The MIT Press

Productivity and interdisciplinary impacts of Organized Research Units

Daniel J. Hicks[1,2] iD

[1]Data Science Initiative, UC Davis
[2]Cognitive and Information Sciences, UC Merced

ABSTRACT

Organized Research Units (ORUs) are nondepartmental units utilized by U.S. research universities to support interdisciplinary research initiatives, among other goals. This study examined the impacts of ORUs at one large public research university, the University of California, Davis (UC Davis), using a large corpus of journal article metadata and abstracts for both faculty affiliated with UCD ORUs and a comparison set of other faculty. Using regression analysis, I find that ORUs appeared to increase the number of coauthors of affiliated faculty, but did not appear to directly affect publication or citation counts. Next, I frame interdisciplinarity in terms of a notion of discursive space, and use a topic model approach to situate researchers within this discursive space. The evidence generally indicates that ORUs promoted multidisciplinarity rather than interdisciplinarity. In the conclusion, drawing on work in philosophy of science on inter- and multidisciplinarity, I argue that multidisciplinarity is not necessarily inferior to interdisciplinarity.

1. INTRODUCTION

While there is no formal definition, Organized Research Units (ORUs) are nondepartmental organizational units utilized by U.S. research universities to support clusters of researchers working on related topics. ORUs are typically organized internally—many researchers are university faculty; other researchers and staff are university employees—but funded externally. Examples might include museums, observatories, research stations, some large physical science labs (Geiger, 1990, p. 6); the numerous small "centers" and "labs" containing just one or two faculty and limited external funding are usually not counted as ORUs (Stahler & Tash, 1994, p. 542). As the examples suggest, some ORUs support research by providing an institutional home for highly capital-intensive research activities such as specimen collections or the maintenance of large or complex instruments. ORUs can also serve as focal points for recruiting external funding, either by demonstrating to funders that the university is actively engaged in a particular area of research (Geiger, 1990, p. 9) or by providing, for example, dedicated support staff for the grant-writing and -administering cycle.

However, at least since Geiger (1990), research policy scholars have theorized ORUs as key sites for bridging barriers between disciplines (interdisciplinarity) and between academic and social interests (extradisciplinarity) (Etzkowitz & Kemelgor, 1998; Geiger, 1990; Sá &

Oleksiyenko, 2011; Sommer, 1994). That is, it is thought that ORUs support research not just *materially* (with resources and support staff) but also *culturally* (creating a certain kind of research community).

The aim of the current project is to examine the impact of ORUs at one large public research university—the University of California, Davis (UC Davis)—in terms of both traditional bibliometric notions of productivity (papers written, citations received) as well as interdisciplinarity. In other words, have the ORUs at UC Davis promoted research productivity? And have they promoted interdisciplinarity?

To answer these questions, I link rosters of faculty affiliated with ORUs to publication metadata retrieved from Scopus. Importantly, I include not only faculty affiliated with ORUs, but also a comparison set of researchers who are affiliated with the same departments but are not affiliated with any ORU. I use regression models to control for variables, such as career length and gender, and a directed acyclic graph (DAG) and a sequence of models to examine the mechanisms by which ORUs might increase productivity.

To examine disciplinarity and interdisciplinarity, I introduce the conceptual framework of "discursive space" and situate researchers in this space by applying topic modeling—a text mining technique—to a large corpus of journal articles abstracts by both ORU-affiliated faculty and comparison faculty.

In brief, I find that the UC Davis ORUs have had productivity impacts, but likely did so by enabling researchers to work with more coauthors[1]. The analysis of "discursive space" suggests that the UC Davis ORUs have promoted multidisciplinarity rather than interdisciplinarity. In the conclusion, drawing on work in philosophy of science on inter- and multidisciplinarity, I argue that multidisciplinarity is not necessarily inferior to interdisciplinarity.

Note that, because I examine ORUs at a single institution, I do not claim that my results generalize to ORUs at any other institution.

1.1. Organized Research Units at UC Davis

At the time data collection began for this study (fall 2018), UC Davis had eight ORUs, each of which describes itself as engaged in interdisciplinary research or education. See Tables 1 and 2.

Four of these ORUs are dedicated to environmental topics, broadly constructed to include ecology, conservation biology, environmental science, and environmental policy: the Air Quality Research Center (AQRC); the Bodega Marine Lab/Coastal and Marine Science Institute (BML/CMSI); the Institute of Transportation Studies (ITS); and the John Muir Institute of the Environment (JMIE). Of these four ORUs, JMIE is the most heterogeneous, with distinct initiatives in climate change, data science in environmental science, energy systems, polar science, and water science and policy.

Three other ORUs are dedicated to biomedical topics: the Comprehensive Cancer Center (CCC); the Center for Healthcare Policy and Research (CHPR); and the Program in International and Community Nutrition (PICN). CCC has both research and clinical aspects,

---

[1] An anonymous reviewer suggests the possibility that these impacts are highly contingent on funding; if funding were to disappear, then these impacts might disappear as well. I agree that this is possible; but as none of the UC Davis ORUs experienced dramatic funding changes during the study period, it is outside of the scope of the current paper.

Table 1. UC Davis Organized Research Units [ORUs] examined in this study

Abbreviation	Full name	Founded in
AQRC	Air Quality Research Center	2005
BML/CMSI	Bodega Marine Laboratory & Coastal Marine Science Institute	1960 (BML)/2013 (CMSI)
CNPRC	California National Primate Research Center	1962
CCC	Comprehensive Cancer Center	2002
CHPR	Center for Healthcare Policy & Research	1994
ITS	Institute for Transportation Studies	1991
JMIE	John Muir Institute of the Environment	1997
PICN	Program in International & Community Nutrition	1987

and as the name indicates CHPR supports both academic research and policy analysis. PICN has a strong global focus, with active research projects in Laos; Haiti; Cameroon and Ethiopia; Burkina Faso, Ghana, and Malawi; the Gambia; Niger; Bangladesh; India; and Kenya.

The eighth ORU, CNPRC, is organized into four "units," devoted to infectious diseases, neuroscience and behavioral research, reproductive science and regenerative medicine, and respiratory diseases. As these labels suggest, CNPRC supports a mix of behavioral and biomedical research.

As Table 1 indicates, the ages of these ORUs vary substantially. BML is the oldest of the current ORUs, founded in 1960. CMSI, the youngest of the current ORUs, was formed in 2013 to coordinate research activities between BML (a single laboratory on the Pacific Ocean north of San Francisco) and other water research (such as the Tahoe Environmental Research Center in Incline Village, Nevada, on the north shore of Lake Tahoe in the Sierra Nevada mountains). BML/CMSI are treated as a single unit for the purposes of this study.

This study only considered ORU affiliations as of fall 2018; researchers who might have been affiliated with an ORU previously, but were no longer affiliated as of fall 2018, and were still actively publishing with a UC Davis affiliation during the period 2015–2017, would be considered not affiliated with any ORU.

1.2. Productivity and Discursive Impacts

1.2.1. Productivity impacts

Research evaluation often focuses on measures such as publication counts, citation counts, and perhaps patents or other indicators of economic impact (Hicks & Melkers, 2012). In the context of evaluating the effects of a particular (set of) programmatic interventions—namely, recruiting faculty to an ORU—I refer to these familiar kinds of outputs as *productivity impacts* of the intervention. Research evaluation might also consider productivity inputs, such as grant application success rate or quantity of external research funds received.

In this study, I examine three productivity impacts of the UC Davis ORUs. Publication and citation counts are familiar measures of research productivity. The third, coauthor count, is not usually used as a primary measure of productivity. However, it is highly plausible that increased collaboration—and so an increased number of coauthors—leads at least to increased

Table 2. Each ORU describes itself as engaged in interdisciplinary research or education. All quotations were retrieved on 28 May 2020

ORU	Interdisciplinary self-description
AQRC	"The AQRC provides support for teams of collaborative researchers to conduct scientific, engineering, health, social and economic research that educates and informs planning and regulations for air quality and climate change." (https://aqrc.ucdavis.edu/about)
BML/CMSI	"CMSI's 100+ affiliated faculty and staff and over 120 graduate students and postdoctoral scholars are internationally recognized for their expertise across the full spectrum of modern marine science, including ecology, evolutionary biology, conservation biology, microbiology, coastal oceanography, environmental toxicology, geochemistry, political science, natural resource management, economics, law, corporate sustainability, and marine wildlife health." (https://marinescience.ucdavis.edu/about)
CCC	"Center members are organized into programs, shared resources and innovation groups to transcend departmental affiliations and to be transdisciplinary, translational and transformative." (https://health.ucdavis.edu/cancer/research/membership/index.html)
CNPRC	"The California National Primate Research Center (CNPRC) is improving health and advancing science with interdisciplinary programs in biomedical research on significant human medical conditions throughout the lifespan." (https://cnprc.ucdavis.edu/about-us/)
CHPR	"Since our inception in 1994, the UC Davis Center for Healthcare Policy and Research (CHPR) has conducted interdisciplinary and collaborative research and research synthesis to improve health outcomes and services, educated the next generation of health services researchers, and assist policymakers in formulating effective health policies." (https://medicine.ucdavis.edu/chpr/)
ITS	"We have a strong commitment not just to research, but interdisciplinary education and engagement with government, industry, and non-governmental organizations." (https://its.ucdavis.edu/about/)
JMIE	"The founding of the John Muir Institute of the Environment in 1997 was a monumental achievement for visionary academics. It was the culmination of over 30 years of strong leadership in environmental research, interdisciplinary collaboration, and campus and system-wide endorsement." (https://johnmuir.ucdavis.edu/about)
PICN	"The educational curriculum provides students with the necessary knowledge and skills to master their chosen primary disciplines, while simultaneously exposing them to interdisciplinary research methods. The training program therefore combines courses in basic biological sciences, behavioral sciences, and social sciences, as well as interdisciplinary seminars in the planning, implementation, and evaluation of nutrition programs at the community and national levels." (https://globalnutrition.ucdavis.edu/about)

publication. Even if increased collaboration does not itself count as increased productivity, it is one potential mechanism by which ORUs might increase productivity. That is, along with providing (or facilitating the provision of) research funds and other material resources, ORUs might serve an important network function, encouraging researchers to work together more than they would have otherwise. I therefore include coauthor count as a potentially significant productivity impact.

1.2.2. Discursive impacts

Evaluating the productivity impacts of an interdisciplinary research program or organizational unit is, methodologically and conceptually, essentially the same as evaluating a disciplinary program or organizational unit: The same kinds of data will be collected and analyzed in the same way. In addition, productivity impacts abstract from the content of research. Counting publications doesn't consider what those publications are about.

But interdisciplinary research is typically justified in terms of distinctive pragmatic goals. For example, an unsigned editorial in *Nature* argues that "tackl[ing] society's challenges through research requires the engagement of multiple disciplines" (*Nature, 2016*). Following the distinctions made by Vanevar Bush and James Conant in the early Cold War period, Geiger contrasts disciplinary research with "programmatic research" (Geiger, 1990, p. 8). Geiger argues that the norms of disciplinary research are enforced by academic departments, making them "inherently [epistemically] conservative institutions." By contrast, ORUs (broadly understood to include museums, observatories, and extension offices) "exist to do what departments cannot do: to operate in interdisciplinary, applied, or capital-intensive areas in response to social demands for new knowledge" (Geiger, 1990, p. 17; see also Sá, 2008).

Interdisciplinary research may also have epistemic goals. Huutoniemi, Klein et al. (2010) distinguish epistemologically oriented and instrumentally oriented interdisciplinary research (and sometimes use a third category of "mixed orientation" interdisciplinary research) (see also Bruun, Hukkinen et al., 2005, pp. 29–30, 90–91). In a qualitative analysis of interdisciplinary research funded by the Academy of Finland, they find that the majority of interdisciplinary funding is directed towards (purely) epistemologically oriented projects (Bruun et al., 2005, p. 104), and that, weighted by funding, epistemologically oriented research is more likely to be deeply integrative (rather than merely multidisciplinary research) than instrumentally oriented research (Bruun et al., 2005, p. 106).

Even when interdisciplinary research has purely epistemic goals, we would expect these goals to be distinctive from those of disciplinary research. "Integration of various disciplinary perspectives is expected to lead to a more profound scientific understanding or more comprehensive explanations of the phenomena under study" (Huutoniemi et al., 2010, p. 85). Different disciplines are assumed to offer complementary perspectives on the phenomenon or subject. Then, bringing these complementary perspectives together is expected to produce qualitatively better knowledge than each could have produced on its own.

However, different disciplinary perspectives are not necessarily complementary. Different disciplines—or even lines of research within a given discipline—may depend on different metaphysical, epistemological, and methodological background assumptions (Cartwright, 1999; Eigenbrode, O'Rourke et al., 2007; Holbrook, 2013; Kuhn, 1996; Longino, 2013; Potochnik, 2017 ch. 7). These sets of background assumptions may be deeply incompatible with each other, and attempts to integrate them might be frustrating and unproductive. In other words, in this case, interdisciplinary research might be *less*, rather than more, than the sum of its disciplinary parts.

Insofar as interdisciplinary integration has been successful in a particular case—that is, insofar as a body of research has been interdisciplinary rather than *multidisciplinary*—a variety of theoretical perspectives predict that researchers will have to have produced a collection of material and linguistic affordances spanning the divide. In a material mode, Star and Griesemer (1989) examine "boundary objects" that circulate across disciplinary communities, serving as both shared objects of inquiry and shared sources of evidence. Work on "trading zones" has drawn on concepts from linguistics, such as pidgins and creoles, to analyze linguistic innovation in successful cross-disciplinary interactions (Collins, Evans, & Gorman, 2007; Galison, 1997). For example, Andersen (2012) analyzes a successful collaboration between chemists and physicists, stressing the need for "interlocking mental models" such that "the same concepts may form part of multiple lexica concerned with, for example, different aspects of a phenomenon" (Andersen, 2012, p. 281ff).

Citation data are often used in quantitative studies of interdisciplinarity (Wagner, Roessner et al., 2011). Cited works are classified into disciplines, often at the journal level (say, any paper published in *Cell* counts as a biology paper), and empirical distributions of citations across disciplines are used in various metrics of *variety*, *balance*, and/or *disparity/similarity* (Rafols & Meyer, 2010).

While citation data can help us assess the extent to which researchers draw on work across disciplines (Youtie, Kay, & Melkers, 2013), these data have limited ability to tell us to what extent researchers engage in different goals, pursue different research topics, or adopt different mental models. Accessing these features of research in large-scale quantitative studies likely requires textual data and methods from text mining and natural language processing (NLP). For example, Hicks, Stahmer, and Smith (2018) suggest that text mining methods might be useful for developing measures relating to what they call "outward-facing goals," defined as "the value of research for other [extra-academic] social practices." They focus specifically on nouns extracted from journal article abstracts, and show how clusters of these nouns can be matched to an existing taxonomy of basic human needs and values. Hofstra, Kulkarni et al. (2020) analyze abstracts of PhD dissertations from 1977 to 2015 to identify novel word associations, which they interpret as measures of innovation. They combine this analysis with name-based automated gender and race attributions and author-level bibliometric data from Web of Science to examine how the relationship between innovation and career success varies across demographic groups. However, neither of these papers examined interdisciplinary research as such.

