TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 13 (2021) 399–413

ISSN:1756-8765 online
DOI: 10.1111/tops.12526

This article is part of the topic "2018 Rumelhart Prize Issue Honoring Michael K. Tanenhaus," John C. Trueswell (Topic Editor). For a full listing of topic papers, see https://onlinelibrary.wiley.com/toc/17568765/2021/13/2.

Integration by Parts: Collaboration and Topic Structure in the CogSci Community

Isabella DeStefano, Lauren A. Oey, Erik Brockbank, Edward Vul

Department of Psychology, University of California, San Diego

Received 16 November 2020; received in revised form 11 February 2021; accepted 12 February 2021

Abstract

Is cognitive science interdisciplinary or multidisciplinary? We contribute to this debate by examining the authorship structure and topic similarity of contributions to the Cognitive Science Society from 2000 to 2019. Our analysis focuses on graph theoretic features of the co-authorship network—edge density, transitivity, and maximum subgraph size—as well as clustering within the space of scientific topics. We also combine structural and semantic information with an analysis of how authors choose their collaborators based on their interests and prior collaborations. We compare findings from CogSci to abstracts from the Vision Science Society over the same time frame and validate our approach by predicting new collaborations in the 2020 CogSci proceedings. Our results suggest that collaboration across authors and topics within cognitive science has become increasingly integrated in the last 19 years. More broadly, we argue that a formal quantitative approach which combines structural co-authorship information and semantic topic analysis provides inroads to questions about the level of interdisciplinary collaboration in a scientific community.

Keywords: Co-authorship networks; Topic modeling; Interdisciplinarity; Multidisciplinarity; Scientometrics

Correspondence should be sent to Isabella DeStefano and Lauren Oey, Department of Psychology, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093. E-mail: idestefa@ucsd.edu; loey@ucsd.edu

1. Introduction

Since its foundation, the Cognitive Science Society sought to unify various disciplines of study under one interdisciplinary research field. Recently, criticism of the success of this mission has sparked debate about whether cognitive science, in its current form, is fundamentally multidisciplinary rather than interdisciplinary (Gray, 2019; Núñez et al., 2019; Schunn, Crowley, & Okada, 1998). The distinction between these community structures is subtle, making any claims favoring one or the other difficult to evaluate. Broadly, the debate centers on the idea that a research community is more *multidisciplinary* if collaborations happen mostly within small groups and there is greater topical isolation of each group from the rest. On the other hand, a more *interdisciplinary* research community will show fewer isolated groups and less separation of research interests across groups. Researchers hoping to promote progress in a field might therefore strive for a more interdisciplinary, rather than multidisciplinary, approach.

But how do we measure interdisciplinarity in a way that captures meaningful differences within diverse communities? Currently, there is no consensus on a single measure that best aligns with this abstract concept. Previous studies quantified interdisciplinarity by looking at the publication record in journals associated with a given discipline. Some of these studies have examined the distribution of journals cited (Goldstone & Leydesdorff, 2006; Núñez et al., 2019; Porter, Cohen, Roessner, & Perreault, 2007), the citation networks (Rafols & Meyer, 2010), and the journals that authors previously published in (Bergmann, Dale, Sattari, Heit, & Bhat, 2017). But this earlier research aiming to quantify interdisciplinarity was primarily targeted at the categorization of disciplines. These measures are subject to inconsistencies across classification systems, leading to variable conclusions (Wagner et al., 2011). Others have used departmental affiliation and educational background (Núñez et al., 2019; Schunn et al., 1998), but research interests often shift over the course of a lifetime, which makes the affiliation label a transient indicator (Porter et al., 2007).

Recent efforts to measure interdisciplinarity or characterize the level of collaboration in a field have sought to address these challenges by incorporating more data-rich, bottom-up measures. For example, the *contents* of scientific work in a number of fields outside cognitive science have been described using text-based clustering (Gowanlock & Gazan, 2013), word co-occurrence (Ravikumar, Agrahari, & Singh, 2015), semantic structural analysis (Parinov & Kogalovsky, 2014), and topic modeling (Nichols, 2014). Further, the *structural* properties of research collaboration have been described using network analysis tools applied to publication in diverse scientific fields (Barabási et al., 2002; Newman, 2001, 2004), in management and organizational research (Acedo, Barroso, Casanueva, & Galaán, 2006), and in international collaborations (Wagner & Leyesdorff, 2005). These measures offer the ability to characterize work in a field without relying on the manual assignment of authors or publications to particular disciplines.

