Preprints.org

Article

Topic Modeling as a Tool to identify Research Diversity: A Study Across Dental Disciplines

Maria Teresa Colangelo , Stefano Guizzardi , Carlo Galli [*]

Article

Topic Modeling as a Tool to Identify Research Diversity: A Study Across Dental Disciplines

Maria Teresa Colangelo, Stefano Guizzardi and Carlo Galli *

Histology and Embryology Laboratory, Department of Medicine and Surgery, University of Parma, Via Volturno 39, 43126 Parma, Italy; mariateresa.colangelo@unipr.it (M.T.C.); stefano.guizzardi@unipr.it (S.G.)

* Correspondence: carlo.galli@unipr.it

Abstract: This study investigates the diversity and evolution of research topics within the dental sciences from 1994 to 2023 using topic modeling and Shannon's entropy as a measure of research diversity. We analyzed a dataset of 412036 scientific articles across six dental disciplines: Orthodontics, Prosthodontics, Periodontics, Implant Dentistry, Oral Surgery, and Restorative Dentistry. This research relies on BERTopic to identify distinct topics within each field. The study revealed significant shifts in research focus over time, with some disciplines exhibiting robust growth in article numbers such as Periodontics and Prosthodontics. The application of Shannon's entropy revealed an increasing diversification of research efforts in disciplines such as Restorative Dentistry, while others, like Prosthodontics, in spite of their size and their high number of research topics, maintain a more specialized research focus. Taken together, our findings describe the dynamic nature of dental research and highlight the balance shifts in research focus across several key areas of Dentistry.

Keywords: research diversity; shannon's entropy; topic modeling

1. Introduction

As research progresses and scholars explore new research lines and publish their most recent results, the range of topics in a science field typically broadens. New discoveries open up new questions that can be explored, often generating novel research directions. Understanding the dynamics and diversity of research topics in a discipline is, potentially, a relevant indicator of how active the area is [1]. For instance, a lower degree of diversity in a scientific field may indicate the presence of one or more prominent topics, which most of the publications focus on; this in turn could be driven by various factors, both intrinsic to the field (such as a compelling scientific conundrum that attracts the attention of many scholars of the community) or external (e.g., socioeconomical incentives) [2,3].

But understanding research diversity entails understanding its epistemic composition, it requires to analyze the different scientific areas that it consists of. To this effect, Topic modeling can be a precious resource to unpack the main lines along which research evolves and is conducted [4]. Topic modeling algorithms are convenient tool to get a sense of the array of topics that are present in a field [5]. These algorithms, such as LDA or, more recently, BERTopic, are capable of scanning large datasets, and segment a whole corpus of documents, according to their semantic content. For the present work we used BERTopic, an algorithm centered on BERT (Bidirectional Encoder Representations from Transformers) embeddings, i.e., numerical representation of the semantics of a sentence or even of a whole document [6]. BERT is a recent instrument developed by Google, which is based on a mechanism known as attention, and has quickly outperformed earlier embedding algorithms in various tasks [7].

To better investigate how research efforts are distributed across research topics, we decided to resort to Shannon's entropy, a concept derived from information theory, and a measure of uncertainty or randomness in a dataset [8]. It was originally formulated by Claude Shannon in 1948,

and it has since been adapted for various domains. Although entropy measurements have been applied to investigate the performance of topic modeling algorithms [9–11], we propose to use Shannon's entropy to quantify the research diversity by assessing the distribution of research topics within a given field as determined by BERTopic. The mere number of topics in a field may not be an accurate measurement of diversity, when taken alone. A field where there is e.g., one predominant topic that absorbs 80% of the global research efforts alongside 19 minor niche research areas is in a very different situation than a field where there are 20 equally relevant research areas. To account for the distribution of the research efforts (measured through their output, i.e., their resulting publications), entropy may be helpful. Higher entropy values indicate a more even distribution of topics, reflecting greater diversity, while lower entropy values suggest that research is concentrated around few dominant topics.

Despite its potential, the use of Shannon's entropy to measure research diversity remains underexplored in many disciplines, including dental sciences. Dental research includes a wide array of specialties, each with its own unique focus and themes, which have undergone a tremendous growth in the last decades of the 20th century [12]. The introduction of groundbreaking techniques (such as implant dentistry) has also opened up new therapy opportunities, challenges and questions [13]. Understanding how research topics evolve and diversify within these specialties could help guide future research efforts and policy decisions.

In this study, we applied Shannon's entropy to analyze the diversity of research topics across multiple dental disciplines over a thirty-year period (1994-2023). Using BERTopic, we extracted and categorized research topics from a comprehensive dataset of scientific articles from MEDLINE. By calculating the entropy of these topics over time, we believe we could provide a quantitative measure of research diversity within each discipline and offer insights into the evolution and distribution of research themes in the field of dental research.