In this paper, I propose that interdisciplinary research, as contrasted with disciplinary and multidisciplinary research, will have distinctive linguistic traces that can be detected using text mining methods.

Conceptually, I begin with the idea of *discursive space*, the space of research topics and conceptual schemes as they manifest in language. Figure 1 suggests how disciplinary and interdisciplinary researchers might be configured in this discursive space. There are two groups of disciplinary researchers, "red" and "blue." These researchers have simple primary colors and are clustered close together, indicating that they work on similar research topics, employ similar conceptual schemes, and more generally use similar language. The circles representing these researchers are small, indicating that they work on a relatively small set of topics. And the clusters are in distinct areas of discursive space, indicating that they differ substantially in their research topics and conceptual schemes. The clusters are internally homogeneous but externally heterogeneous.

Figure 1 also includes two interdisciplinary researchers. These researchers are shades of purple and are located in the space between the red and blue clusters, indicating that they
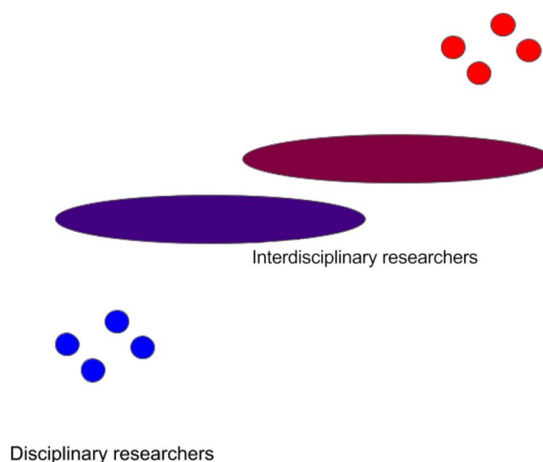
Figure 1. Conceptual model of "discursive space." Disciplinary researchers (red and blue dots) tend to be located close to members of the same discipline but far from members of other disciplines. Interdisciplinary researchers (purple ellipses) are located in between these disciplinary clusters.

use a mix of research topics, conceptual schemes, and language more generally from the two disciplines. The ellipses representing the interdisciplinary researchers are larger, indicating that they work on a relatively large set of topics. The shading and position of the researchers suggests that they have home departments or disciplines: One is a bluish purple, and is closer to the blues; the other is a redish purple, and is closer to the reds. But these interdisciplinary researchers are closer to each other than they are to their home disciplinary clusters.

More precisely, I suggest three hypotheses based on Figure 1 and the conceptual framework of "discursive space": compared to their departmental peers, interdisciplinary researchers will:

H1. have greater *discursive breadth*, that is, work on a wider variety of issues or use a wider variety of methods, and so have more linguistic diversity;
H2. be further from the *discursive central tendency* of their home departments, that is, the center of departmental clusters in discursive space; and
H3. be closer to their interdisciplinary peers than their departmental peers.

In a context where we expect an intervention to promote interdisciplinary research—for example, recruiting a faculty member to an ORU—I refer to these three hypotheses as the expected *discursive impacts* of the intervention. This concept of discursive impacts provides a framework for evaluating ORUs and other interdisciplinary research initiatives. Insofar as my hypotheses are correct and the UC Davis ORUs have effectively promoted interdisciplinary research, ORU-affiliated researchers should exhibit discursive impacts.

Bibliometricians and researchers in related areas of quantitative science studies have developed various similarity measures that might be interpreted as situating researchers (or other units of analysis, such as documents or journals) relative to each other in "space" (Boyack, Klavans, & Börner, 2005; Leydesdorff & Rafols, 2011; Wagner et al., 2011). Several common metrics are based on citations (Boyack et al., 2005, p. 355; Lu & Wolfram, 2012, p. 1974). *Bibliographic coupling* operationalizes similarity in terms of outgoing citations: Two units of analysis (documents, authors, journals) are similar insofar as they cite the same sources. *Cocitation analysis* works in the opposite direction, operationalizing similarity in terms of

incoming citations: Two units are similar insofar as they are cited by the same sources. A third citation-based metric is *intercitation analysis*, on which two units are similar insofar as they cite each other. (Calculations of intercitation similarity usually combine citations from *x* to *y* with citations from *y* to *x* to create a symmetric statistic; Boyack et al., 2005, p. 356.) Again, while citation data carry certain kinds of information, they do not seem to tell us much about the goals, research topics, or mental models used by researchers.

Another common similarity metric, *coword analysis* is based on word usage: Two units are similar insofar as they use the same terms. Coword data is usually combined with cosine similarity, which automatically adjusts for differences in total word count (for example, a senior researcher is likely to have a much greater total word count than a junior researcher; Leydesdorff & Rafols, 2011, p. 88).

In this study I use topic models to assess similarity. Topic models begin with document-term frequency data, as in coword analysis; for example, the term "researcher" appears in a certain document five times. The models then interpolate probability distributions of topics conditional on documents and terms conditional on topics. For example, a given document may "contain" 50% topic 1, 25% topic 2, 10% topic 3, and so on. Each topic, in turn, has a probability distribution over terms. For example, topic 1 might have "researcher" with probability 1%, "starfish" with probability 0.5%, "cancer" with probability 0.0001%, and so on. More formally, topic models begin with observed conditional probability distributions over terms, $p(term_w|document_i)$, and fit two probability distributions $\beta_{w,t} = p(term_w|topic_t)$ and $\gamma_{t,i} = p(topic_t|document_i)$.

Examining the quantitative science studies literature, I was unable to find any systematic reviews that compare topic models with other approaches to measuring similarity, such as coword or cocitation analysis. Lu and Wolfram (2012) found modest correlations (0.40–0.48, Kendall's τ-b) among coword, cocitation, and topic model-based measures of similarity; however, they used a small sample of only "the 50 most prolific authors" in a data set and selected the particular value of *k* used in their topic model "because it produced the most reasonable outcome by our judgment," with no further explanation or justification. Yan and Ding (2012) compare similarity networks constructed using bibliographic coupling, cocitation, coword, and topic model similarities. They find that the topic model networks are highly similar to all of the other networks (0.93–0.99, cosine similarity). However, this finding is difficult to interpret. Most of the other networks are pairwise quite dissimilar to each other (for example, the cocitation networks' similarities for most other network types range from 0.01 to 0.65); and Yan and Ding (2012) dichotomize the topic network by connecting nodes if and only if their topic similarities (measured using cosine similarity) are greater than 0.8[2].

In quantitative science studies, topic models are often used diachronically, to examine the ways research foci have changed over time (Han, 2020; Malaterre, Lareau et al. 2020; Rehs,

---

[2] The term *similarity* is also used in information retrieval (IR), where it is used for scoring schemes that find the documents in the corpus that are most relevant or related to a given document (or search query). Boyack, Newman et al. (2011) compare several different clustering methods based on different similarity scores, in this sense, including one derived from topic modeling. In their application—several years of papers indexed by PubMed—the coherence of the topic model-based clusters is generally comparable to the two best methods, BM25 (a widely used IR similarity score) and PRMA (the IR similarity score used by PubMed). The only cases where BM25 and PRMA clusters were substantially more coherent than those of the topic model-based method were a few very large clusters. Note that IR similarity scores are generally not symmetric: For a score *s* and documents d_1, d_2, it can be the case that $s(d_1, d_2) \neq s(d_2, d_1)$. For this reason these similarity scores do not correspond to metrics or distance functions. All of the similarity measures discussed in the main text do correspond to metrics.

2020). Nichols (2014) used a (previously fitted) topic model synchronically, to examine traces of interdisciplinary research in projects funded by the U.S. National Science Foundation. In all of these examples, a single topic model was used for analysis, and topics were interpreted as disciplines, fields, or research areas, depending on the scope of the corpus. For example, Nichols (2014) assigned almost all topics from a previously fitted 1000-topic model to NSF directorates (which roughly correspond to high-level scientific fields, such as biology vs. computer science vs. social and behavioral science), then calculated interdisciplinarity scores based on whether these topics indicated interdisciplinarity within or between directorates. Malaterre et al. (2020) used a corpus comprising articles from eight journals within a single field, philosophy of science, and interpreted the 25-topic model in terms of research areas.

As the examples in the last paragraph suggest, the topics in topic models can be interpreted as disciplines (or subdisciplinary units of analysis such as subfields). Because the topics are clusters of terms (or, strictly speaking, probability distributions over terms), "disciplinary topics" correspond immediately to research areas, methods, and other features that readily appear in the language used within disciplines, rather than to features of disciplines as social organizations, such as power hierarchies or patterns of funding or training. However, it is highly plausible that these two kinds of features are related, as the social organization enforces norms that regulate research areas, methods, technical terminology, and other linguistic features. For example, Morgan, Economou et al. (2018) combined academic hiring data with the occurrence of keywords in article titles to examine the effect of prestige on the spread of research areas across a discipline.

As they are usually applied, topic models are known to have two significant researcher degrees of freedom. First, the number of topics, k, is a free parameter, and there is no consensus on the appropriate way to choose a value for k. As far as I am aware, the current state of the art in topic model development requires fitting multiple models across a range of values for k, then calculating various goodness-of-fit statistics on each model, such as semantic coherence, exclusivity (Roberts, Stewart et al., 2014, pp. 6–7), the log likelihood of a holdout subset of terms (i.e., not used to fit the model), the standard deviation of the residuals (which should converge to 1 at the "true" value of k; Taddy, 2012), and the number of iterations required by the algorithm to fit the model. However, these statistics all have known limitations. Semantic coherence favors a small number of "large" topics; exclusivity favors a large number of "small" topics. The log likelihood and residual methods both assume that there is a "true" correct number of topics; but we would expect different numbers of topics at different levels of conceptual granularity (for example, at a coarse level of granularity "biology" might be a single topic, while at a finer level "ecology" and "molcular biology" might be distinct topics). This last conceptual mismatch is directly related to what Carole Lee has called the "reference class problem for credit valuation in science" (Lee, 2020). So in general these goodness-of-fit statistics will not agree on a single "best" model to use in further analysis.

Topic interpretation introduces a second major research degree of freedom. Topics are usually interpreted by extracting short term lists—usually the five or 10 highest-probability terms from each topic—and manually assigning a label to each topic based on these term lists. Sometimes interpretation also involves a review of the documents with the highest probabilities for each topic. Topic labels are almost always assigned by the authors of the topic model study—who may or may not have subject matter expertise in the areas covered by the corpus—and reports often provide little or no detail about how labels were validated (for example, to what extent there were substantive disagreements between labelers about how to interpret a given topic and how such disagreements were reconciled).

To mitigate these concerns about researcher degrees of freedom, the current project does not select a single model for analysis, and does not lean on topic interpretation. Instead, all

fitted topic models are analyzed purely quantitatively to compare and situate authors relative to each other in discursive space. That is, treating authors as "documents" in the topic model, for authors i and j we can locate them relative to each other in discursive space by comparing the distributions $\gamma_{.,i}$ and $\gamma_{.,j}$. Then, insofar as the UC Davis ORUs have promoted interdisciplinary research, I expect this space to have the features predicted by the three hypotheses above. Here topic models function as a technique of dimensionality reduction, moving from the high-dimensional space of all terms to the relatively low-dimensional space of topics. Comparisons of these quantitative analyses across all fitted topic models allow us to assess the robustness of results.

2. DATA AND METHODS

In this study, my unit of analysis is individual researchers or authors, as individuated by the Scopus Author Identifier system[3], except for a few analytical moments in which I compare individual researchers to organizational entities (departments or ORUs). My unit of observation is publications—paradigmatically, journal articles—retrieved from Scopus and aggregated as either author-level totals or concatenated blocks of text (specifically, the abstracts of an author's published work, treated as a single block of text).

Unless otherwise noted, all data used in this project was retrieved from Scopus, using either the web interface or application programming interface (API), between November 2018 and June 2019. Due to intellectual property restrictions, the data cannot be made publicly available. Some downstream analysis files may be provided upon request.

All data collection and analysis was conducted in R (R Core Team, 2018). The RCurl package was used for API access (CRAN Team & Temple Lang, 2020); the `spaCy` Python library was used for tokenizing, lemmatizing, and tagging abstract texts with parts of speech (spaCy, 2018); the `spacyr` package was used as an interface between R and `spaCy`; the `stm` package was used to fit topic models (Roberts et al., 2014); and numerous tools in the `tidyverse` suite were used to wrangle data (Wickham & RStudio, 2017). Because work on this project was interrupted for a period of approximately 18 months, software versions were not consistent across the lifespan of this project.

All code used in data collection and analysis is available at https://github.com/dhicks/orus.

2.1. Author Identification

In November 2018, the UC Davis Office of Research provided me with then-current rosters for each ORU. These rosters included faculty (tenured/tenure-track faculty), "other academics" (primarily staff scientists), and "collaborators" (other researchers, who may or may not be affiliated with UC Davis and who generally did not receive funding from the ORU). I extracted the names and ORU affiliation for all 134 affiliated faculty. In the remainder of this paper, I refer to these ORU-affiliated faculty interchangeably as "ORU faculty" and "ORU researchers."

In January 2019, I conducted searches using the Scopus web interface for all papers published with UC Davis affiliations in 2016, 2017, and 2018. These searches returned 7,460, 7,771, and 8,066 results, respectively, totaling 23,297 publications. The metadata retrieved for these papers included names, affiliation strings, and Scopus author IDs for each author.

---

[3] See https://service.elsevier.com/app/answers/detail/a_id/11212/supporthub/scopus/ for current details on this system.

Using a combination of automated and manual matching, I identified author IDs for ORU-affiliated faculty, matching 125 out of 134 affiliated faculty. I next searched the affiliation strings (from the publication metadata) for "Department of" to identify departmental affiliations for these ORU faculty.

To identify a comparison set of researchers, I first identified all authors in the 2016–2018 Scopus results with the same departmental affiliations; that is, an author was included in this stage if they shared at least one departmental affiliation string with an ORU researcher. This resulted in 5,645 "candidate" authors for the comparison set. However, many of these candidates were likely graduate students and postdoctoral researchers. Because ORU researchers are generally tenured faculty, including students and postdoctoral researchers would confound the analysis of differences between ORU and non-ORU researchers. For example, students and postdoctoral researchers have much stronger incentives than tenured faculty to engage in narrowly disciplinary research.

I therefore used the Scopus API (application programming interface) to retrieve author-level metadata for both the ORU faculty and the candidate comparison researchers. Specifically, I examined the number of publications and the year of first publication. After exploratory data analysis, I filtered both ORU faculty and comparison researchers, including them for further analysis only if they met two conditions: 15 or more total publications (as a proxy for student/postdoc status), and first publication after 1970. The first condition removed 60% of the candidate comparison authors, and the second was used to exclude a small number of researchers with very early first publication years (e.g., 1955) that were plausibly due to data errors.

