The Landscape of Neuroscience

Mario Senden
Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616 Maastricht, 6200 MD, Limburg, The Netherlands

Abstract

Neuroscience emerged as a distinct academic discipline during the 20th century and has undergone rapid expansion and diversification. This study leverages text-embedding and clustering techniques together with large language models to analyze 461,316 articles published between 1999 and 2023 to provide a snapshot of neuroscience's landscape. Inter-cluster citation analysis uncovers a surprisingly integrated picture. An analysis of how research clusters align with pre-defined dimensions demonstrates a strong experimental focus, widespread reliance on specific mechanistic explanations rather than unifying theoretical frameworks, and a growing shift towards applied research.

Keywords: research trends; text embedding analysis; scientific mapping; large language models (LLMs); topic modeling

Introduction

Modern neuroscience has witnessed dramatic expansion and diversification since its inception in the 20th century. While this is a natural tendency of scientific disciplines (Andersen, 2016), it can hide the interconnectedness of phenomena and research questions and may thus hinder further progress (Popper, 1963, 1985). Consequently, there is a need to provide a high-level perspective on the evolving landscape of neuroscience that may help to integrate the field by identifying unifying principles and methodological approaches as well as structural limitations that need to be overcome. To address this, I have compiled an extensive dataset of neuroscientific abstracts and their metadata and performed a content-driven examination of the field's trajectory over a 25 year period.

Results

Neuroscientific Research Domains

I identified neuroscience journals ranked in the top two quartiles in each year from 1999 to 2023 according to the SCIMago Journal Rank. I supplemented these neuroscientific journals with Q1 multidisciplinary journals that publish neuroscientific research such as Nature, Science and Plos ONE. The next step involved querying PubMed for a maximum of 5,000 articles per year for each selected journal for metadata and abstracts of research and review articles. After removing non-neuroscientific articles, the dataset comprised 461,316 articles published in 375 journals.

To identify the distinct research domains that divide neuroscientific work, I clustered abstracts based on their semantic similarity measured as the cosine similarity between their domain specific text embeddings. To achieve this, I first embedded abstracts into domain-specific latent space. I then constructed a semantic graph wherein each abstract is connected to its 50 semantically nearest neighbors, with each link weighted by the cosine similarity between its vertices. I then applied the Leiden community detection algorithm (Traag, Waltman, & van Eck, 2019) on this graph to obtain clusters. For each cluster, I submitted abstracts of the 200 nearest neighbors to the cluster's centroid to a gpt-4o large language model (LLM) from OpenAI to describe the cluster.

Clustering applied to embedded abstracts identified 175 unique clusters ranging in size from 9,155 (cluster 0) to 117 (cluster 174) articles (see Supplementary Figure 1a). The largest cluster involves research on the mechanisms of neuropathic pain. The smallest cluster is concerned with the effects of electromagnetic fields emitted by mobile devices on brain function. While most clusters are dominated by research articles, intriguingly two clusters contain more review articles. Cluster 171, which investigates the pathophysiology and potential long-term effects of SARS-CoV-2 on the nervous system, and cluster 173, which investigates the role of exosomes in neurodegenerative disease. Cluster 171 also presents the highest median citation rate for both research (8.1 citations per year) and review (11.1 citations per year) articles, (Supplementary Figure 1b) likely explained by the immense interest in SARS-CoV-2 during the global pandemic (Mondal et al., 2023; Samim et al., 2024; Schor, Cudkowicz, & Banwell, 2023). Several clusters exhibit some degree of thematic overlap. A total of 14 clusters are devoted to Alzheimer's disease (AD), each investigating distinct but complementary aspects. For example, cluster 1 investigates the role of amyloid beta peptides in the pathogenesis of AD whereas Cluster 47 focuses on the pathophysiological mechanisms involving tau protein modifications. This pattern is not unique to AD. There are nine clusters devoted to Parkinson's, indicating that neurodegenerative diseases form their own group of clusters. Apart from clusters devoted to specific conditions, I also observed modality (e.g., vision and audition), cognitive/behavioral (e.g., decision-making, language, and memory), and methodological (functional neuroimaging, brain stimulation) groups of clusters. My examination did not reveal any theory-specific clusters.