Based on this work, what conclusions can be drawn about cognitive science specifically? A recent comprehensive attempt to assess whether cognitive science reached the interdisciplinary status it aspired to comes from Núñez et al. (2019). The authors combine bibliometric indicators—the affiliation of authors in the journal *Cognitive Science* and the disciplines of journals cited therein—as well as socio-institutional ones: the doctoral training of faculty in cognitive science departments and the coursework requirements of cognitive science undergraduate cores. Both of these latter measures are already constrained by the few institutions that offer undergraduate training or have separate departments in cognitive science at all. Núñez et al. (2019) conclude that there is an imbalanced contribution of the constituent disciplines to cognitive science, suggesting that cognitive science remains premature in its efforts to forge a coherent interdisciplinary field. The results sparked controversy and a range of responses (see overview in Gray, 2019; Núñez et al., 2020), both theoretical and empirical. Many of these addressed the inherent challenges of measuring interdisciplinarity, noting for example that an author's departmental affiliation provides at best "a useful proxy for a scholar's background" (Bender, 2019). Thus, the discussion about the level of interdisciplinary work in cognitive science may benefit from more fine-grained measures of author affiliations and research areas.

In the current work, we aim to contribute to the debate over interdisciplinarity in cognitive science by using a range of data-driven, bottom-up methods which do not require the domain-specific analysis of journals and curricula and which may therefore represent a more generalized approach to addressing the interdisciplinary nature of the field. Though we do not claim to resolve the question of whether cognitive science is fundamentally interdisciplinary or multidisciplinary, we argue that the discussion benefits from the novel measurements we present here, which suggest that collaborations and topics within the field have become increasingly integrated in the last 19 years. Specifically, we address the challenges of defining and measuring interdisciplinarity in cognitive science through a combination of co-authorship network features and topic analysis. We validate our measures using full papers from the Cognitive Science Society proceedings between 2000 and 2019 and abstracts from the Vision Science Society (VSS; only abstracts are submitted) over a similar time frame (2001–2019). We further show that measures derived from network structure and research topics offer a viable means of studying interdisciplinary collaboration and the movement of the field more broadly by using a combination of structure and topic measures to predict new and persisting collaborations in an out-of-sample data set of the 2020 CogSci proceedings.

First, the degree to which a community is interdisciplinary or multidisciplinary may in large part be revealed by who collaborates with whom. Scientific collaboration can be represented as an undirected graph, in which nodes correspond to individual authors and edges between nodes indicate whether any two authors co-authored a paper together (Barabási et al., 2002; Newman, 2001, 2004). Co-authorships within a community containing multiple areas of study can range from highly integrated to highly modular, and the structure of the resulting co-authorship network will reflect this spectrum of possibilities (Fig. 1).

Second, while the collaboration structure of a community no doubt reveals something about the modularity of interdisciplinary work that occurs within it, the ways in which

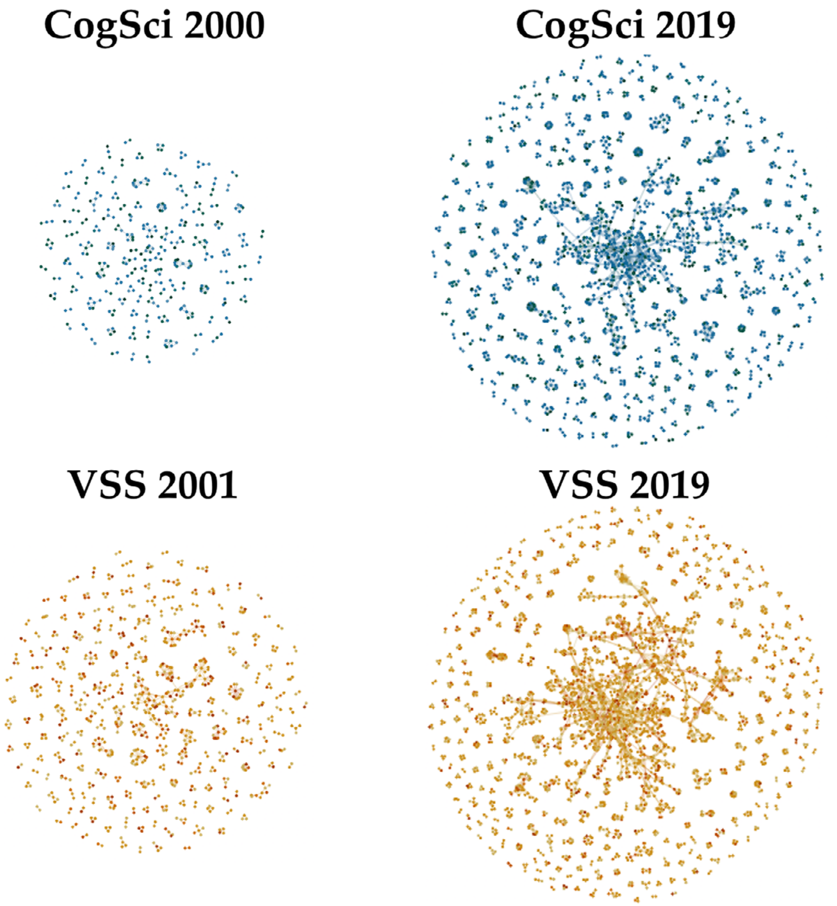

Fig. 1. The co-authorship network of CogSci in 2000 and 2019 and the network of Vision Science Society (VSS) in 2001 and 2019.

research interests combine must play a role as well. To better understand how the *content* of collaborations informs the interdisciplinarity of the field, we use a topic model (Griffiths & Steyvers, 2004) to extract high-level patterns in cognitive science research over the last 19 years. Topic models have been used in previous research to capture trends in the published work within a discipline, including within cognitive science (Cohen Priva & Austerweil, 2015; Rothe, Rich, & Zhi-Wei, 2018). Studies specifically addressing interdisciplinarity have used topic models to complement pre-defined discipline tagging (Nichols, 2014). In the present work, we apply clustering algorithms to the topics that authors study, addressing the separability of the interests and methods of researchers in the field. More distinct clusters in topic space imply greater division between disciplines.