Note that, in the analysis below, departmental affiliations for all authors are based on the 2016–2018 Scopus results, not entire publication careers.

After applying these filters, 2,298 researchers had been selected for analysis, including 116 ORU-affiliated researchers and 2,182 "codepartmental" comparison researchers. Figure 2 shows the number of researchers in the analysis data set for each ORU and the comparison set[4], and Figures 3 and S1 show the structure of organizational relationships in the data. (Note that, because some researchers are affiliated with multiple ORUs, the ORU counts are greater than 116.)

A substantial body of work in social studies of science finds consistent evidence that female academics have lower publication rates and typically receive fewer citations per publication than male academics (Beaudry & Larivière, 2016; Cameron, White, & Gray, 2016; Chauvin, Mulsant et al., 2019; Ghiasi, Larivière, & Sugimoto, 2015; Larivière, Ni et al., 2013; Symonds, Gemmell et al., 2006; van Arensbergen, van der Weijden, & van den Besselaar, 2012). (I was unable to find any literature that reported findings for nonbinary, genderqueer, or transgender identities. Chauvin et al. (2019) note that they "planned to include faculty identifying as nonbinary or genderqueer in a separate group," but "were unable to identify any such faculty from publicly available data sources.") To control for these gender effects, the online tool `genderize.io` was used to attribute gender to all authors based on their first or given name. This tool resulted in a trinary gender attribution: woman, man, and "missing" when the tool was not sufficiently confident in either primary gender attribution. While extremely limited, this tool allows us to account for a known confounder given the

---

[4] At the time of data collection, the AQRC roser also included nine "other academics," all of whom had non-faculty titles such as "Researcher," "Research Professor," or "Operations Manager." For completeness, AQRC is included in all analyses, except those comparing researchers within a given ORU, such as distance to members of the same ORU.
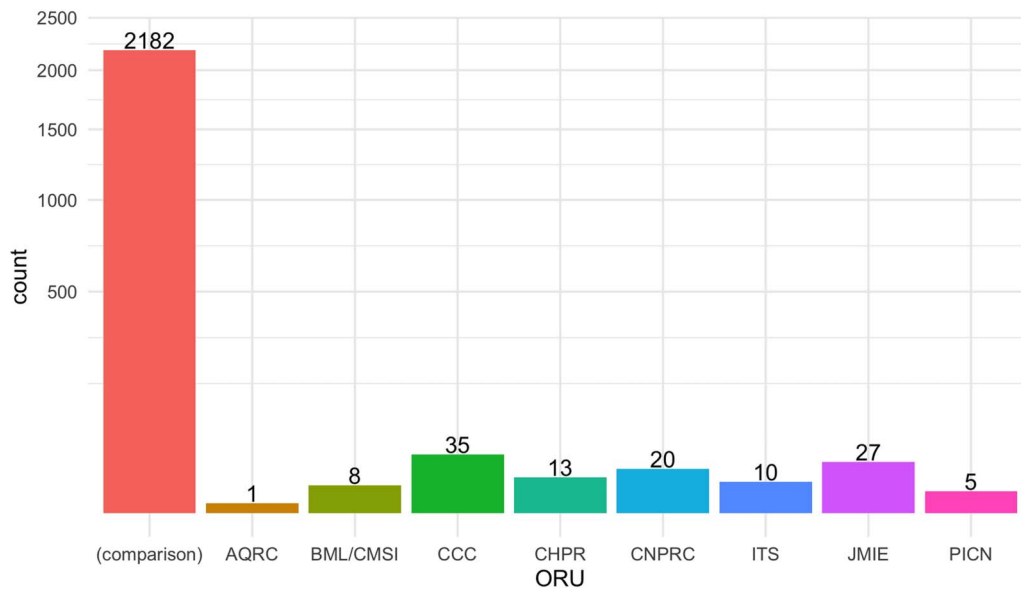
Figure 2. Number of researchers in the analysis data set for each ORU and the comparison set. Note that the *y*-axis uses a square-root scale.

limited time available for this project. Figure 4 shows the distribution of attributed gender for ORU- and non-ORU-affiliated researchers and each ORU separately. All together, ORUs appear to be slightly more male-dominated than the comparison group. However, there is substantial variation across ORUs; the single AQRC-affiliated faculty member is attributed as a man, more than half of PICN-affiliated faculty are attributed as women, and most ORUs have
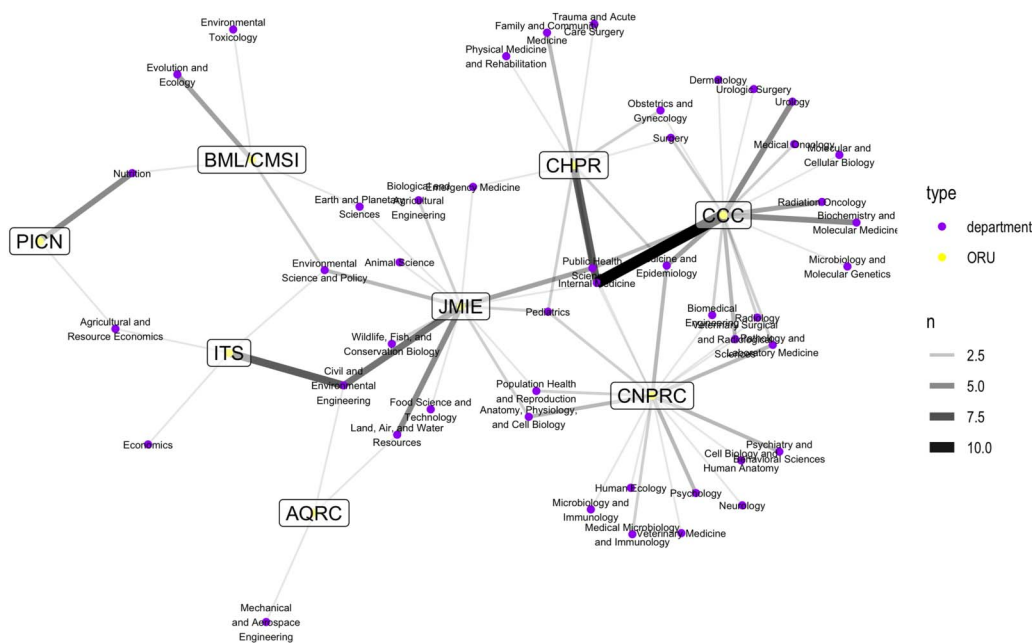
Figure 3. Organizational relationships in the data. Nodes are either ORUs or departments at UC Davis. Edges connect an ORU to a department if the two organizations share a common faculty member. Edge width and color indicate the number of such individuals.
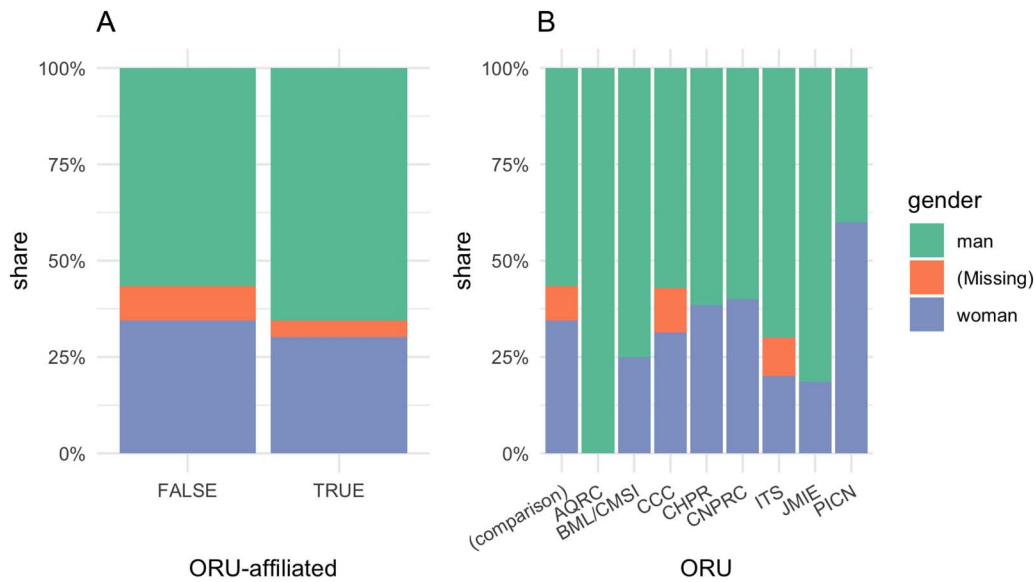
Figure 4. Distribution of attributed gender for ORUs and comparison researchers. A: All ORU-affiliated researchers grouped together. B: Gender distribution within each ORU. Gender attributions were made using an automated online tool based only on first or given name.

20–40% faculty attributed as women. In the regression analyses below, men (attributed gender) are used as the reference level for estimated gender effects.

2.2. Productivity Impacts

To investigate the productivity impacts of ORUs, I used author-level metadata from the Scopus API. Specifically, all-career publication counts and (incoming) citation counts are both reported in the Scopus author retrieval API, and so these data were retrieved prior to the filtering step above. Coauthor counts (total number of unique coauthors) were calculated from the article-level metadata retrieved for the text analysis steps discussed below. Because coauthor counts, publication counts, and citation counts all varied over multiple orders of magnitude, I used the log (base 10) value for these variables in all analyses.

I fit regression models for each of these three dependent variables, using ORU affiliation as the primary independent variable of interest and incorporating controls for gender, first year of publication (centered at 1997, which is the rounded mean first year in the analysis data set), and dummy variables for departmental affiliation.

Because of the log transformation of the dependent variables, the regression model coefficients can be exponentiated and interpreted as multiplicative associations. For example, a coefficient of 0.5 can be interpreted as an association with a $10^{0.5} \approx 3.16$-fold or $3.16 \times 100\% - 100\% = 216\%$ increase in the dependent variable.

To account for relationships between the three dependent variables, I use the simplified DAG shown in Figure 5. According to this model, the number of coauthors influences the number of publications, which influences the number of citations. The number of publications thus mediates between coauthors and citations, and coauthors mediates between the independent variables and publications; I also allow that coauthors might directly influence citations.
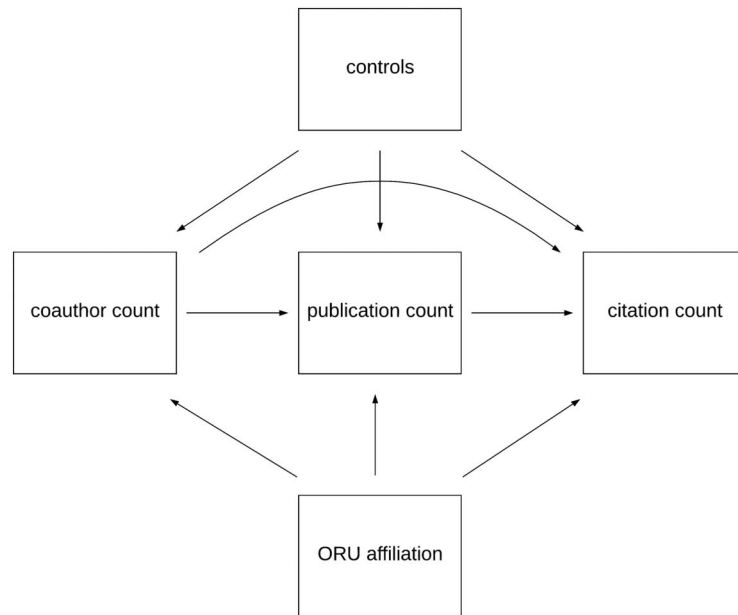
Figure 5. A simplified directed acyclic graph (DAG) used to account for relationships between the productivity dependent variables.

Both ORU affiliation and all of the included control variables (first year of publication, gender, department affiliation) might directly influence all three dependent variables.

2.3. Discursive Impacts

I use topic models and related text analysis methods to examine the discursive impacts of ORUs.

2.3.1. Topic modeling

Specifically, I first used the Scopus API to retrieve paper-level metadata for all authors in the analysis data set. I aimed to collect complete author histories—metadata for every publication each author had written in their entire career. Metadata were retrieved for 128,778 distinct papers in June 2021, of which 114,461 had abstract text[5].

Abstract texts were aggregated within individual authors, treating each individual author as a single "document." For example, suppose researcher A was an author on documents 1 and 2, and researcher B was an author on documents 2 and 3. Researcher A, as a single "document," would be represented for text analysis by adding together the term counts of abstracts 1 and 2; while researcher B would be represented by adding together the term counts of abstracts 2 and 3.

Vocabulary selection began by using part-of-speech tagging to identify noun phrases in each paper abstract. Nouns are more likely than other parts of speech to carry substantive information about research topics and methods. Noun phrases, such as "random effects model" or "unintended pregnancy," are more specific and informative than nouns alone. 414,423 distinct

---

5 In review, changes to the inclusion criteria used to select comparators required me to reretrieve these meta-data. Only papers published through 2019 were included at this stage. One Scopus author identifier no longer existed when the metadata were reretrieved. This identifier corresponded to one comparison author, who was excluded from analysis.

noun phrases were extracted from the corpus. I then counted occurrences for each noun phrase for each author. (In the remainder of this paper, I generally use *terms* to refer to noun phrases.) I used these author-term counts to calculate an entropy-based statistic for each term, keeping the top 11,490 terms to achieve a 5:1 ratio between authors ("documents") and terms. Note that stopwords were not explicitly excluded, though typical lists of English stopwords do not include many nouns.

To score the "importance" of each term, consider the task of identifying an author (or, more generally, a "document"), selected uniformly at random from the analysis data set. For N authors, the unconditional probability of drawing any given author is $p_N = \frac{1}{N}$; this distribution has entropy

$$H(p_N) = \sum_N - p_N \log_2 p_N$$
$$= -N\frac{1}{N} \log_2 \frac{1}{N}$$
$$= \log_2 N$$

Now suppose we are given a term w ("word") drawn from the token distribution of the unknown author. Because the uniform distribution has maximal entropy, the conditional author distribution given the term, $p_w = p(author_j | term_w)$, has a lower entropy $H_w = H(p_w) \leq \log_2 N$. Let $\Delta H_w = \log_2 N - H_w$. ΔH_w measures the information about the identity of the author gained when we are given the term w. (This formula derives from the Kullback-Leibler divergence from the uniform distribution p_N to p_w.) A high-information term dramatically narrow downs the range of possible authors. That is, terms have higher information insofar as they are specific to a smaller group of authors.