I followed this up with a systematic investigation of the underlying dimensions that characterize neuroscientific research. To that end, I defined 10 dimensions (Appliedness, Methodological Approach, Species, Spatial Scale, Temporal Scale, Modality, Cognitive Complexity, Theory Engagement, Theory Scope, and Interdisciplinarity) and submitted abstracts of the 250 nearest neighbors to a cluster's centroid to the LLM to characterize each cluster along these dimensions. The descriptions generated by the LLM where then in a second step submitted again to the LLM to judge whether a particular cluster qualifies for categories that I defined for each dimension such as ”human”, ”non-human primates”, and ”rodents” for the Species dimension. Note that the categories that characterize a dimension are not mutually exclusive. A quantitative overview of how many clusters qualify for categories is shown in Supplementary Figure 1c and reveals a predominantly experimental focus (96% of clusters) that employs both hypothesis-driven (77%) and data-driven (67%) research. Neuroscientific research often contains a theoretical element in the sense that it employs computational modeling (46%). This aligns with the observation that neuroscience tends to employ micro-theories (64%), i.e., narrowly scoped mechanistic accounts. Theories of intermediate scope in the form of domain-specific (35%) and disease-specific (39%) theories are also prevalent. However, only 17% of clusters contain research that employs overarching theoretical frameworks aiming to explain fundamental principles of brain function. I observed that translational (55%) and clinical (30%) work sit atop a broad base of fundamental science (81%). Many clusters contain work on rodents (74%) or humans (71%), though non-human primates (33%) and other mammals (30%) also feature prominently. Neuroscience generally shows a balanced division between spatiotemporal scales, though work at the microsecond scale is sparse (6%).

I next sought to understand the extent to which categories co-occur by examining the Matthews correlation (phi) coefficient between pairs of categories. Both molecular and cellular spatial scales are negatively correlated with regional scale ($\phi = -0.4048$, $t(173) = -5.822$, $p \ll 0.0001$ and $\phi = -0.3631$, $t(173) = -5.126$, $p = 0.00148$, respectively). The cellular scale also exhibits a negative correlation with the whole-brain scale ($\phi = -0.3773$, $t(173) = -5.359$, $p = 0.000497$). For temporal scales, there is often pairwise but never full integration. Fundamental research is significantly anticorrelated with clinical ($\phi = -0.5496$, $t(173) = -8.6535$, $p \ll 0.0001$), translational ($\phi = -0.3441$, $t(173) = -4.8206$, $p = 0.00587$), and method development ($\phi = -0.4553$, $t(173) = -6.7260$, $p \ll 0.0001$) approaches. For theoretical scope, there is a negative correlation between micro theories and overarching theoretical frameworks ($\phi = -0.3752$, $t(173) = -5.3234$, $p = 0.000588$).

To examine how clusters interact with each other, I examined their citation structure. To that end, I computed the Krackhardt coefficient (Krackhardt & Stern, 1988) for each cluster's incoming links (citations its articles receive from external articles) and outgoing links (references outside the cluster). The Krackhardt coefficient provides a measure of the outward (positive) versus inward (negative) focus of a cluster. Notably, 74.86% of clusters exhibit positive Krackhardt coefficients for both citations and references. These clusters lean on research from other clusters (indicated by a positive Krackhard coefficient for their references) but also provide insights for other clusters (indicated by a positive Krackhardt coefficient for their citations). By contrast, 6.86% of clusters contain articles that frequently cite and are cited internally, indicated by negative Krackhardt coefficients for both their reference and citation patterns. The majority of neuroscientific domains are thus well-integrated and share knowledge.

