/2378023119862409.

Evans, Eliza D., Gomez, Charles J., McFarland, Daniel A., 2016. Measuring Paradigmaticness of disciplines using text. Sociological Science 3, 757–778. https://doi.org/10.15195/v3.a32.

Gaziano, Emanuel, 1996. Ecological metaphors as scientific boundary work: innovation and Authority in Interwar Sociology and Biology. Am. J. Sociol. 101 (4), 874–907. https://doi.org/10.1086/230783.

Glänzel, Wolfgang, Thijs, Bart, 2004. Does co-authorship inflate the share of self-citations? Scientometrics 61 (3), 395–404.

Gomez, Charles J., Herman, Andrew C., Parigi, Paolo, 2022. Leading countries in global science increasingly receive more citations than other countries doing similar research. Nat. Hum. Behav. 6 (7), 919–929. https://doi.org/10.1038/s41562-022-01351-5.

González, Martín, Ana M., Vázquez, Diego P., Ramos-Jiliberto, Rodrigo, Lee, Sang Hoon, Miele, Vincent, 2020. Core-periphery structure in mutualistic networks: an epitaph for Nestedness? BioRxiv. https://doi.org/10.1101/2020.04.02.021691.

Graf, Holger, Kalthaus, Martin, 2018. International research networks: determinants of country embeddedness. Research Policy 47 (7), 1198–1214.

Grimm, Volker, Revilla, Eloy, Berger, Uta, Jeltsch, Florian, Mooij, Wolf M., Railsback, Steven F., Thulke, Hans-Hermann, Weiner, Jacob, Wiegand, Thorsten, DeAngelis, Donald L., 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310 (5750), 987–991.

Hazelkorn, Ellen, 2015. Rankings and the Reshaping of Higher Education: The Battle for World-Class Excellence. Springer.

Hazelkorn, Ellen, Gibson, Andrew, 2017. Global science, National Research, and the question of university rankings. Palgrave Communications 3 (1), 21. https://doi.org/10.1057/s41599-017-0011-6.

Hidalgo, César A., Hausmann, Ricardo, 2009. The building blocks of economic complexity. Proc. Natl. Acad. Sci. 106 (26), 10570–10575.

Hoeppke, Christoph, Simmons, Benno I., 2021. Maxnodf: an R package for fair and fast comparisons of Nestedness between networks. Methods Ecol. Evol. 12 (4), 580–585.

Hofstra, Bas, Kulkarni, Vivek V., Galvez, Sebastian Muñoz-Najar, He, Bryan, Jurafsky, Dan, McFarland, Daniel A., 2020. The Diversity–Innovation Paradox in Science. Proc. Natl. Acad. Sci. 117 (17), 9284–9291.

Jones, Benjamin F., Wuchty, Stefan, Uzzi, Brian, 2008. Multi-university research teams: shifting impact, geography, and stratification in science. Science 322 (5905), 1259–1262. https://doi.org/10.1126/science.1158357.

Kapturkiewicz, Agata, 2021. Varieties of entrepreneurial ecosystems: a comparative study of Tokyo and Bangalore. Research Policy 104377.

Kwiek, Marek, 2020. Internationalists and locals: international research collaboration in a resource-poor system. Scientometrics 124 (1), 57–105.

Lafuente, Esteban, Ács, Zoltán J., Szerb, László, 2021. A composite Indicator analysis for optimizing entrepreneurial ecosystems. Research Policy 104379.

Larivière, Vincent, Gong, Kaile, Sugimoto, Cassidy R., 2018. Citations strength begins at home. Nature 564 (7735), S70–S71. https://doi.org/10.1038/d41586-018-07695-1.

Leendertse, Jip, Schrijvers, Mirella, Stam, Erik, 2021. Measure twice, cut once: entrepreneurial ecosystem metrics. Res. Policy 104336.

Leydesdorff, Loet, 2000. The triple Helix: an evolutionary model of innovations. Research Policy 29 (2), 243–255.

Leydesdorff, Loet, Carley, Stephen, Rafols, Ismael, 2012. Global maps of science based on the new web-of-science categories. Scientometrics 1–5. https://doi.org/10.1007/s11192-012-0784-8.