Finally, in an effort to unify both the structure and the content of collaboration within the cognitive science community and to illustrate how these variables contribute to our understanding of multi- and interdisciplinary work, we analyze the degree to which structure and topic measures predict future co-authorship. Patterns of interdisciplinary and multidisciplinary collaboration are ultimately revealed in the ways authors form new collaborations and maintain existing ones; thus, the value of topic and network analysis measures for characterizing collaboration in a field can be measured in part by how well they predict novel and continued collaboration. Drawing on our earlier analyses of topic space and co-authorship structure, we assess the role that prior collaboration and topic similarity play in determining whether two authors will collaborate, using the data from the last 19 years to fit a model which we test with out-of-sample data from papers presented at the 2020 meeting of the Cognitive Science Society.

Together, our analyses address (a) interconnectedness in the co-authorship network structure, (b) clusters in the author topic space, and (c) how collaborations arise from a combination of co-authorship network and topic space measures. Not only do these metrics quantitatively illustrate how authorship within cognitive science has changed over time, but we also believe these measures may provide a meaningful contribution to the multidisciplinary–interdisciplinary debate across science.[1]

2. Data

We retrieved 11,553 full text PDFs (with 12,203 unique authors) from the published *Proceedings of the Annual Meeting of the Cognitive Science Society* from 2000 to 2019.[2] The data are primarily full text conference proceedings papers but also include submitted abstracts. In addition, we retrieved 22,504 VSS Annual Meeting abstracts (with 23,842 unique authors) published in the *Journal of Vision* from 2001 to 2019.[3] Both data sets were processed to extract unique authors, publication year, and the full text of each paper or abstract.

3. Co-authorship network

Using the publication data collected from CogSci and VSS proceedings, we generated a co-authorship network for each year of the conferences with nodes representing authors and edges representing co-authored publications by pairs of authors in that year's proceedings. The graphs were unweighted; that is, edges represented whether two authors published together *at all* in a given year. We analyze three graph-theoretical measures which, when applied to the collaboration networks, provide insight into the level of interdisciplinarity within these conference communities: edge density, transitivity, and maximum subgraph size.

3.1. Edge density

Edge density refers to the proportion of edges within the network relative to the theoretical maximum. Here, the theoretical maximum is determined by the number of edges possible given the total number of publications in that year. For every paper, there exists a fully connected subgraph of the paper's authors with $n(n-1)/2$ edges, where n is the number of authors on that paper. Thus, the full set of N papers, and their associated number of co-authors, sets a theoretical maximum number of edges at $\sum_{i=1}^{N} \frac{n_i(n_i-1)}{2}$. We define edge density for a given year by normalizing the observed number of edges by this theoretical maximum (Eq. 1).

$$\text{edge density} = \frac{|E(G)|}{\sum_{i=1}^{N} \frac{n_i(n_i-1)}{2}} \tag{1}$$

where $|E(G)|$ is the total number of edges in the co-authorship network G for that year, N is the total number of papers published in that year, and n_i is the number of authors on any given paper i. Our edge density metric measures the degree of repeated collaboration between any two authors, as a proportion of the amount of possible collaboration: A higher edge density indicates a higher rate of unique co-authorships. In an interdisciplinary community, we expect a *higher edge density*, indicating that authors tend to publish with a broad set of collaborators.

The edge density metric is shown in Fig. 2. The edge density for both CogSci and VSS appears relatively stable over the range considered. Critically, we note that the edge density for VSS is significantly lower than CogSci ($\hat{\beta} = -0.05$, $p < .001$) and, perhaps