My last goal was to identify trends in neuroscience. At the individual cluster level, I examined growth trends in terms of the size-adjusted annual growth rate, which quantifies the yearly increase in article count for each cluster relative to its total number of articles. A majority (52.0%) of clusters have increased their output above what is expected based on increases in output of the entire discipline. One third of clusters exhibit stable output as they neither decline nor exceed the growth of the discipline. Finally, 15.4% of clusters exhibit a decline in their output. Notably, the SARS-CoV-2 cluster (171) is among the ten fastest growing clusters. Generally, it appears that growing clusters share a strong applied focus and target overarching themes such as neurodegeneration (clusters 69 and 141), neuromodulation (clusters 167 and 168), and technological advancements in neuroscience (clusters 94 and 120). By contrast, declining clusters reflect predominantly fundamental research with a focus on receptor dynamics (clusters 9, 55, 60, 75, 89, 101, 124, and 135) and signaling pathways (clusters 48 and 102). These extremes of the spectrum reflect larger trends across clusters. While clusters containing fundamental research together are growing at a compound annual growth rate of 1.89%, this is less than the growth exhibited by neuroscience globally (2.39%). By contrast, clusters involving translational research, clinical research, method development, and technological exploitation exhibit compound annual growth rates of 3.57%, 4.78%, 7.51%, and 7.80%, respectively. See Supplementary Figure 1d,e for the publication trajectories of the ten most growing and most declining clusters. A qualitative analysis performed by an LLM on recently published articles extracted trends for every cluster. Aggregating these into shared themes across clusters further confimed that trends are primarily driven by applied research concerns.

Discussion

Neuroscience appears to be thriving as it maintains high output across diverse topics ranging from neurodegenerative diseases, neuromodulation, cognitive functions, to technological advances. Despite this diversity, neuroscience achieves a high level of integration and extensive knowledge exchange across its domains. The field currently exhibits a good balance between hypothesis-driven and data-driven approaches and between fundamental and applied research. However, growth trends show that fundamental research is losing ground. Neuroscience spans all levels of organization, from molecular and cellular studies to whole-brain dynamics. However, integration across spatiotemporal scales remains limited. There is, for instance, a clear divide between small and large spatial scales. A particularly concerning revelation of this study is the field's predominant reliance on highly specific micro theories rather than broader theoretical frameworks. While computational models are widely used to test mechanistic hypotheses, theoretical work that develops and refines overarching frameworks is notably scarce. The findings of this study highlight that ensuring that fundamental research remains valued, greater integration across spatiotemporal scales, and increased theoretical synthesis could strengthen the field.

References

Andersen, H. (2016). Collaboration, interdisciplinarity, and the epistemology of contemporary science. *Studies in History and Philosophy of Science Part A*, *56*, 1–10.

Krackhardt, D., & Stern, R. N. (1988). Informal networks and organizational crises: An experimental simulation. *Soc Psychol Q*, 123–140.

Mondal, S., et al. (2023). Covid-19 pandemic and neurosciences in india–the coin study: Impact on neuroscience research and early career researchers. *J Neurol Sci*, *455*.