However, typically, the most high-information terms will be unique to a single author, such as typos or idiosyncratic terms. To account for this, I also calculate the order of magnitude of the occurrence of a term across the entire corpus, $\log_{10} n_w$. We then take the product $\log_{10} n_w \Delta H_w$, which I represent in the code as `ndH`, and select the top terms according to this $\log_{10} n_w \Delta H_w$ statistic. Table S1 shows the top 50 terms selected for the analysis vocabulary. As the term list suggests, this statistic is effective at identifying terms that are clearly distinctive (in this case, to different disciplines and research fields), meaningful, and frequent. The term list also illustrates how $\log_{10} n_w \Delta H_w$ balances information gain with word occurrence. Some terms, such as "cuttlefish," have extremely high information gain (very low $H = 0.17$) but are common enough (occurring 118 times across the corpus) that they are not typos or idiosyncratic to a single author. Other terms, such as *schizophrenia,* have more modest information gain ($H = 4.59$) but are extremely common (occurring 2,374 times).

As discussed above, topic models require setting the number of topics k before fitting the model, and there is no consensus on the appropriate way to choose a value for k. Exploratory analysis of the author-term distributions using principal components found that 50% of the variance could be covered by 24 principal components, 80% required 167 principal components, and 90% required more than 300 principal components. I also speculated that small-k topic models might capture coarse disciplinary distinctions, but would also be less stable. Given these considerations, I fit models with 5, 10, 15, 20, 25, and then 50, 75, 100, 125, and 150 topics. I calculated five goodness-of-fit statistics for each of these models: semantic coherence, exclusivity, the log likelihood of a holdout subset of terms (i.e., not used to fit the model), the standard deviation of the residuals, and the number of iterations required by the algorithm to fit the model. As expected, these statistics did not indicate a uniformly "best"

topic model; though $k = 50$ minimized both the number of iterations and the residuals, and had a greater coherence and approximately the same exclusivity as the larger models.

Rather than selecting a single "best" topic model, in the analysis below I either (a) conduct and report analyses using all of the topic models, highlighting $k = 50$, or (b) conduct analyses for $k = 5, 25, 50, 100$, reporting all four equally. Approach (b) is generally used when the analysis involved a complex visualization component, to keep the number of plots manageable.

2.3.2. Analyses

I focused my analysis on the topic distribution $\gamma_{t,i} = p(topic_t|author_i)$ for each topic model. Recall the three hypotheses for discursive impacts, introduced in Section 1.2.2. For H1, I calculated "discursive breadth" for author i as the entropy of the topic distribution $H_i = H(\gamma_{.,i}) = \Sigma_t - \gamma_{t,i} \log_2 \gamma_{t,i}$. In information theory, entropy is understood as a measure of the "width" or "breadth" of a distribution (McElreath, 2016, p. 267). Rafols and Meyer (2010) examine the use of diversity concepts in studies of interdisciplinarity, and analyze them into the "attributes" or "categories" of variety, balance, and disparity/similarity (Rafols & Meyer, 2010, p. 266). They note that entropy combines variety and balance (Rafols & Meyer, 2010, p. 268). Rosen-Zvi, Griffiths et al. (2004) use the entropy of author-topic distributions (from an author-document-topic model) "to assess the extent to which authors tend to address a single topic in their work, or cover multiple topics" (Rosen-Zvi et al., 2004, p. 8). At the journal level, Leydesdorff and Rafols (2011) consider the entropy of citation distributions as a measure of interdisciplinarity. In a factor analysis, they find it is related to a Rao–Stirling diversity measure, and conclude that "Shannon entropy qualifies as a vector-based measure of interdisciplinarity" (Leydesdorff & Rafols, 2011, p. 96).

H2 and H3 both appeal to a measure of "discursive distance." I operationalize this distance between any pair of topic distributions γ_1, γ_2 as Hellinger distance ("Hellinger Distance," 2021):

$$h(\gamma_1, \gamma_2) = \frac{1}{\sqrt{2}} \sqrt{\sum_t \left(\sqrt{\gamma_{t,1}} - \sqrt{\gamma_{t,2}} \right)^2}$$

$$= \sqrt{1 - \sum_t \sqrt{\gamma_{t,1} \cdot \gamma_{t,2}}} \qquad\qquad = \sqrt{1 - \left\langle \sqrt{\gamma}_1, \sqrt{\gamma}_2 \right\rangle}$$

Hellinger distances range from 0 to 1, where 0 indicates that two distributions are the same and 1 indicates that the two distributions have completely different support. Hellinger distance can be understood as a scaled version of the Euclidean distance between the square root vectors $\sqrt{\gamma}_1$, $\sqrt{\gamma}_2$; or, because the square root vectors are all unit length, as a distance measure corresponding to the cosine similarity $\frac{\langle x,y \rangle}{|x||y|}$ between the square root vectors. Cosine similarity is widely used in bibliometrics (Mingers & Leydesdorff, 2015).

H2 requires constructing department-level topic distributions. The stm package provides functions that, given a fitted topic model and an observed term distribution for a "document," estimate a conditional topic distribution γ for that "document" (somewhat like using a fitted regression model to predict outcomes for new observations). One simple way to construct a department-level "document" would be to aggregate the work (take the sum of term counts) of all of the authors associated with that department. However, for the purposes of investigating H2, this construction would lead to various problems. First, papers by multiple authors in the department would contribute to the department-level distribution multiple times. Second, insofar as ORU faculty are distant from the other members of their department, their contributions to the department distribution will act as outliers, and the resulting distance measures will

be biased towards the ORU faculty, leading to underestimates of the effect for H2. On the other hand, if all and only non-ORU faculty members contribute to the department-level distribution, then their work would be counted twice: First they would be used to construct the department-level distribution, and then second we would calculate their distances from this distribution. In this case, the distance measures will be biased towards the non-ORU faculty.

To avoid these problems, I constructed department-level distributions as follows. I first borrowed an approach from machine learning (James, Witten et al., 2013, p. 176ff), and randomly separated non-ORU authors into two discrete subsets. The first subset—referred to as the "training" set in machine learning—was used to construct the department-level topic distributions. The second subset—the "testing" subset—as used to make the distance comparisons, using Hellinger distance. 50% of non-ORU authors by departmental affiliation were allocated to the training set, selected uniformly at random, and the remaining non-ORU authors were assigned to the testing set. (This means that a non-ORU author affiliated with multiple departments had the same role—testing or training—across all of their affiliations.) Then, for each department, I aggregated all of the papers that (a) had at least one training set author and (b) did not have an ORU-affiliated author.

After allocating authors to these subsets and constructing department-level topic distributions, I calculated "departmental distance" using Hellinger distance for all ORU-affiliated faculty and all comparison authors in the testing subset. I used these departmental distance values as dependent variables in a series of regression models, one for each value of k, including first publication year, gender, log number of documents and coauthors, and department dummies as controls.

For H3, I made both "individual" and "organizational" comparisons for each ORU-affiliated faculty member. At the individual level, I calculated the Hellinger distance between the ORU-affiliated researcher and other individuals, (a) in the same ORU and (b) in the same departments, and then took the minimum for both (a) and (b). At the organizational level, I calculated the Hellinger distance between the ORU-affiliated researcher and (a) an ORU-level topic distribution, constructed by aggregating the papers authored by affiliates, and (b) the department topic distribution described above. (Note that this means ORU-affiliated authors contribute to the ORU topic distribution, but not the department distribution. So this construction might tend to bias distance estimates towards ORUs and away from departments.) At both levels, (a) gives us a measure of distance within the researcher's ORU and (b) gives us a measure within the researcher's departments. If the ORU distance is less than the departmental distance, this indicates that the ORU faculty member is closer to their ORU than to their home department, consistent with H3. Using both "individual" and "organizational"-level comparisons accounts for the possibility that an ORU-affiliated researcher may be quite close (in discursive space) to one or a few non-ORU-affiliated departmental colleagues but still relatively far from the "core" or "mainstream" of their department.

3. RESULTS

3.1. Productivity Impacts

Regression analyses indicate that ORU affiliation is associated with a substantial increase in the order-of-magnitude number of coauthors, 1.5–2.1-fold (1.8-fold)[6]. See Figure 6. ORU

---

[6] Statistical estimates are reported as 95% confidence intervals followed by the maximum likelihood point estimate in parentheses. No statistical hypothesis testing was done in this paper, so no *p*-values are calculated or reported. Confidence intervals should be interpreted as a range of values that are highly compatible with the observed data, given the modeling assumptions (Amrhein, Trafimow, & Greenland, 2018). A confidence interval that contains zero can still be evidence of a substantial (nonzero, say, positive) relationship insofar as the bulk of the interval is greater than zero.
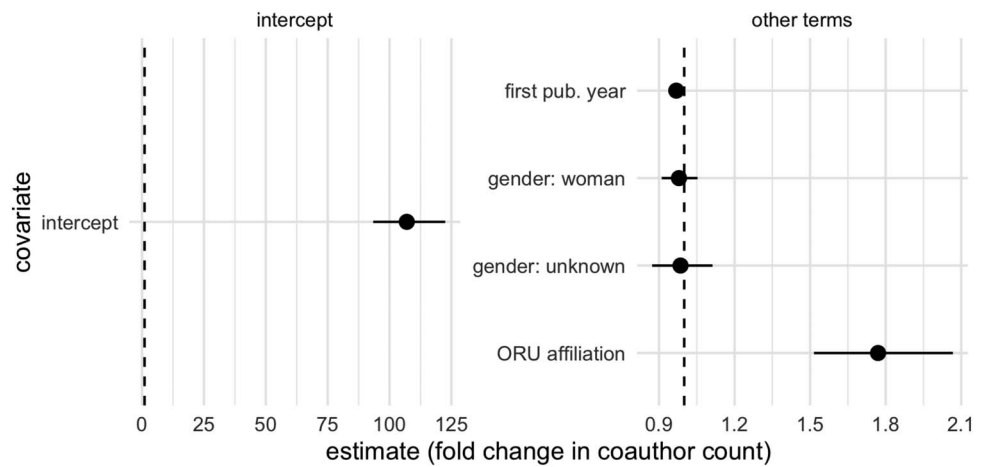
Figure 6. Regression estimates for number of coauthors.

affiliation had a much weaker direct association with the number of publications, 1.0–1.2-fold (1.1-fold), while an order-of-magnitude increase in number of coauthors had a much stronger association, 2.9–3.3-fold (3.1-fold). See Figure 7.

The estimates for citations are similar. Order-of-magnitude increases in number of publications and number of coauthors are both associated with substantial increases in the number of citations a researcher has received to date: 4.0–5.3-fold (4.6-fold) for publications, 2.4–3.0-fold (2.7-fold) for coauthors. When controlling for these midstream dependent variables, the association between ORU affiliation and citations received is small or perhaps even negative, 0.9–1.2-fold change (1.0-fold). See Figure 8.

All together, in causal terms, these regression results suggest that ORU affiliation has a substantial direct effect only on the number of coauthors. This increase in coauthors in turn leads to increased publications and increased citations; but ORU affiliation has a much smaller direct effect on these two downstream productivity measures. On this interpretation of the findings, ORU affiliation makes faculty more productive primarily by connecting them with collaborators.
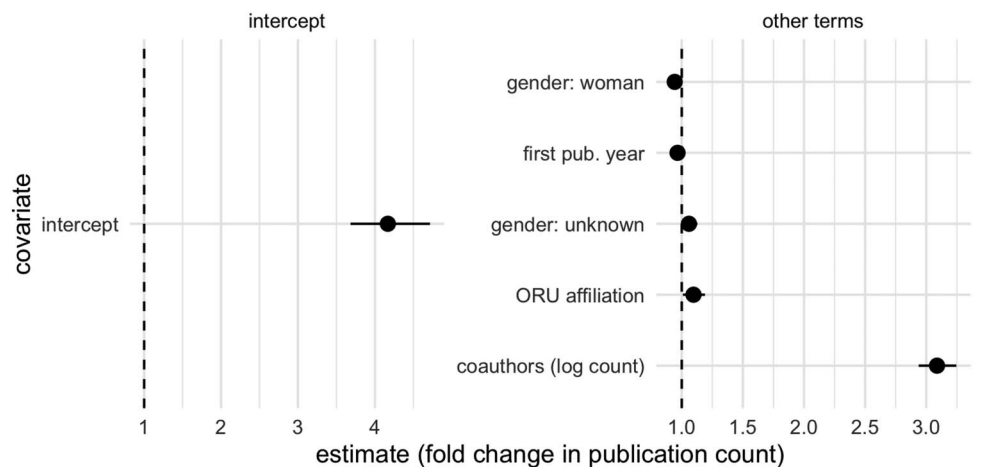

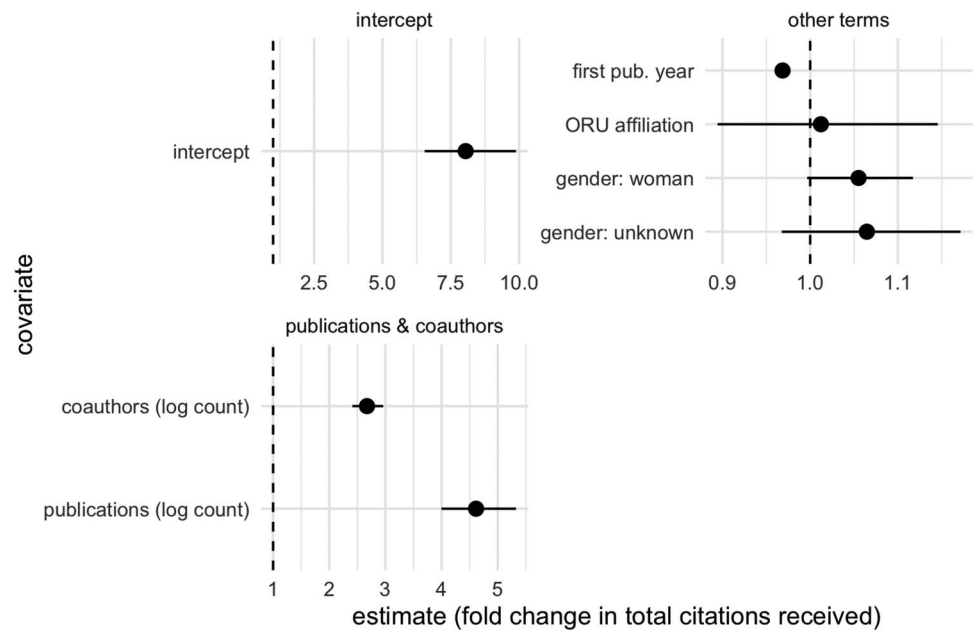**Figure 7.** Regression estimates for number of publications.

Figure 8. Regression estimates for number of citations.

However, the evidence for this causal interpretation is limited, because we do not have the data to compare a researchers's number of coauthors before and after they join the ORU[7]. The available data are consistent with a pattern where some unobserved variable is a common cause of both ORU affiliation and coauthor count. For example, highly gregarious and extroverted faculty members might tend to have more coauthors and also be more likely to be invited to join an ORU. Or, an interdisciplinary group of researchers might have already been working together, then formed an ORU to provide institutional support for their collaboration. For example, the AQRC "About Us" page states that "The Air Quality Research Center was established in the summer of 2005, although our faculty, staff and student affiliates had been working together for many years prior" (https://airquality.ucdavis.edu/about)[8].