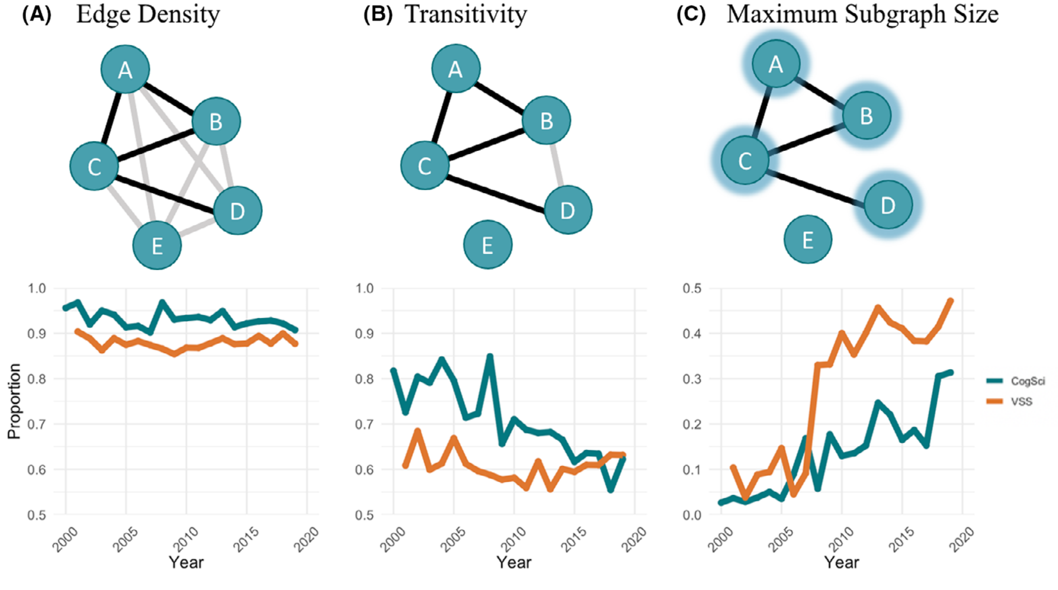

Fig. 2. For each of the network representations, nodes are connected by the black edges. (a) Edge density, or the proportion of edges in the graph to the theoretical maximum given the number of papers and authors per paper. (b) Transitivity, or the proportion of authors whose co-authors also publish together. (c) The maximum subgraph size, or how many authors are in the largest island relative to the full graph.

more importantly, the CogSci edge density measure is relatively close to the theoretical maximum for this measure. This suggests that on average, CogSci authors publish with many unique authors.

3.2. Transitivity

Transitivity measures the probability of a node's adjacent nodes also being connected by an edge, that is, closed triads. Also referred to as the clustering coefficient, transitivity approximates the commonality of local clustering in the graph, such that higher transitivity indicates more clustering. Thus, we would expect an interdisciplinary community to have *lower transitivity*—authors publish with authors across group boundaries.

The transitivity for CogSci appears to decrease over time, whereas the transitivity of VSS remains low over the range considered. Indeed,the slope of a regression against year is significantly negative ($\hat{\beta} = -0.012$, $p < .001$), suggesting that the transitivity of the CogSci network is decreasing meaningfully. This could be influenced by a number of factors, including the possibility that authors have published more papers in the proceedings over time. Nonetheless, the decreasing transitivity suggests that collaborations are often between a more diverse set of individuals: that is, CogSci has become less "clique-y."

3.3. Maximum subgraph

The *size of the maximum subgraph* specifies the proportion of nodes in the graph that are connected to the largest island. A network with a large island relative to the overall size of the graph indicates that many authors are connected to many other authors through their co-authors' and their co-authors' co-authors' collaborations. We would expect an interdisciplinary community to have a *large maximum subgraph*, reflecting the tendency of a large subset of the field to be connected in the same collaboration network.

Across both VSS and CogSci, the maximum subgraph appears to grow over the analyzed time period. Broadly, this suggests that the network of authors within the CogSci community has become increasingly interconnected: The positive slope of this increase in the CogSci data is significant ($\hat{\beta} = 0.014$, $p < .001$).

4. Topic space

To extract the research topics studied by the cognitive science community, we used the stm package in R (Roberts, Stewart, & Tingley, 2014) to fit a topic model to the full text of the papers from the CogSci and VSS proceedings. stm provides functions for cleaning the data by removing punctuation, stopwords, and numbers, then lemmatizing the remaining text. Finally, we fit a latent Dirichlet allocation topic model to the full text documents (Blei, Ng, & Jordan, 2003; Griffiths & Steyvers, 2004). In the model-fitting process we specified 100 topics, which yielded niche yet enduring topics and methods, for example, theory formation (Gopnik & Sobel, 2000), rational analysis (Chater &

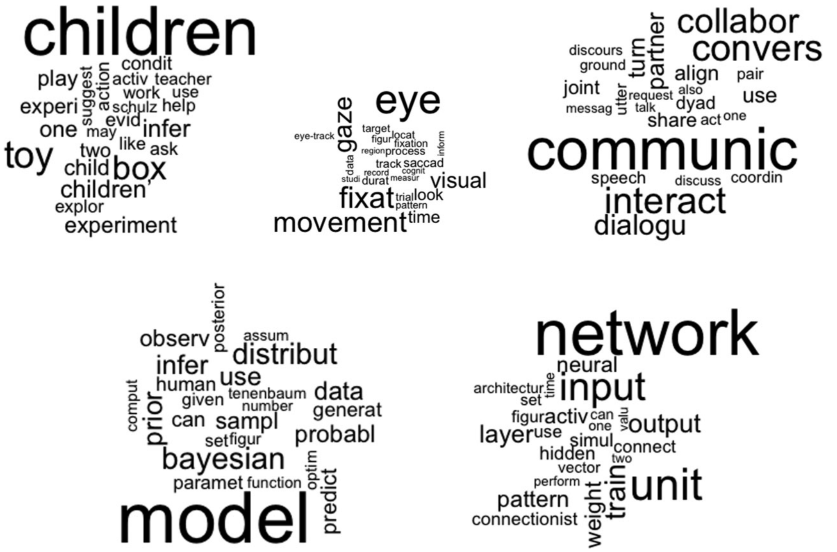

Fig. 3. Frequent lemmas from selected topics: These examples illustrate the level of granularity that the topic model is able to extract from the CogSci texts with 100 topics.