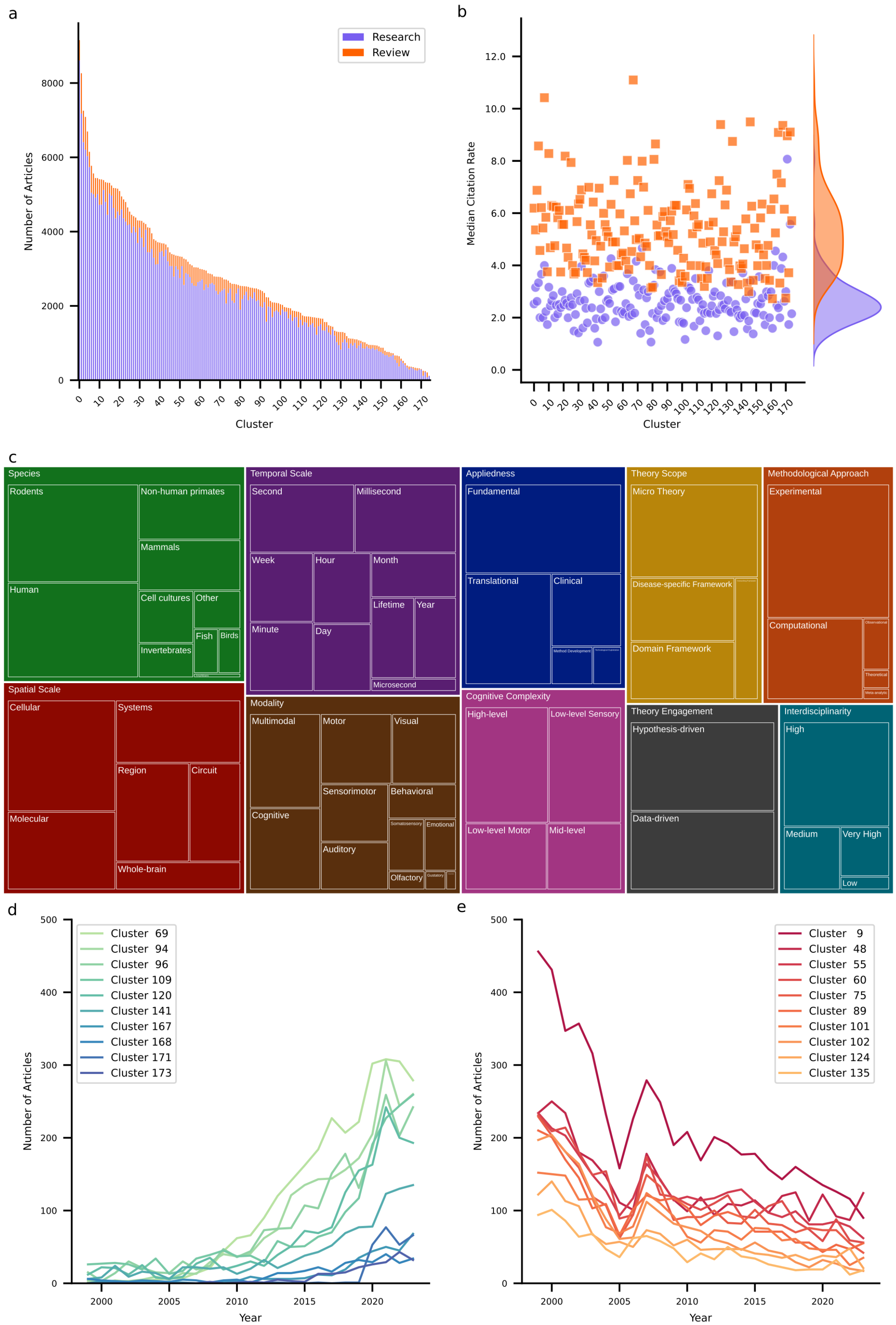
Popper, K. (1963). *Science: Conjectures and refutations*. London and New York.

Popper, K. (1985). *Realism and the aim of science*. Routledge. doi: 10.4324/9780203713969

Samim, M. M., et al. (2024). Covid-19 pandemic & neurosciences in india– the coinstudy: Impact of the pandemic on research related to the neurosciences. *Indian Journal of Medical Research*, *159*, 557.

Schor, N. F., Cudkowicz, M. E., & Banwell, B. (2023). Academic neurology and the covid-19 pandemic. *Neurology*, *100*, 430–436.

Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From louvain to leiden: guaranteeing well-connected communities. *Sci Rep*, *9*, 5233.



Supplementary Figure 1: **a**, Number of articles within each cluster with review articles (orange) stacked on top of research articles (purple). **b**, Median citation rates of research and review articles per cluster. Distributions of MCRs are superimposed on the right. **c**, Tree map visualizing number of clusters qualifying for a given category within each dimension. Size of a region reflects the number of clusters that contain work exhibiting the category. Note that categories are not exclusive, and the same cluster may count towards several categories even within the same dimension. **d** Number of articles in the dataset for the ten most growing clusters broken down by year. **e** Number of articles in the dataset for the ten most declining clusters broken down by year.