3.2. Discursive Space

In this section, I report the results of an exploratory analysis of the topic model results. I focus primarily on the department- and ORU-level distributions, and evaluating the suitability of the

---

[7] Some readers might object that, without this data, this study cannot make any causal claims. This objection involves a common mistake about causal inference, confusing a *sufficient* condition for *strong* evidence for causal claims (satisfying the assumptions of causal inference theory) with a *necessary* condition for (possibly *weaker*) evidence for causal claims. Any correlation between two variables provides a reason to believe—that is, evidence—that there is a causal relationship between them (compare Reichenbach's common cause principle; Hitchcock & Rédei, 2021). The reason we conduct observational studies is that correlations provide evidence of causation. But the inference from correlation to causation is not very reliable, and thus mere correlation provides only rather weak, highly defeasible evidence for causation. The point of causal inference theory is to give sufficient conditions for more reliable inferences and thus better evidence. Unfortunately, the data needed to satisfy these conditions for this study—the dates when researchers first became affiliated with ORUs—do not exist. Still, even weak, defeasible evidence is evidence, and so I sometimes put forward—defeasible, qualified—causal claims in this paper. Readers may, at their discretion, reject the causal claims even if they accept the correlational results.

[8] Thanks to an anonymous reviewer for pointing out this example confounder.

topic model results for analyzing interdisciplinarity using the "discursive space" conceptual framework.

A list of the five highest-probability terms from each topic in the $k = 50$ model is provided in Table S2. As discussed above, I do not label these topics, and my analysis does not depend on the content of the term lists.

Figure 9 visualizes "discursive space" based on the pairwise Hellinger distance for the topic distributions of authors, departments, and ORUs. In the visualization, these pairwise distances are represented in two-dimensional space using the t-SNE algorithm (van der Maaten & Hinton, 2008). This algorithm uses an information-theoretic approach to represent high-dimensional relationships (here, the Hellinger distances) in two dimensions. The algorithm is widely used in fields such as computational biology. But it is designed to emphasize local topology rather than global geometry. This means that a t-SNE visualization indicates nearness, but that distances in the visualization do not necessarily correspond to distances in the original
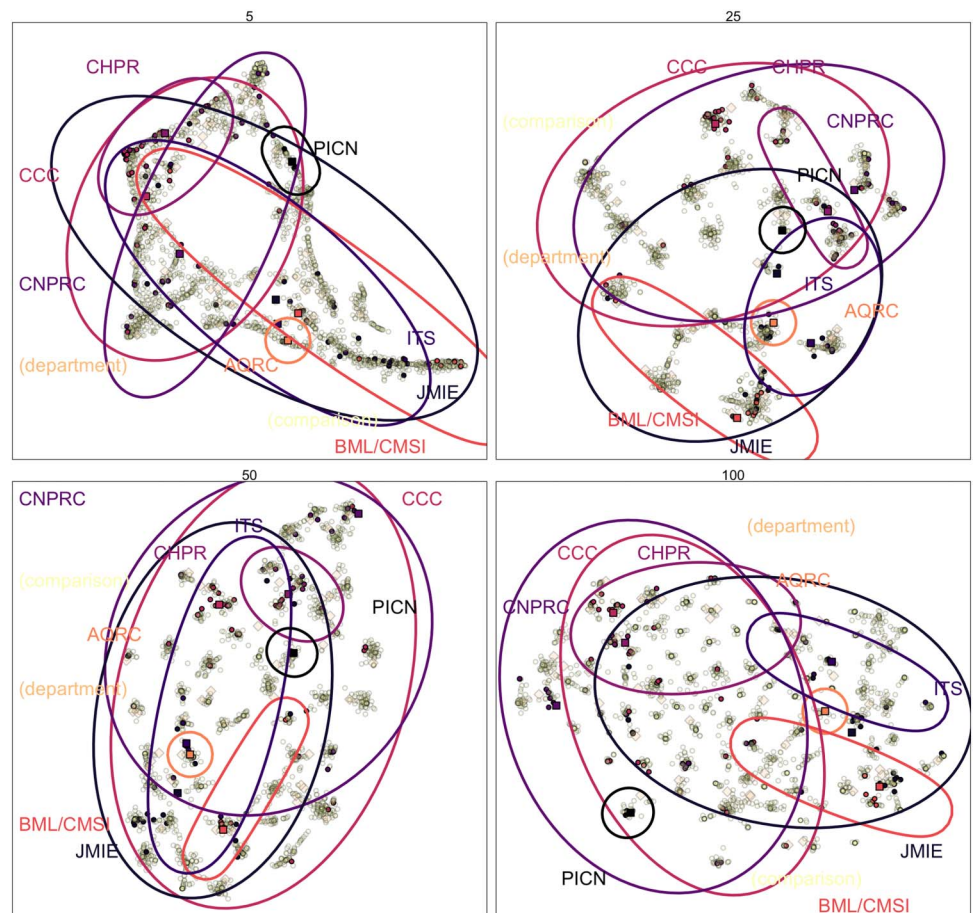
Figure 9. Visualization of "discursive space." Panels correspond to different values of k (number of topics). Circles are authors; non-ORU-affiliated authors are indicated with translucent yellow points. Squares and diamonds represent ORU- and department-level topic distributions, respectively. Point positions are calculated using t-SNE on the pairwise Hellinger distance between author-topic distributions. Ellipses indicate bounds on the researchers affiliated with each ORU (based on the convex hull).

high-dimensional space. In particular, t-SNE tends to organize points into visual clusters. Two clusters might be visually far apart but relatively close in the original high-dimensional space.

The t-SNE visualizations of "discursive space" suggest complex archipelagoes of researchers. Some ORUs, such as PICN and ITS, have all of their affiliates in one or a few clusters (recall that AQRC has a single faculty affiliate in these data). For others, such as BML/CMSI and CHPR, most of their affiliates are clustered near the ORU-level topic distribution (indicated by the square), with a few further-flung affiliates. Again, note that the t-SNE algorithm can place two clusters visually far apart even when they are close in the original high-dimensional space.

Figure 10 breaks out the visualization of "discursive space" by department for large departments (50 or more authors in the data set). While a few departments are tightly clustered into a single island, most are somewhat scattered, with authors distributed across the visualization.

Figures S2 and S3 show the department- and ORU-level topic distributions, and Figure 11 shows the entropy of these distributions. These figures indicate that, except for the highest values of k, most departments and ORUs have entropies of 1–3 bits, roughly corresponding
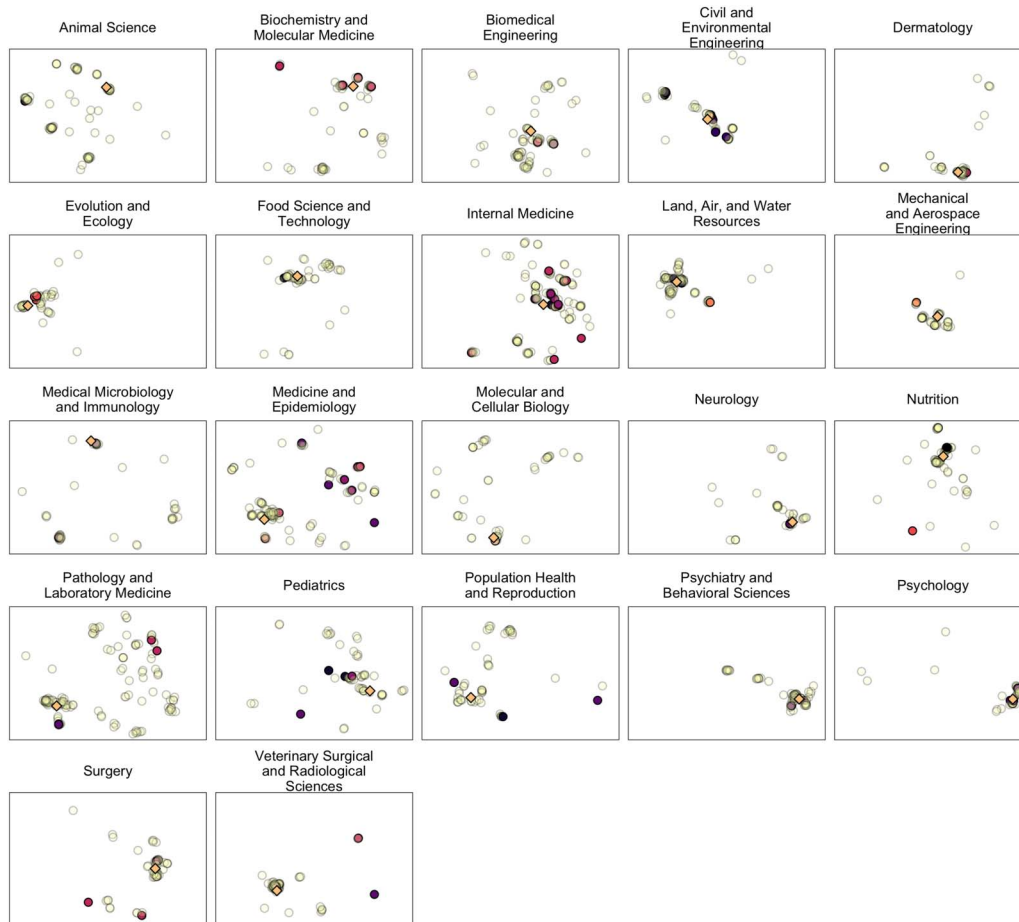
Figure 10. Visualization of "discursive space" for departments with 50 or more authors in the data set, k = 50 topic model. Circles are authors; non-ORU-affiliated authors are indicated with translucent yellow points. Diamonds represent the department-level topic distribution. Positions within each panel are the same as those calculated in Figure 9.
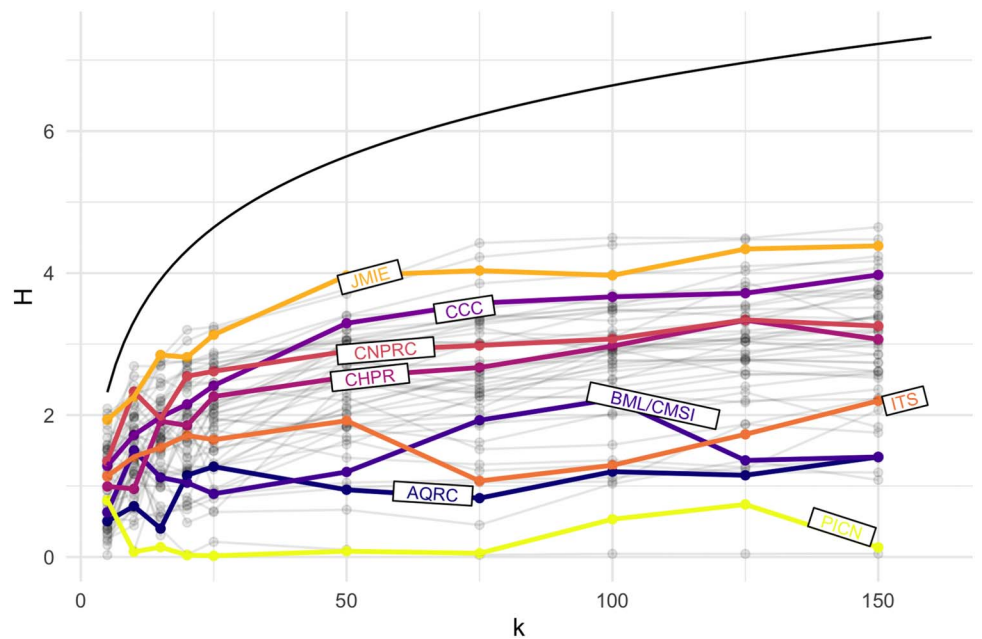
Figure 11. Entropy of the ORU and department-level topic distributions. Thin grey lines correspond to departments; colored lines correspond to ORUs. Each point corresponds to the entropy of the distribution shown in Figure S2 or S3. The solid black curve corresponds to the maximum possible entropy at that value of k (a uniform distribution across all topics).

to 2–8 topics. For example, BML/CMSI has an entropy of about 2 at $k = 100$, and four colored bars are visible in the corresponding panel in Figure S3.

Figure 12 visualizes the similarity network among the departments and ORUs based on the Hellinger distance between their topic distributions for the $k = 50$ topic model; Figure S4 shows networks across four values of k. (All pairs of departments and ORUs are used in this network analysis; that is, no edges with "low" values are trimmed to zero. This eliminates a common researcher degree of freedom in this kind of network analysis.) For $k > 5$, edge weights/similarity scores are uniformly low, indicating that these units are generally far from each other in "discursive space." Connections between ORUs and related departments are among the strongest in each network, although even these are still only moderate in absolute terms (for example, < 0.50 for $k = 100$).

The similarity network visualizations also include clusters produced using the Louvain community detection algorithm (Blondel, Guillaume et al., 2008; while this algorithm is widely used in network analysis, Hicks [2016] reviews fundamental problems with Louvain and other modularity-based community detection algorithms), using Hellinger similarity scores as edge weights. The cluster results suggest that the topic models are effectively encoding higher-level relationships between disciplines, with higher values of k enabling the detection of more fine-grained relationships. For example, Human Ecology, Psychology, Psychology and Behavioral Sciences, and CNPR are consistently clustered together. Similarly, another consistent cluster is Evolution and Ecology; Wildlife, Fish, and Conservation Biology; Environmental Science and Policy; and BML/CMSI.