Oaksford, 1999), and connectionism (Rumelhart & McClelland, 1986). See Fig. 3 for several examples of high probability lemmas belonging to particular topics fit by the model. The topic model estimates a distribution over the 100 topics for each paper (or abstract); author locations in topic space were computed to be the overall distribution of their topics across all papers they had published in a given year. To alleviate unusually high spikes within topic distributions resulting from authors who publish only one paper, we smoothed the distributions by regularizing individual authors' topic distributions in a given year to the overall topic distribution for each year.

To understand how *integrated* the topics were year over year, we first applied multidimensional scaling to the authors' distributions across the 100 topics to reduce the space to two dimensions, which is easier to visualize. We computed clusters on the scaled topic space of authors via k-means clustering (we used $k = 5$ which seemed to balance resolution of salient clusters and consistency across years). If authors are more clustered in topic space, that reflects less connectivity between disciplines and suggests a multidisciplinary community. To measure the separability of clustering across years, we computed the ratio of the within-cluster sum of squares to the between-cluster sum of squares based on the k-means centroids. A higher ratio reflects greater dispersion within clusters compared to between clusters, indicating that the clusters are not very separated—in other words, that authors are less siloed in disciplinary enclaves, as would be the case in a more interdisciplinary field.

The central plot in Fig. 4 shows the ratio of the total within-cluster sum of squares to the between-cluster sum of squares for CogSci and VSS between 2000 and 2019. While

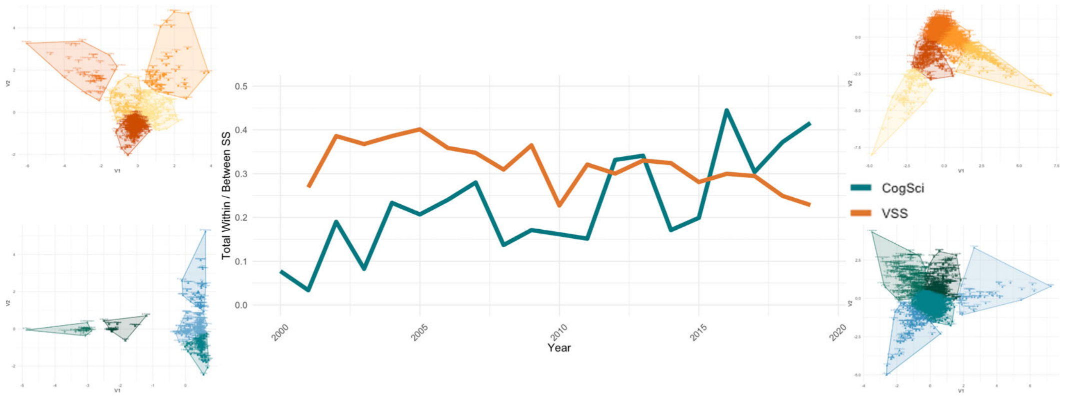

Fig. 4. k-means cluster analysis ($k = 5$) on topic space of authors, mapped onto two dimensions via multidimensional scaling. The cluster maps show the clustering of topics studied by authors in the earliest (left) and most recent year (right) for both CogSci (blue) and Vision Science Society (VSS; orange). The line graph shows the ratio of within- to between-cluster sums of squares for each year. CogSci is becoming less clustered over time.

VSS appears relatively stable (a regression on the data during this range is in fact negative: $\hat{\beta} = 0.006$, $p = .004$), the CogSci data have increased dramatically during this time ($\hat{\beta} = 0.014$, $p < .001$). Our results suggest that clusters in topic space have become less separable over time. The left and right sides of Fig. 4 are the author clusters for the earliest and most recent years of the CogSci and VSS data sets. The increase in topic overlap (decreased separability) in the set of CogSci authors is apparent in the two plots while topic consolidation in VSS does appear more nominal.

5. Combining topic space and network structure

In the previous sections, we argue that structural measures of collaboration and general trends in topic space are both useful in trying to quantify interdisciplinarity. However, interdisciplinarity is not only about community structure and topic distributions alone, but about the distribution of topics studied *within* the co-authorship structure. Here, we ask how topic similarity and prior collaboration structure combine to contribute to new and persisting collaborations between authors. In this way, we test the putative role that structural and topic space variables play in determining the overall collaboration landscape of the field. Concretely, we frame the contribution of topic similarity and co-authorship structure in the CogSci network as a link prediction problem: How do these variables contribute to the likelihood that two authors will publish a paper together in a given year? To the degree that both variables can be combined to produce high-fidelity predictions of new and ongoing collaborations, this represents a novel means of synthesizing research content and structure to predict the overall movement of the field. In addition, this validates the use of measures derived from topic space and network analysis to better describe the level of interdisciplinary work that authors are engaged in.