Liu, Bin, Gao, Jianbo, 2022. Normality in the distribution of revealed comparative advantage index for international trade and economic complexity. Appl. Sci. 12 (3), 1125.

Lundberg, Jonas, 2007. Lifting the crown—citation z-score. J. Informet. 1 (2), 145–154.

McFarland, Daniel A., Moody, James, Diehl, David, Smith, Jeffrey A., Thomas, Reuben J., 2014. Network ecology and adolescent social structure. Am. Sociol. Rev. 79 (6), 1088–1121. https://doi.org/10.1177/0003122414554001.

Moed, H.F., Burger, W.J.M., Frankfort, J.G., Van Raan, A.F.J., 1985. The use of bibliometric data for the measurement of university research performance. Research Policy 14 (3), 131–149. https://doi.org/10.1016/0048-7333(85)90012-5.

Mok, Ka Ho, Chan, Ying, 2008. International benchmarking with the best universities: policy and practice in mainland China and Taiwan. High Educ. Pol. 21 (4), 469–486.

Nielsen, Mathias Wullum, Andersen, Jens Peter, 2021. Global citation inequality is on the rise. Proc. Natl. Acad. Sci. 118 (7) https://doi.org/10.1073/pnas.2012208118.

Papaioannou, Theo, Wield, David, Chataway, Joanna, 2009. Knowledge ecologies and ecosystems? An empirically grounded reflection on recent developments in innovation systems theory. Environment and Planning C: Government and Policy 27 (2), 319–339. https://doi.org/10.1068/c0832.

Pardo-Guerra, Juan Pablo, 2022. The Quantified Scholar: How Research Evaluations Transformed the British Social Sciences. COLUMBIA University Press.

Penner, Orion, Petersen, Alexander M., Pan, Raj K., Fortunato, Santo, 2013. Commentary: the case for caution in predicting Scientists' future impact. Phys. Today 66 (4), 8–9. https://doi.org/10.1063/PT.3.1928.

Petersen, Alexander M., Pan, Raj K., Pammolli, Fabio, Fortunato, Santo, 2019. Methods to account for citation inflation in research evaluation. Research Policy 48 (7), 1855–1865. https://doi.org/10.1016/j.respol.2019.04.009.

Pölönen, Janne, and Otto Auranen. 2021. "Research Performance and Scholarly Communication Profile of Competitive Research Funding: The Case of Academy of Finland." Pp. 901–12 in *INTERNATIONAL CONFERENCE ON SCIENTOMETRICS AND INFORMETRICS*. Vol. 18.

Priem, Jason, Piwowar, Heather, Orr, Richard, 2022. OpenAlex: A Fully-Open Index of Scholarly Works, Authors, Venues, Institutions, and Concepts.

Puuska, Hanna-Mari, Muhonen, Reetta, Leino, Yrjö, 2014. International and domestic co-publishing and their citation impact in different disciplines. Scientometrics 98, 823–839.

Ren, Zhuo-Ming, Zeng, An, Zhang, Yi-Cheng, 2020. Bridging Nestedness and economic complexity in multilayer world trade networks. Humanities and Social Sciences Communications 7 (1), 1–8. https://doi.org/10.1057/s41599-020-00651-3.

Saavedra, Serguei, Reed-Tsochas, Felix, Uzzi, Brian, 2009. A simple model of bipartite cooperation for ecological and organizational networks. Nature 457 (7228), 463–466. https://doi.org/10.1038/nature07532.

Saavedra, Serguei, Stouffer, Daniel B., Uzzi, Brian, Bascompte, Jordi, 2011. Strong contributors to network persistence are the Most vulnerable to extinction. Nature 478 (7368), 233–235. https://doi.org/10.1038/nature10433.

Saracco, Fabio, Di Clemente, Riccardo, Gabrielli, Andrea, Squartini, Tiziano, 2015. Randomizing bipartite networks: the case of the world trade web. Sci. Rep. 5 (1), 10595 https://doi.org/10.1038/srep10595.