These figures enable us to align the topic models and the conceptual framework of "discursive space" with the institutional account of disciplines, namely, departments as the primary
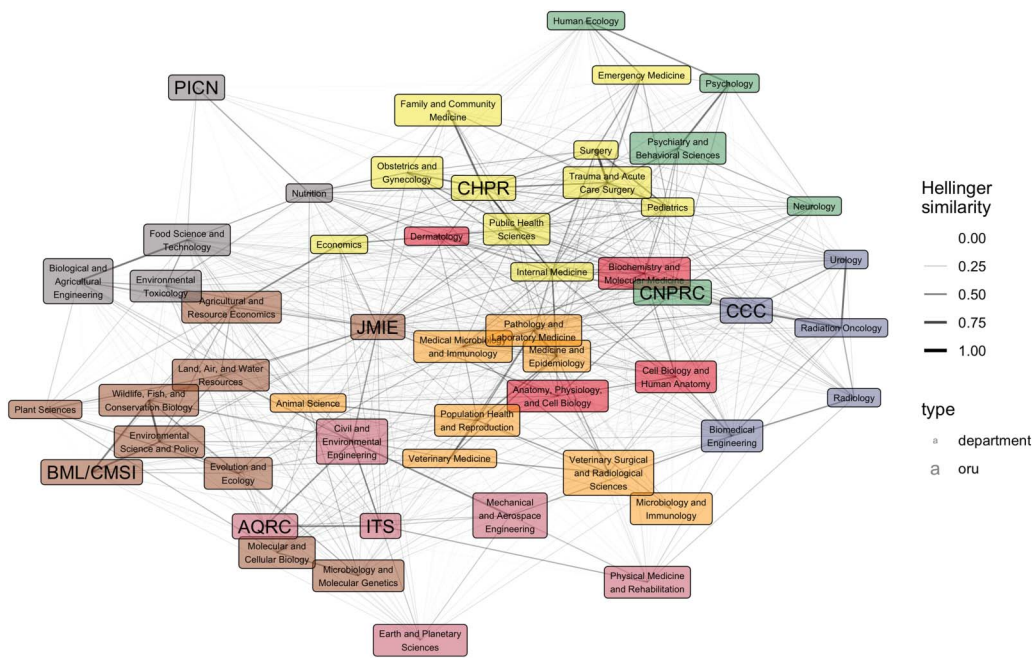
Figure 12. Similarity network for departments and ORUs, based on Hellinger distance between department- and ORU-level topic distributions (Figures S2 and S3) for the $k = 50$ topic model. Nodes (labels) correspond to departments/ORUs, with ORUs indicated by slightly larger labels. Edge widths and shading correspond to similarity values, and edges for all pairs are used in each network (no values are trimmed or rounded to 0). Node placement uses the sparse stress majorization algorithm (Ortmann, Klimenta, & Brandes, 2016), with similarity scores used as edge weights. Label color corresponds to results of Louvain community detection (Blondel et al., 2008) with similarity scores used as edge weights.

sites where disciplinary standards are enforced and codified, and thus the sites where disciplines are brought into being (Geiger, 1990; Hicks & Stapleford, 2016; Pence & Hicks, n.d.). Departments specialize in just a few topics, and at higher values of k these topics separate departments from each other (with high Hellinger distance/low Hellinger similarity). That is, topics are distinctive and characteristic of departments. At the same time, higher-level disciplinary relationships between departments can be recovered using clustering methods. Thus the topic models distinguish, for example, behavioral science or conservation science and policy. For the ORUs, most appear to specialize in distinctive combinations of topics that are well represented in departments. That is, the ORUs do not seem to work on ORU-specific topics, but instead combine disciplinary topics (in either interdisciplinary or multidisciplinary ways). These combinations of topics may correspond to an ORU's distinctive research aims or linguistic affordances that have been developed to facilitate interdisciplinary research. The one apparent exception to this pattern is PICN; here the topic models do seem to have identified a distinctive topic for this ORU, locating it on the margins of the similarity networks.

3.3. Discursive Impacts

3.3.1. Discursive breadth

H1 states that ORU interdisciplinarity may lead to increased "discursive breadth," operationalized as the entropy of the topic distribution. Figure 13 visualizes the entropy for each individual researcher in the data set, by ORU affiliation status and across selected values of k.
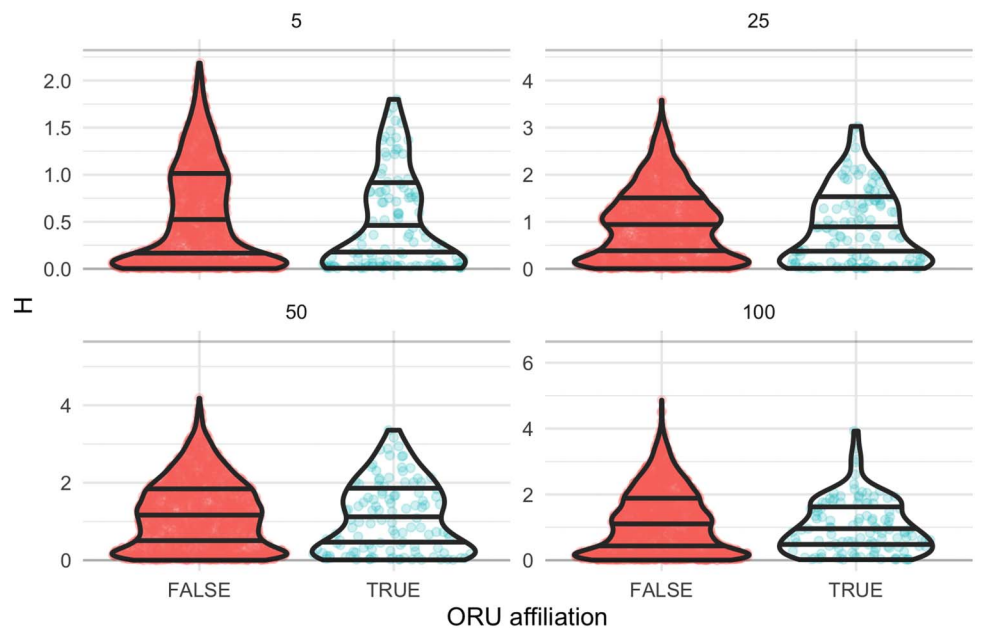
Figure 13. Researcher entropies, by ORU affiliation status and across selected values of *k* (number of topics). Violin plots include 25th, 50th (median), and 75th percentiles. Color fills are points for individual researchers, staggered horizontally to correspond to the violin plots. Panels correspond to values of *k*. Gray horizontal lines at the top of each plot indicate the maximum possible entropy (uniform distribution across all topics) for that value of *k*.
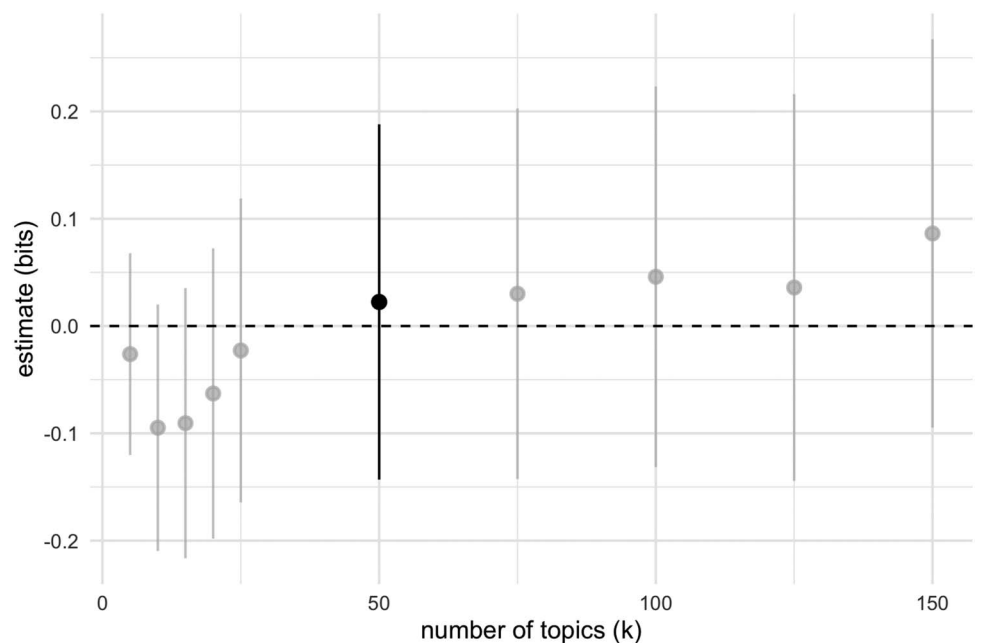

Figure 14. Coefficient estimates for association between topic entropy and ORU affiliation across values of *k* (number of topics). Whiskers are 95% confidence intervals. Regression models include as controls attributed gender, first year of publication, number of publications, number of coauthors, and departmental dummies. *k* = 50 is highlighted because it was judged "best" by some but not all goodness-of-fit statistics.

(Figure S5 shows entropies across all values of k.) Across values of k, the distributions for ORU-affiliated and nonaffiliated researchers are similar: For $k > 5$ the median researcher for both has entropy $H \approx 1$, roughly corresponding to two topics, 75% of researchers have entropy less than $H \approx 2$, roughly corresponding to four topics, and the modal researcher has $H \approx 0$, roughly corresponding to a single topic. In other words, most researchers work in only a handful of topics, whether they are affiliated with an ORU or not. This pattern is consistent even for topic models with high values of k, though the right tail (the "neck" of the violin) is longer as k increases (especially for non-ORU-affiliated researchers), meaning there are a few researchers in the data set who work on a very wide range of topics.

Figure 14 shows the regression coefficient estimates for the association between topic entropy and ORU affiliation across all topic models, controlling for first year of publication, gender, department affiliation, and logged number of documents and coauthors. Across all models, confidence intervals generally cover from −0.15 to 0.2 bits, with point estimates in the range −0.1 to 0.1 bits. Using 0.5 bits ("half of a coin flip") as a threshold for substantive difference (that is, treating any value between −0.5 and 0.5 as too small to be interesting), the models uniformly indicate that the difference in discursive breadth between ORU authors and their peers is *trivial*.

H1 does not seem to be supported by these results. This hypothesis posits a connection between interdisciplinary, the notion of "discursive breadth," and topic distribution entropy as an operationalization of this notion of breadth. The apparent failure of H1 might be due to failures at any of these three points. First, the ORUs at UC Davis might not have effectively promoted interdisciplinarity. Second, "discursive breadth" might be an inapt way of characterizing interdisciplinarity. And third, entropy might be an inapt operationalization of
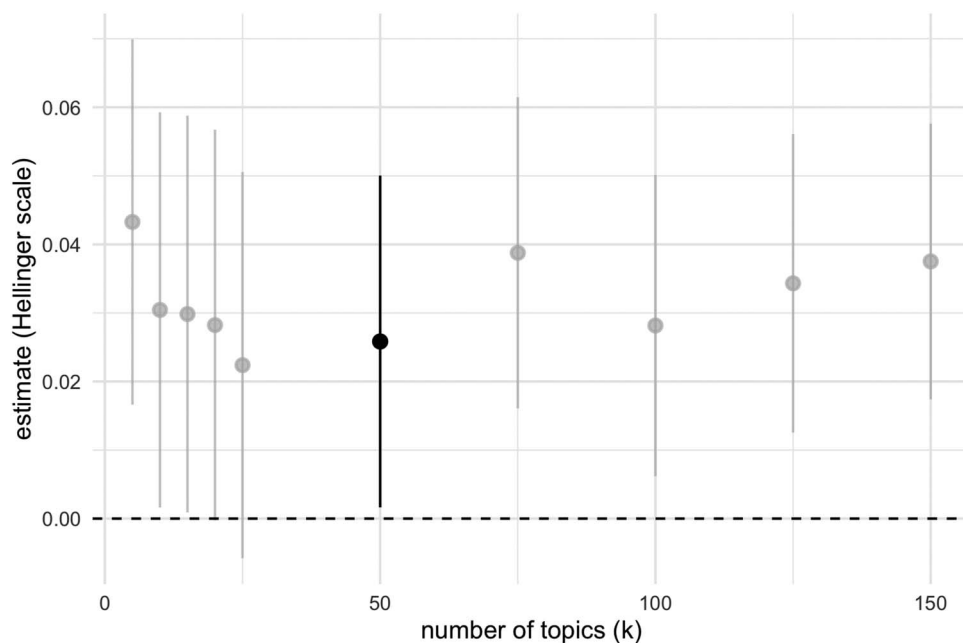

Figure 15. Coefficient estimates for association between distance to departmental mean distribution and ORU affiliation across values of k (number of topics). Whiskers are 95% confidence intervals. Regression models include as controls attributed gender, first year of publication, number of publications, number of coauthors, and departmental dummies. $k = 50$ is highlighted because it was judged "best" by some but not all goodness-of-fit statistics.
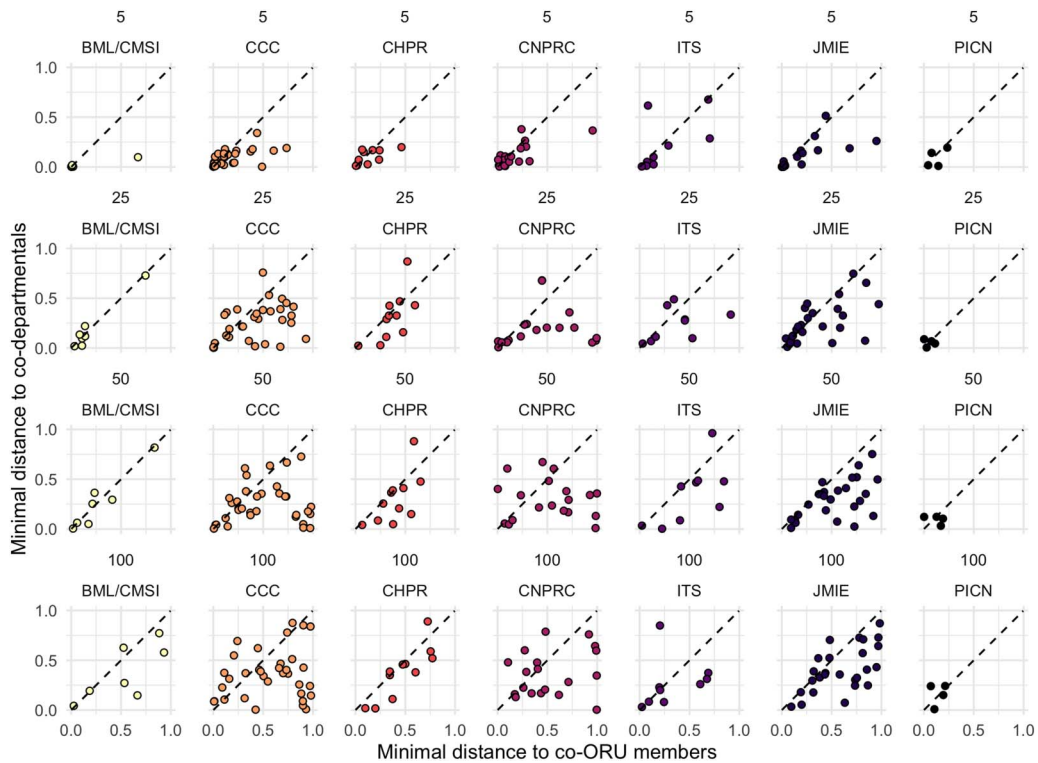
Figure 16. Minimal distance to ORU peers vs. departmental peers. Both *x*- and *y*-axes are on the Hellinger distance scale (0–1). The dashed line in each panel indicates *y* = *x*. Points above this line are closer to ORU peers than departmental peers, supporting H3. Note that comparisons to the dashed line should be made *vertically* or *horizontally*.

"discursive breadth." For example, if the topic model included an ORU-specific "interdisciplinary topic," then an interdisciplinary researcher might have a narrower distribution than their disciplinary peers[9]. However, the examination of department- and ORU-level topic distributions in Figures 11, 12, S2, S3, and S4 indicated that, except for PICN, the topic models did not include ORU-specific "interdisciplinary topics," suggesting that the third link is not the problem. If the other analyses of discursive impacts give indications that the UC Davis ORUs have effectively promoted interdisciplinarity, then this would suggest that the problem with H1 is in the second or third link. So this would imply that the problem is the second link. On the other hand, if these other analyses indicate that the ORUs have not effectively promoted interdisciplinarity, this would indicate that the problem is with the first link, and consequently, "discursive breadth," operationalized by the entropy of topic distributions, might still be apt for measuring interdisciplinarity.