The first measure we use to predict co-authorship is the topic similarity between potential collaborators. Our earlier measures of separability in topic space suggest that author research topics offer insight into the level of interdisciplinary collaboration in cognitive science. Building on this insight, we investigate how the relative position of individual authors in topic space affects their probability of forming a new collaboration or maintaining an existing one. We measured similarity in topic space between two authors in a given year using their cosine similarity: the cosine of the angle θ between two authors' 100-topic vectors fitted by the topic model.

$$cos(\theta) = \frac{a \times b}{||a||||b||}. \tag{2}$$

The cosine similarity between each pair of authors was log-transformed to eliminate skewedness.

We fit a logistic regression to the co-authorships during each year with the topic similarity between authors from the previous year and whether the authors published together the previous year as predictors. We find that prior collaboration, topic similarity, and their interaction are all significant predictors of future collaborations between authors (*prior collaboration*: $\hat{\beta} = 4.592$, *topic similarity*: $\hat{\beta} = 2.248$, *interaction*: $\hat{\beta} = -3.862$, all ps < .001). Though the magnitude of the coefficients themselves might offer insight into the nature of collaboration in cognitive science, we refrain from interpreting the $\hat{\beta}$ values for the simple reason that confirming the independence of these predictors remains a challenge. Insofar as particular patterns of collaboration in the network are associated with corresponding patterns of topic similarity, the magnitudes assigned to these coefficients may not reflect the magnitude of their predictive power.

Instead, to ensure the strength of all the predictors in our model, we compare the full model described above—which predicts new collaborations on the basis of prior publication, topic similarity, and their interaction—to lesioned models with only prior publication and only main effects. Alignment of research topics across collaborators should be inevitable to some degree, assuming coherent and stable research interests. So, it is perhaps unsurprising that collaboration is strongly predicted by both prior collaboration and topic similarity. However, it is less clear whether both variables are necessary. Put another way, does topic similarity and its interaction with prior publication predict collaborations above and beyond having previously collaborated? The full model outperformed both lesioned models (topic similarity: *deviance* = 1,150, $p < .001$; interaction: *deviance* = 518, $p < .001$), suggesting that topic similarity and the interaction between topic similarity and prior publication improve predictions of novel collaborations.

5.1. Predicting 2020 collaborations

Using the regression with prior publication, topic similarity, and their interaction as predictors and training data from CogSci 2000 to 2019, we generate predictions about who co-authors together in CogSci 2020. A subset of these predictions is shown in

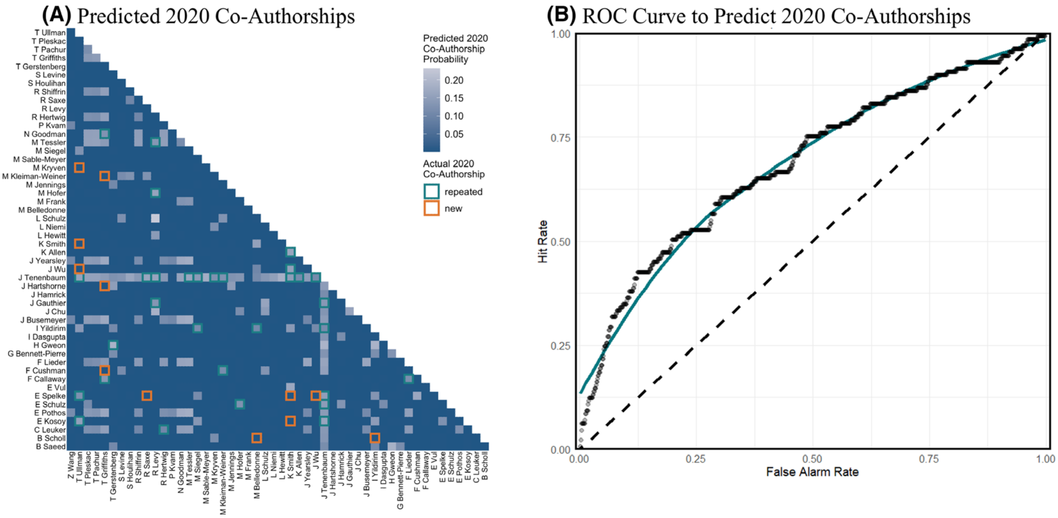

Fig. 5. (a) Prediction of co-authorships in 2020 for the 50 most eigencentral authors of 2019. The lighter the tile, the more likely our model predicts two authors will publish together. Highlighted tiles were co-authorships that indeed occurred in 2020, where tiles highlighted in teal were repeated collaborations from 2019 and tiles highlighted in orange were new collaborations. (b) Receiver operator characteristic (ROC) curve created by using different thresholds on the probability of new publication to make binary predictions. We evaluate only cases where authors did not publish together in the previous year. The dotted line shows where an ROC curve would fall for a model making predictions at chance.