Saxenian, AnnaLee, 1996. Regional Advantage: Culture and Competition in Silicon Valley and Route 128. Harvard University Press, Cambridge, Mass.

Saxenian, AnnaLee, 2007. The New Argonauts: Regional Advantage in a Global Economy. Harvard University Press, Cambridge, Mass.

Scholl, Maarten P., Calinescu, Anisoara, Doyne Farmer, J., 2021. How market ecology explains market malfunction. Proc. Natl. Acad. Sci. 118 (26), e2015574118.

Selten, Friso, Neylon, Cameron, Huang, Chun-Kai, Groth, Paul, 2020. A longitudinal analysis of university rankings. Quantitative Science Studies 1 (3), 1109–1135.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27 (3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Sherwin, William B., Prat, Narcis, i Fornells., 2019. The introduction of entropy and information methods to ecology by Ramon Margalef. Entropy 21 (8), 794.

Shin, Jung Cheol, Toutkoushian, Robert K., Teichler, Ulrich, 2011. University Rankings: Theoretical Basis. Springer Science & Business Media, Methodology and Impacts on Global Higher Education.

Song, Chuliang, Rudolf P. Rohr, and Serguei Saavedra. 2017. "Why Are Some Plant–Pollinator Networks More Nested than Others?" edited by A. Eklöf. *Journal of Animal Ecology* 86(6):1417–24. doi:https://doi.org/10.1111/1365-2656.12749.

Tacchella, Andrea, Cristelli, Matthieu, Caldarelli, Guido, Gabrielli, Andrea, Pietronero, Luciano, 2012. A new metrics for Countries' fitness and Products' complexity. Sci. Rep. 2 (1), 723.

Tahamtan, Iman, Bornmann, Lutz, 2019. What do citation counts measure? An updated review of studies on citations in scientific documents published between 2006 and 2018. Scientometrics 121 (3), 1635–1684. https://doi.org/10.1007/s11192-019-03243-4.

Tang, Li, Shapira, Philip, Youtie, Jan, 2015. Is there a clubbing effect underlying Chinese research citation increases? J. Assoc. Inf. Sci. Technol. 66 (9), 1923–1932.

Ulrich, Werner, Almeida-Neto, Mário, Gotelli, Nicholas J., 2009. A Consumer's guide to Nestedness analysis. Oikos 118 (1), 3–17. https://doi.org/10.1111/j.1600-0706.2008.17053.x.

Uzzi, Brian, 1997. Social structure and competition in interfirm networks: the paradox of embeddedness. Adm. Sci. Q. 35–67.

Wagner, Caroline S., Leydesdorff, Loet, 2005. Network structure, self-organization, and the growth of international collaboration in science. Research Policy 34 (10), 1608–1618. https://doi.org/10.1016/j.respol.2005.08.002.

Wagner, Caroline S., Whetsell, Travis A., Leydesdorff, Loet, 2017. Growth of international collaboration in science: revisiting six specialties. Scientometrics 110 (3), 1633–1652.

Wagner, Caroline S., Whetsell, Travis A., Mukherjee, Satyam, 2019. International research collaboration: novelty, conventionality, and Atypicality in knowledge recombination. Research Policy 48 (5), 1260–1270. https://doi.org/10.1016/j.respol.2019.01.002.

Wagner, Caroline S., Zhang, Lin, Leydesdorff, Loet, 2022. A discussion of measuring the Top-1% Most-highly cited publications: quality and impact of Chinese papers. Scientometrics 1–15.

Wuestman, Mignon L., Hoekman, Jarno, Frenken, Koen, 2019. The geography of scientific citations. Research Policy 48 (7), 1771–1780. https://doi.org/10.1016/j.respol.2019.04.004.

Xie, Qingnan, Freeman, Richard B., 2019. Bigger than you thought: China's contribution to scientific publications and its impact on the global economy. Chin. World. Econ. 27 (1), 1–27.

Yuan, Huiai, and Lihong Zhao. 2022. "A Review of Revealed Comparative Advantage and the Measurement Tools." Pp. 132–35 in *Proceedings of the 2022 International Conference on E-business and Mobile Commerce*.