3.3.2. Departmental distance

H2 states that ORU interdisciplinarity may lead to increased departmental distance, that is, increased Hellinger distance from the department-level topic distribution. Figure 15 shows co-efficient estimates for the association between departmental distance and ORU affiliation across values of *k* (number of topics), along with 95% confidence intervals. Here the estimates may appear to support H2, as they are generally positive. However, most confidence intervals

---

[9] Thanks to an anonymous reviewer for suggesting this potential problem.

end well below 0.06, and (except for $k = 5$) point estimates are all in the the range 0.02–0.04. Recall that Hellinger distance is on a 0–1 scale. On this scale, distances less than 0.05 would seem to be trivial. That is, there does not seem to be a meaningful difference between ORU faculty and the mean of their departmental peers, and so H2 does not appear to be supported either.

It might be suspected that departmental distance effects could vary across ORUs. Figure S6 reports coefficient estimates for ORU dummy variables, rather than the binary yes/no ORU affiliation used above; "no ORU affiliation" is used as the contrast value for the ORU dummies.

Figure S6 does indeed suggest that the potential association between ORU affiliation and departmental distance does vary across ORUs, albeit still to a limited extent. Except for CHPR and JMIE, for at least some values of k the point estimate is greater than 0.10. There does seem to be some evidence of a nontrivial association for CNPRC, PICN, and AQRC (though recall that this last had only a single faculty affiliate during the analysis period). So H2 might be true for these three ORUs, but not the others.

3.3.3. ORU-department relative distance

H3 proposes that ORU interdisciplinarity leads researchers to be closer to their ORU peers than their (non-ORU-affiliated) departmental peers in discursive space. Figure 16 shows scatterplots for minimal distances to both kinds of peers, for each ORU and four values of k. In these scatterplots, the dashed line indicates $y = x$. Points above this line are closer to ORU peers than departmental peers, so these points would be compatible with H3.
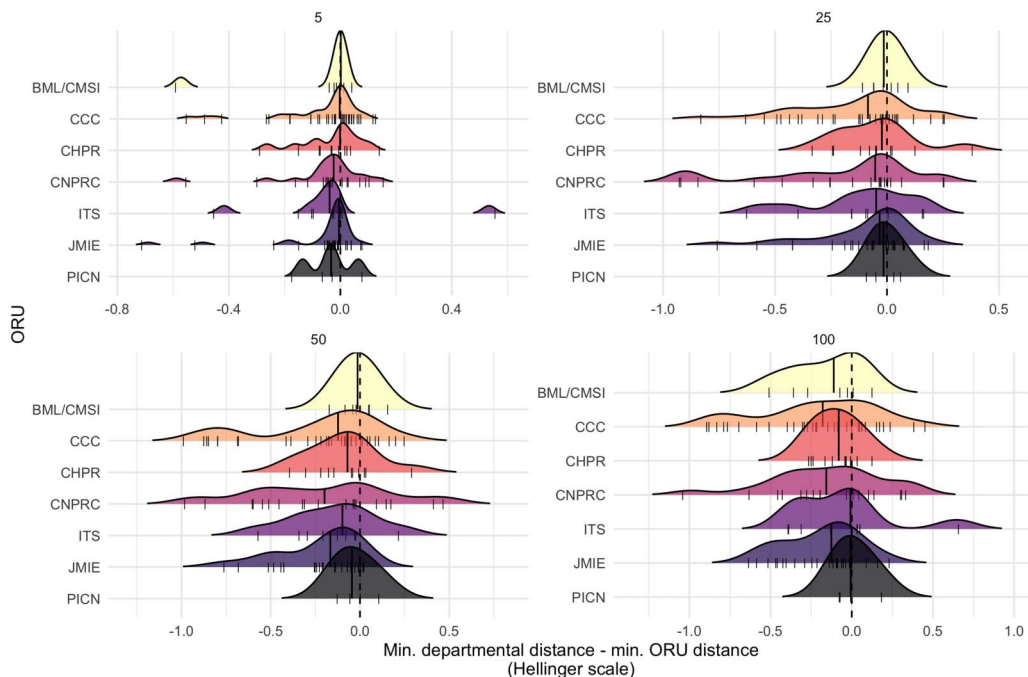

Figure 17. Comparison of minimal ORU and departmental distances, across ORUs and values of k. Positive x-axis values indicate that departmental distance is greater than ORU distance, supporting H3. Dashed vertical line indicates 0; solid lines within densities indicate median values; and small vertical dashes indicated individual values.

For most ORUs, across values of k, most researchers are located near or somewhat below the dashed line. This means that researchers are typically *equidistant from or closer to their closest departmental peers* than their closest ORU peers.

Because distance comparisons in scatterplots can be misleading (comparing vertical distance to the dashed line, not Euclidean distance), Figure 17 shows the distribution of these comparisons. In this figure, positive x-axis values would indicate that departmental distance is greater than ORU distance, which would support H3. In this figure, modal and median values are all negative or near 0. While there are a few exceptional individuals, ORU faculty are generally *equidistant from or closer to their closest departmental peers*. As with the other two hypotheses, these findings conflict with H3.

Figures 18 and 19 show distance comparisons to the ORU- and department-level distributions. For most ORUs and most values of k, the median and modal differences are positive, indicating that generally *ORU researchers are closer to the ORU than their departments*, which is consistent with H3. This tendency is most obvious for PICN, with a median > 0.50 on the Hellinger scale for $k > 5$. The one exception is JMIE, which has a negative median across all values of k.

All together, the distance comparisons suggest the following. ORU faculty are often quite close to some individual non-ORU researchers in their home departments, but still relatively far from the "center" of the department as a whole. This suggests that ORU faculty are not the only members of their departments engaged in cross-disciplinary research, but they are still
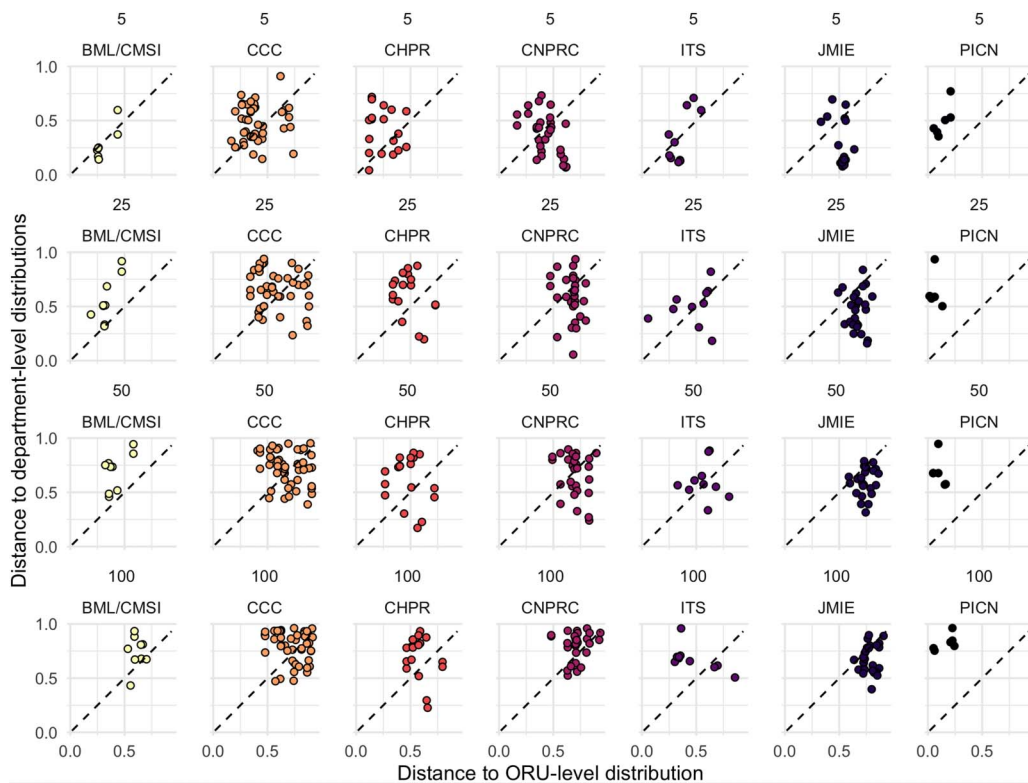

Figure 18. Distance to ORU- and department-level topic distributions. Both x- and y-axes are on the Hellinger distance scale (0–1). The dashed line in each panel indicates $y = x$. Points above this line are closer to ORU peers than departmental peers, supporting H3. Note that comparisons to the dashed line should be made *vertically* or *horizontally*.
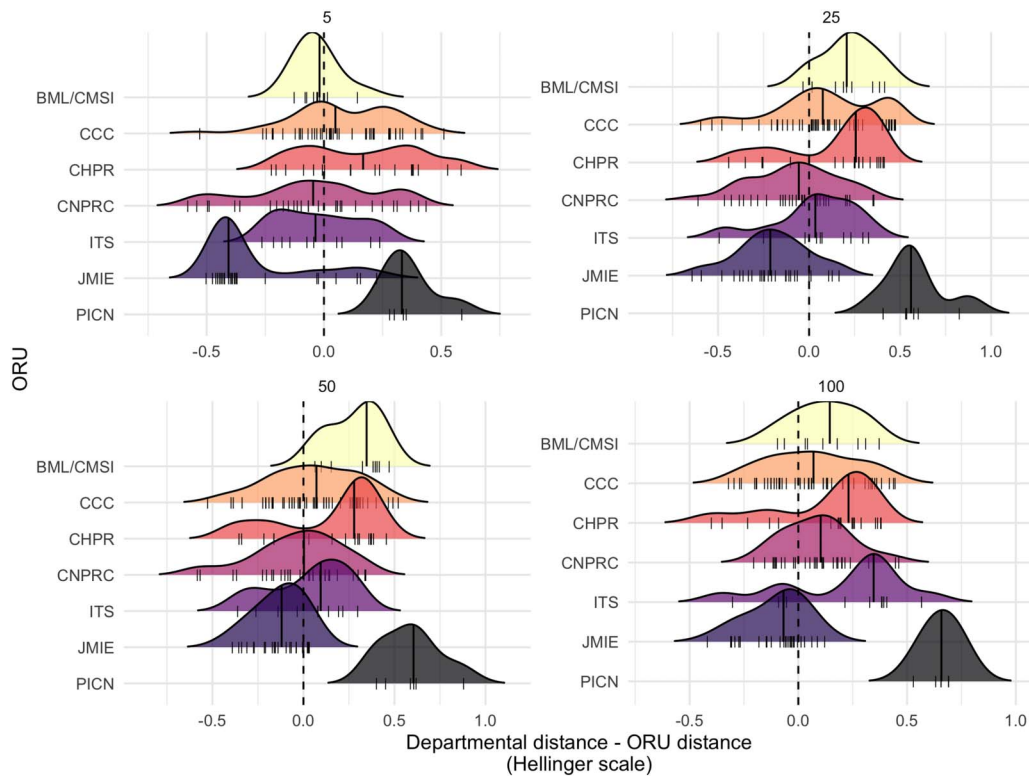
Figure 19. Comparison of distance to ORU- and department-level topic distributions, across ORUs and values of *k*. Positive *x*-axis values indicate that departmental distance is greater than ORU distance, supporting H3. Dashed vertical line indicates 0; solid lines within densities indicate median values; and small vertical dashes indicated individual values.

somewhat outside of the disciplinary mainstream. As originally posed, H3 is ambiguous between individual-level and organizational-level comparisons. The individual reading is not supported here, but there is support for the organizational-level reading.

4. DISCUSSION

The analysis of productivity impacts provides some evidence that ORUs have increased the productivity of affiliated faculty at UC Davis. Specifically, the sequence of regression models suggests that ORUs have increased the number of coauthors that affiliated faculty have; that this increased collaboration leads to increased publications and citations; but that ORUs have had at most a small direct effect on publications and citations.

Due to the limitations of the study design and data, we cannot be sure whether this relationship is indeed a causal effect of ORUs on productivity. The data are also compatible with an effect of productivity on ORUs (that is, perhaps ORUs have tended to recruit faculty who already tend to be more productive), an unmeasured common cause (for example, perhaps more extroverted faculty have tended both to be more productive and to be recruited to ORUs), or indeed a combination of multiple causal relationships. A more complex possibility, in line with the Matthew Effect (DiPrete & Eirich, 2006; Merton, 1968), is that ORUs may have enhanced pre-existing trends, by directing more resources towards faculty who already tended to be more productive and have more coauthors. To sort out these plausible causal relationships, we would need data on when researchers began their affiliation with ORUs. Unfortunately, the UC Davis Office of Research does not keep such data.

Keeping these limitations of any causal interpretation in mind, the productivity findings of this study suggest that ORUs—and similar infrastructure for interdisciplinary research—may have a key social network formation role. Hicks and Simmons (2019) and Hicks, Coil et al. (2019) also analyzed specific interdisciplinary research funding programs, finding evidence that these programs appeared to be effective at supporting novel collaborations and stimulating the formation of a new research community, respectively.

Turning to discursive impacts, the data do not appear to be compatible with hypotheses 1 or 2, discursive breadth or departmental distance; nor with hypothesis 3, relative distance from ORU vs. department, when read individualistically. Here, the expected effects of interdisciplinarity (rather than multidisciplinarity) do not appear. These results suggest that, if ORUs at UC Davis have any effect on researchers' location and distribution in "discursive space," they have fostered multidisciplinarity rather than interdisciplinarity.

There is some support for hypothesis 3, relative distance from ORU vs. department, if this comparison is interpreted at an organizational rather than individual level. These findings indicate that ORU affiliated researchers are located "away" from the mainstream of their home disciplines, and are closer to ORU-distinctive research questions and methods. But, thinking of the results for hypothesis 2, non-ORU-affiliated researchers are not homogeneous, and many of them are located just as far from the disciplinary "center" as ORU-affiliated researchers. That is, both ORU-affiliated and non-ORU-affiliated faculty are distributed around the disciplinary "center," with comparable *magnitudes*; but ORU-affiliated faculty have a distinct *direction*.

JMIE is an important exception to the pattern observed for hypothesis 3. The median JMIE researcher is closer to their home department than to the JMIE topic distribution. JMIE also has the widest topic distribution. Where ORUs such as PICN and BML/CMSI appear to be tightly focused on a small set of issues (represented by a single topic), JMIE appears to be much more heterogeneous (spread across several topics). The evidence from JMIE consistently points to multidisciplinarity rather than interdisciplinarity.