Fig. 5a. To evaluate the model's effectiveness, we compare the model's predictions to holdout data: the full set of collaborations from CogSci 2020 (879 papers, 1,929 unique authors). Fig. 5b shows a receiver operator characteristic (ROC) curve (Swets, 1988) for the model's predictions. A model that predicts co-authorships might attain reasonably high accuracy simply by assuming that authors who previously collaborated together will do so again. For a more stringent test, we consider *new collaborations only* (when there was no prior collaboration the previous year): a prediction which is made based on only the authors' topic similarity. We use the area under the curve (AUC) to evaluate how well our model predicted new publications; an AUC of 0.5 indicates chance performance and an AUC of 1 indicates perfect classification accuracy. We found our model had an AUC = 0.689, which indicates our model is well above chance when making predictions about new collaborations. At the optimal threshold, the model's predictions have specificity (i.e., true negatives) of 0.802 and sensitivity (i.e., true positives) of 0.504. If we instead evaluate all co-authorships, including authors with and without prior collaboration, we obtain an AUC = 0.869, indicating that stability of collaboration networks plays an outsized role in publications. The ability to predict new collaborations from out-of-sample data based purely on topic similarity, and to predict all collaborations using a combination of topic similarity and prior collaboration, suggests that variables related to both authorship network structure and research topic space play a role in the ways collaborations form and persist. This bolsters the claim that questions related to interdisciplinary and multidisciplinary research, which are tied to collaborations both new and old in a given field, can be addressed using data-rich, bottom-up measures derived from co-authorship patterns and topics.

6. Discussion

It has been argued that science is becoming more interdisciplinary across a broad range of research areas (Porter & Rafols, 2009). However, a recent debate in the cognitive science community raises questions about whether the diverse fields that contribute to cognitive science pursue integrated research or are better described as multidisciplinary (Gray, 2019; Núñez et al., 2019). We argue that this discussion—and further investigations into the interdisciplinary nature of research more broadly—is strengthened by the use of formal, bottom-up measures of the collaboration structure and content within the field. Using the full text and author data from 19 years of published proceedings of the Cognitive Science Society, we analyze the evolution of the co-authorship network and assess changes in topic space year over year. Furthermore, we examine the distribution of topics within the co-authorship structure by querying how authors select their collaborators based on their interests and prior collaborations. Since these methods are novel in their application, we further validate their use by comparing the CogSci results to the full set of abstracts published in the VSS over a similar time period.

Our bottom-up approach yields converging support for the claim that cognitive science researchers have become more integrated over the past two decades. First, the co-authorship network shows that researchers published in the CogSci proceedings have become (structurally) less clustered and more interconnected, as evidenced by the decreasing transitivity of co-authorships and increasing maximum subgraph size. Second, co-authorship edge density, though more stable over time, is consistently higher for CogSci than VSS, suggesting that CogSci authors tend to publish with more unique authors. Third, beyond the structure of collaboration networks in CogSci, we find that the clustering of authors by topic within the CogSci proceedings has become less separable over time. This suggests that distinctions among disciplines may be shrinking. Finally, by combining co-authorship network and topic information, we find that prior collaboration and topic similarity are both significant predictors of collaboration in subsequent years; the significant interaction between them suggests that this is not a simple additive relationship. These variables allow us to predict new collaborations in out-of-sample data from CogSci 2020. Critically, this validates the use of measures derived from both the co-authorship network and the research topic space to characterize interdisciplinary collaboration. More broadly, it suggests that the combination of topic modeling and network analysis provides a window into the ongoing developments in a scientific field from one year to the next.

The use of topic modeling, characteristics of the co-authorship network, and the combination of the two offers a novel set of measures for understanding interdisciplinarity in a given field. The strength of these measures, aside from their formality, is the degree to which they are sensitive to the data in the research itself. Rather than pre-specifying the unique disciplines or fields within the community, we let graph clusters and topic separability speak to the connectedness of the research being done. This may allow for broader application across a range of other fields.

Critically, however, the measures we outline here are only part of the larger discussion about whether cognitive scientists are conducting interdisciplinary research. Key to understanding the progression of research in a field over time is not just how interconnected authors become or how creatively existing topics are combined, but how the reach of the network and the topics themselves evolve. Intuitively, interdisciplinarity is about both the *integration* of authors and topics over time, as well as maintaining or even increasing their *diversity* (e.g., Feng & Kirkley, 2020). A field that becomes more interconnected by barring certain subfields or methodologies can hardly be said to have accomplished the goal of interdisciplinary work. The measures we outline here provide a precise and nuanced picture of the integration of authors and topics in cognitive science, but they fall short of allowing us to draw strong inferences about the diversity of topics and disciplines represented over time. Indeed, the contrast between the current results and those of works like Núñez et al. (2019) may be in large part attributable to this distinction. Future work should explore ways that the data-driven approach we outline here might be expanded to reflect the goals of broad affiliation and diverse interests that interdisciplinary fields aspire to. The present results provide a step in this direction by showing how the tools of network analysis and machine learning can inform questions about the ways in which collaborations and research topics reflect meaningful integration.