Multidisciplinarity is often seen as inferior to interdisciplinarity; for example, Holbrook defines multidisciplinarity as "the (*mere*) juxtaposition of two or more academic disciplines focused on a single problem" (Holbrook, 2013, p. 1867, parentheses in original, my emphasis). However, interdisciplinary research faces serious challenges that multidisciplinary research might avoid. First, genuine integration across interdisciplinary lines is extremely difficult, not merely for pragmatic/logistical reasons but for deep conceptual reasons as well. Eigenbrode et al. (2007) argue that research disciplines are divided not only by subject matter, but also by conceptual frameworks, research methods, standards of evidence, emphasis on the social (rather than narrowly epistemic) value of research, and metaphysical assumptions about the nature of the world. Drawing on philosophy of language and philosophy of science, they develop a workshop-style intervention designed to make these divides explicit in cross-disciplinary teams (O'Rourke & Crowley, 2012); although clarifying the problem is not the same as solving it. Holbrook (2013) takes this kind of insight further, drawing on the work of Kuhn and MacIntyre to argue that interdisciplinary collaboration is conceptually incoherent because it would require crossing the boundaries of incommensurable conceptual schemes. Even if we do not follow Holbrook (2013) to this radical conclusion, we can recognize the point that interdisciplinary research will tend to face deep communicative-conceptual challenges.

In addition, Brister (2016) notes that cross-disciplinary collaborations can exemplify the same status and power hierarchies as academia more generally, leading to a phenomenon that she calls "disciplinary capture." For example, as a natural science, biology generally has

higher status than anthropology; this hierarchy appeared in a collaboration between biologists and anthropologists, with the result that "Both groups of scientists … perceived that conservation activities are dominated by biological research" (Brister, 2016, p. 86). Fernández Pinto (2016) makes similar points in terms of "scientific imperialism".

Because of these challenges, interdisciplinary research may be difficult, to the point of being highly impractical, without specific interventions or institutional designs. Even if successful interdisciplinary research can bring the epistemic and social benefits that it is supposed to, these may not be worth the costs required in particular cases. So, all things considered, in particular cases multidisciplinarity might be preferable to interdisciplinarity.

ACKNOWLEDGMENTS

COMPETING INTERESTS AND FUNDING

DATA AVAILABILITY

Unless otherwise noted, all data used in this project was retrieved from Scopus, using either the web interface or application programming interface (API), between November 2018 and June 2019. Due to intellectual property restrictions the data cannot be made publicly available. Some downstream analysis files may be provided upon request. All code used in data collection and analysis is available at https://github.com/dhicks/orus.

REFERENCES

Amrhein, V., Trafimow, D., & Greenland, S. (2018). Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication. *The American Statistician*, *73*(sup1), 262–270. https://doi.org/10.1080/00031305.2018.1543137

Andersen, H. (2012). Conceptual development in interdisciplinary research. In U. Feest & F. Steinle (Eds.), *Scientific concepts and investigative practice* (pp. 3–271). De Gruyter. https://doi.org/10.1515/9783110253610.271

Beaudry, C., & Larivière, V. (2016). Which gender gap? Factors affecting researchers' scientific impact in science and medicine. *Research Policy*, *45*(9), 1790–1817. https://doi.org/10.1016/j.respol.2016.05.009

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics*, *64*(3), 351–374. https://doi.org/10.1007/s11192-005-0255-6

Boyack, K. W., Newman, D., Duhon, R. J., Klavans, R., Patek, M., … Börner, K. (2011). Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches. *PLOS ONE*, *6*(3), e18029. https://doi.org/10.1371/journal.pone.0018029, PubMed: 21437291

Brister, E. (2016). Disciplinary capture and epistemological obstacles to interdisciplinary research: Lessons from central African conservation disputes. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *56*, 82–91. https://doi.org/10.1016/j.shpsc.2015.11.001, PubMed: 26651422

Bruun, H., Hukkinen, J., Huutoniemi, K., & Klein, J. T. (2005). *Promoting interdisciplinary research: The case of the Academy of Finland* (p. 206). Helsinki, Finland: Academy of Finland.

Cameron, E. Z., White, A. M., & Gray, M. E. (2016). Solving the productivity and impact puzzle: Do men outperform women, or are metrics biased? *BioScience*, *66*(3), 245–252. https://doi.org/10.1093/biosci/biv173

Cartwright, N. (1999). *The dappled world*. Retrieved from https://books.google.com/books?id=tOFv_i9oiAgC&printsec=frontcover. https://doi.org/10.1017/CBO9781139167093

Chauvin, S., Mulsant, B. H., Sockalingam, S., Stergiopoulos, V., Taylor, V. H., & Vigod, S. N. (2019). Gender differences in research productivity among academic psychiatrists in Canada. *The Canadian Journal of Psychiatry*, *64*(6), 415–422. https://doi.org/10.1177/0706743718802798, PubMed: 30616409

Collins, H., Evans, R., & Gorman, M. (2007). Trading zones and interactional expertise. *Studies in History and Philosophy of Science Part A*, *38*(4), 657–666. https://doi.org/10.1016/j.shpsa.2007.09.003

CRAN Team & Temple Lang, D. (2020). RCurl: General Network (HTTP/FTP/…) Client Interface for R (Version 1.98-1.2). Retrieved from https://CRAN.R-project.org/package=RCurl

DiPrete, T. A., & Eirich, G. M. (2006). Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual Review of Sociology*, *32*(1), 271–297. https://doi.org/10.1146/annurev.soc.32.061604.123127

Eigenbrode, S. D., O'Rourke, M., Wulfhorst, J. D., Althoff, D. M., Goldberg, C. S., … Bosque-Pérez, N. A. (2007). Employing philosophical dialogue in collaborative science. *BioScience*, *57*(1), 55. https://doi.org/10.1641/B570109

Etzkowitz, H., & Kemelgor, C. (1998). The role of research centres in the collectivisation of academic science. *Minerva*, *36*(3), 271–288. https://doi.org/10.1023/A:1004348123030

Fernández Pinto, M. (2016). Economics imperialism in social epistemology: A critical assessment. *Philosophy of the Social Sciences*, *46*(5), 443–472. https://doi.org/10.1177/0048393115625325

Galison, P. (1997). *Image and logic: A material culture of microphysics*. University of Chicago Press.

Geiger, R. L. (1990). Organized Research Units—Their role in the development of university research. *The Journal of Higher Education*, *61*(1), 1–19. https://doi.org/10.1080/00221546.1990.11775088

Ghiasi, G., Larivière, V., & Sugimoto, C. R. (2015). On the compliance of women engineers with a gendered scientific system. *PLOS ONE*, *10*(12), e0145931. https://doi.org/10.1371/journal.pone.0145931, PubMed: 26716831

Han, X. (2020). Evolution of research topics in LIS between 1996 and 2019: An analysis based on latent Dirichlet allocation topic model. *Scientometrics*, *125*(3), 2561–2595. https://doi.org/10.1007/s11192-020-03721-0

Hellinger Distance. (2021). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Hellinger_distance&oldid=1001077827

Hicks, D. J. (2016). Bibliometrics for social validation. *PLOS ONE*, *11*(12), e0168597. https://doi.org/10.1371/journal.pone.0168597, PubMed: 28005974

Hicks, D. J., Coil, D. A., Stahmer, C. G., & Eisen, J. A. (2019). Network analysis to evaluate the impact of research funding on research community consolidation. *PLOS ONE*, *14*(6), e0218273. https://doi.org/10.1371/journal.pone.0218273, PubMed: 31211808

Hicks, D., & Melkers, J. (2012). Bibliometrics as a tool for research evaluation. In A. Link & N. Vornatas (Eds.), *Handbook on the theory and practice of program evaluation* (pp. 323–349). Cheltenham, UK, and Northampton, MA: Edward Elgar. Retrieved from https://works.bepress.com/diana_hicks/31/

Hicks, D., & Simmons, R. (2019). The National Robotics Initiative: A five-year retrospective. *IEEE Robotics Automation Magazine*, *26*(3), 2–9. https://doi.org/10.1109/MRA.2019.2912860

Hicks, D. J., Stahmer, C., & Smith, M. (2018). Impacting capabilities: A conceptual framework for the social value of research. *Frontiers in Research Metrics and Analytics*, *3*, 24. https://doi.org/10.3389/frma.2018.00024

Hicks, D. J., & Stapleford, T. A. (2016). The virtues of scientific practice: MacIntyre, virtue ethics, and the historiography of science. *Isis*, *107*(3), 449–472. https://doi.org/10.1086/688346, PubMed: 28707850

Hitchcock, C., & Rédei, M. (2021). Reichenbach's common cause principle. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2021). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/sum2021/entries/physics-Rpcc/

Hofstra, B., Kulkarni, V. V., Munoz-Najar Galvez, S., He, B., Jurafsky, D., & McFarland, D. A. (2020). The diversity–innovation paradox in science. *Proceedings of the National Academy of Sciences*, 201915378. https://doi.org/10.1073/pnas.1915378117, PubMed: 32291335

Holbrook, J. B. (2013). What is interdisciplinary communication? Reflections on the very idea of disciplinary integration. *Synthese*, *190*(11), 1865–1879. https://doi.org/10.1007/s11229-012-0179-7

Huutoniemi, K., Klein, J. T., Bruun, H., & Hukkinen, J. (2010). Analyzing interdisciplinarity: Typology and indicators. *Research Policy*, *39*(1), 79–88. https://doi.org/10.1016/j.respol.2009.09.011

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York, NY: Springer New York. Retrieved from https://link.springer.com/10.1007/978-1-4614-7138-7

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago, IL: University of Chicago Press. https://doi.org/10.7208/chicago/9780226458106.001.0001

Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Bibliometrics: Global gender disparities in science. *Nature News*, *504*(7479), 211–213. https://doi.org/10.1038/504211a, PubMed: 24350369

Lee, C. J. (2020). The reference class problem for credit valuation in science. *Philosophy of Science*, *87*(5), 1026–1036. https://doi.org/10.1086/710615

Leydesdorff, L., & Rafols, I. (2011). Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Informetrics*, *5*(1), 87–100. https://doi.org/10.1016/j.joi.2010.09.002

Longino, H. E. (2013). *Studying human behavior: How scientists investigate aggression and sexuality*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226921822.001.0001

Lu, K., & Wolfram, D. (2012). Measuring author research relatedness: A comparison of word-based, topic-based, and author cocitation approaches. *Journal of the American Society for Information Science and Technology*, *63*(10), 1973–1986. https://doi.org/10.1002/asi.22628

Malaterre, C., Lareau, F., Pulizzotto, D., & St-Onge, J. (2020). Eight journals over eight decades: A computational topic-modeling approach to contemporary philosophy of science. *Synthese*. https://doi.org/10.1007/s11229-020-02915-6

McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC Press.

Merton, R. K. (1968). The Matthew Effect in science: The reward and communication systems of science are considered. *Science*, *159*(3810), 56–63. https://doi.org/10.1126/science.159.3810.56, PubMed: 5634379

Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European Journal of Operational Research*, *246*(1), 1–19. https://doi.org/10.1016/j.ejor.2015.04.002

Morgan, A. C., Economou, D. J., Way, S. F., & Clauset, A. (2018). Prestige drives epistemic inequality in the diffusion of scientific ideas. *EPJ Data Science*, *7*(1), 40. https://doi.org/10.1140/epjds/s13688-018-0166-4

Nature. (2016). Meet the challenge of interdisciplinary science. *Nature*, *534*(7609), 589–590. https://doi.org/10.1038/534589b, PubMed: 27357757

Nichols, L. G. (2014). A topic model approach to measuring interdisciplinarity at the National Science Foundation. *Scientometrics*, *100*(3), 741–754. https://doi.org/10.1007/s11192-014-1319-2

O'Rourke, M., & Crowley, S. (2012). Philosophical intervention and cross-disciplinary science: The story of the Toolbox Project. *Synthese*, 1–18. https://doi.org/10.1007/s11229-012-0175-y

Ortmann, M., Klimenta, M., & Brandes, U. (2016). A sparse stress model. *arXiv:1608.08909 [Cs]*. Retrieved from https://arxiv.org/abs/1608.08909

Pence, C., & Hicks, D. (n.d.). *Challenges for 'community' in science and values: Cases from robotics research*.

Potochnik, A. (2017). *Idealization and the aims of science*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226507194.001.0001

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, 82(2), 263–287. https://doi.org/10.1007/s11192-009-0041-y

Rehs, A. (2020). A structural topic model approach to scientific reorientation of economics and chemistry after German reunification. *Scientometrics*, 125(2), 1229–1251. https://doi.org/10.1007/s11192-020-03640-0

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., … Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. (2004). The author-topic model for authors and documents. *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence* (pp. 487–494). AUAI Press.

Sá, C. M. (2008). 'Interdisciplinary strategies' in U.S. research universities. *Higher Education*, 55(5), 537–552. https://doi.org/10.1007/s10734-007-9073-5

Sá, C. M., & Oleksiyenko, A. (2011). Between the local and the global: Organized research units and international collaborations in the health sciences. *Higher Education*, 62(3), 367–382. https://doi.org/10.1007/s10734-010-9393-8

Sommer, R. (1994). Serving two masters: Center for Consumer Research at the University of California, Davis, 1976–1992. *Journal of Consumer Affairs*, 28(1), 170–186. https://doi.org/10.1111/j.1745-6606.1994.tb00821.x

spaCy. (2018). Industrial-strength Natural Language Processing (NLP) with Python and Cython (Version 2.0.3) [Python]. Explosion AI. Retrieved from https://github.com/explosion/spaCy

Stahler, G. J., & Tash, W. R. (1994). Centers and institutes in the research university: Issues, problems, and prospects. *The Journal of Higher Education*, 65(5), 540–554. https://doi.org/10.1080/00221546.1994.11778519

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science*, 19(3), 387–420. https://doi.org/10.1177/030631289019003001

Symonds, M. R. E., Gemmell, N. J., Braisher, T. L., Gorringe, K. L., & Elgar, M. A. (2006). Gender differences in publication output: Towards an unbiased metric of research performance. *PLOS ONE*, 1(1), e127. https://doi.org/10.1371/journal.pone.0000127, PubMed: 17205131

Taddy, M. A. (2012). On estimation and selection for topic models. *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 10. La Palma, Canary Islands.

van Arensbergen, P., van der Weijden, I., & van den Besselaar, P. (2012). Gender differences in scientific productivity: A persisting phenomenon? *Scientometrics*, 93(3), 857–868. https://doi.org/10.1007/s11192-012-0712-y, PubMed: 23162173

van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(86), 2579–2605. Retrieved from https://jmlr.org/papers/v9/vandermaaten08a.html

Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., … Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*, 5(1), 14–26. https://doi.org/10.1016/j.joi.2010.06.004

Wickham, H., & RStudio. (2017). tidyverse: Easily install and load the 'Tidyverse' (Version 1.2.1). Retrieved from https://CRAN.R-project.org/package=tidyverse

Yan, E., & Ding, Y. (2012). Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other. *Journal of the American Society for Information Science and Technology*, 63(7), 1313–1326. https://doi.org/10.1002/asi.22680

Youtie, J., Kay, L., & Melkers, J. (2013). Bibliographic coupling and network analysis to assess knowledge coalescence in a research center environment. *Research Evaluation*, 22(3), 145–156. https://doi.org/10.1093/reseval/rvt002