Acknowledgments

We thank Rafael Núñez, Carson Miller Rigoli, Michael Allen, and Richard Gao for helpful discussion. We also thank Jamal Williams and Hayden Schill for their assistance with an earlier version of the analyses presented here. Finally, we thank the CogSci 2020 organizers for graciously sharing an early list of the authors published in the 2020 proceedings. This material is based upon work supported by a UCSD Research Grant to I.D., the NSF Graduate Research Fellowship under Grant No. DGE-1650112 to L.A.O. and UCSD Research Grant No. RG095178 to E.B.

Open Research badges

This article has earned Open Data and Open Materials badges. Data and materials are available at https://github.com/isabelladestefano/formalizing_interdisciplinary_collaboration.

Notes

1. All code used in this analysis can be found at: https://github.com/isabelladestefano/formalizing_interdisciplinary_collaboration.
2. (a) 2000–2014 papers are hosted at https://escholarship.org/uc/cognitivesciencesociety/, retrieved December 9, 2018; (b) 2010–2019 papers are hosted at https://cogsci.mindmodeling.org/, retrieved December 9, 2018, and CogSci 2019, retrieved December 3, 2019. Processed paper data are hosted at https://osf.io/qwzgd/.
3. All abstracts hosted at https://jov.arvojournals.org/, retrieved January 6–8, 2020.

References

Acedo, F. J., Barroso, C., Casanueva, C., & Galaán, J. L. (2006). Co-authorship in management and organizational studies: An empirical and network analysis. *Journal of Management Studies*, *43*(5), 957–983.

Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and Its Applications*, *311*(3–4), 590–614.

Bender, A. (2019). The value of diversity in cognitive science. *Topics in Cognitive Science*, *11*(4), 853–863.

Bergmann, T., Dale, R., Sattari, N., Heit, E., & Bhat, H. S. (2017). The interdisciplinarity of collaborations in cognitive science. *Cognitive Science*, *41*(5), 1412–1418.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, *3*(2), 57–65.

Cohen Priva, U., & Austerweil, J. L. (2015). Analyzing the history of *Cognition* using topic models. *Cognition*, *135*, 4–9.

Feng, S., & Kirkley, A. (2020). Mixing patterns in interdisciplinary co-authorship networks at multiple scales. *Scientific Reports*, *10*(1), 1–11.

Goldstone, R. L., & Leydesdorff, L. (2006). The import and export of *Cognitive Science*. *Cognitive Science*, *30*(6), 983–993.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, *71*(5), 1205–1222.

Gowanlock, M., & Gazan, R. (2013). Assessing researcher interdisciplinarity: A case study of the University of Hawaii NASA Astrobiology Institute. *Scientometrics*, *94*(1), 133–161.

Gray, W. D. (2019). Welcome to cognitive science: The once and future multidisciplinary society. *Topics in Cognitive Science*, *11*(4), 838–844.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(1), 5228–5235.

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(2), 404–409.

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(Suppl. 1), 5200–5205.

Nichols, L. G. (2014). A topic model approach to measuring interdisciplinarity at the national science foundation. *Scientometrics*, *100*(3), 741–754.

Núñez, R., Allen, M., Gao, R., Miller Rigoli, C., Relaford-Doyle, J., & Semenuks, A. (2019). What happened to cognitive science? *Nature Human Behaviour*, *3*(8), 782–791.

Núñez, R., Allen, M., Gao, R., Miller Rigoli, C., Relaford-Doyle, J., & Semenuks, A. (2020). For the sciences they are a-changin': A response to commentaries on Núñez et al's (2019) "what happened to cognitive science?" *Topics in Cognitive Science*, *12*(3), 790–803.

Parinov, S., & Kogalovsky, M. (2014). Semantic linkages in research information systems as a new data source for scientometric studies. *Scientometrics*, *98*(2), 927–943.

Porter, A. L., Cohen, A. S., Roessner, D., & Perreault, M. (2007). Measuring researcher interdisciplinarity. *Scientometrics*, *72*(1), 117–147.

Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, *81*(3), 719–745.

Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, *82*(2), 263–287.

Ravikumar, S., Agrahari, A., & Singh, S. N. (2015). Mapping the intellectual structure of scientometrics: A co-word analysis of the journal scientometrics (2005–2010). *Scientometrics*, *102*(1), 929–955.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2014). Stm: R package for structural topic models. *Journal of Statistical Software*, *10*(2), 1–40.

Rothe, A., Rich, A. S., & Zhi-Wei, L. (2018). Topics and trends in cognitive science (2000–2017). In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 979–984). Austin, TX: Cognitive Science Society.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1: Foundations*. Cambridge, MA: MIT Press.

Schunn, C. D., Crowley, K., & Okada, T. (1998). The growth of multidisciplinarity in the cognitive science society. *Cognitive Science*, *22*(1), 107–130.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*(4857), 1285–1293.

Wagner, C. S., & Leyesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy*, *34*(10), 1608–1618.

Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., Rafols, I., & Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*, *165*, 14–26.