2. Materials and Methods

2.1. Data Collection

Data were collected and analyzed with Google Colab Pro notebooks powered by Python 3.10.12 [14] and running on T4 GPUs [15]. The corpora we used for the investigation were generated with the Biopython library [16] through a query-driven exploration of MEDLINE facilitated by the Entrez.esearch function. The disciplines included were Orthodontics, Prosthodontics, Periodontics, Implant Dentistry, Oral Surgery, and Restorative Dentistry, based on the authors' domain knowledge. For each discipline, relevant scientific articles were retrieved from PubMed using a series of discipline-specific search queries. The search terms were designed to capture a broad range of publications within each field, utilizing both Medical Subject Headings (MeSH) and title/abstract keywords to ensure comprehensive coverage, as follows:

- Orthodontics: "Orthodontics"[MeSH] OR "Orthodontics"[Title/Abstract] OR "Orthodontic Treatment"[Title/Abstract] OR "Orthodontic Appliances"[MeSH] OR "Orthodontic Brackets"[MeSH]) OR ("Malocclusion"[MeSH] OR "Teeth Misalignment"[Title/Abstract]
- Prosthodontics: "Dental Prosthesis"[MeSH] OR "dental prostheses"[Title/Abstract] OR "dental prosthesis"[Title/Abstract] OR "Prosthodontics"[MeSH] OR "Prosthodontics"[Title/Abstract]
- Periodontics: "Periodontics"[MeSH] OR "Periodontal"[Title/Abstract] OR "Periodontics"[Title/Abstract] OR "Periodontology"[Title/Abstract]) OR ("Periodontal Diseases"[MeSH] OR "Periodontitis"[MeSH] OR "Gingivitis"[MeSH] OR "Periodontal Pocket"[MeSH] OR "Gum Disease"[Title/Abstract]
- Implant Dentistry: "Dental Implants"[MeSH] OR "Dental Implantation"[MeSH] OR "Dental Implant*"[Title/Abstract] OR "Implant Dentistry"[Title/Abstract] OR "Implantology"[Title/Abstract]
- Oral Surgery: "Oral Surgical Procedures"[MeSH] OR "Oral Surgery"[Title/Abstract] OR "Oral Surgeons"[Title/Abstract] OR "Maxillofacial Surgery"[Title/Abstract] OR "Oral and Maxillofacial Surgery"[Title/Abstract]

- Restorative Dentistry: "Restorative Dentistry"[Title/Abstract] OR "Tooth Filling"[Title/Abstract] OR "Dental Restoration"[Title/Abstract] OR "Restorative Treatments"[Title/Abstract] OR "Dental Caries"[MeSH] OR "Tooth Cavity"[Title/Abstract] OR "Dental Cavities"[Title/Abstract] OR "Tooth Decay"[Title/Abstract]

The searches were limited to articles published between 1994 and 2023 to try and capture the big scientific developments that occurred in the field at the end of the 20th century and have enough articles for extensive analysis.

2.2. Data Cleaning and Preprocessing

The data retrieved from MEDLINE were formatted into *pandas* dataframes [17], which contained the relevant bibliographic details for the analysis, i.e., author names, article titles, publication years, abstracts, publication type, and journal names. The datasets underwent several cleaning steps:

- Deduplication: Duplicate entries were identified and removed based on article titles.
- Missing Data: Articles without titles were excluded.
- Standardization: The publication years were standardized to four-digit integers. Articles published before 1994 or after 2023 were excluded from further analysis.

2.3. Topic Modeling

2.3.1. Sentence Embedding and Topic Modeling

To identify and analyze research topics across disciplines, we employed BERTopic, a state-of-the-art topic modeling technique [18]. BERTopic leverages sentence embeddings [19], UMAP (Uniform Manifold Approximation and Projection) for dimensionality reduction [20], and HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) for clustering [21]. We decided to use the "*all-MiniLM-L6-v2*" model from SentenceTransformers, which balances speed and accuracy, making it suitable for generating sentence embeddings in large-scale text analysis [22].

2.3.2. Topic Extraction

The titles from the articles of each discipline's dataset were processed using BERTopic. The UMAP model reduced the dimensionality of the embeddings for better clustering, and HDBSCAN was used to cluster these reduced embeddings into distinct topics. We used the following parameters:

- UMAP metric: cosine distance;
- size of the neighborhood: 50;
- number of components: 5;
- HDBSCAN clustering metric: Euclidean;
- cluster_selection_epsilon=0.5
- minimum cluster size: 50.

Those articles that did not fit well into any specific topic were labelled as noise topics (Topic -1) and were excluded from further analysis. To get a better representation of the topics, we relied on the integration between BERTopic and Large Language Models and used OpenAI's GPT 3.5 turbo to generate labels for the topics [23].

2.4. Shannon's Entropy Analysis

Shannon's entropy H(X) was used as a measure of research diversity over time within each dental discipline [8], and it was calculated using the following formula:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i)$$

where $p(x_i)$ is the proportion of articles assigned to topic *i* in a particular year. Higher entropy values indicate a more diverse distribution of topics, while lower values suggest that research focus was concentrated on fewer topics. The calculated entropy values were plotted over time to visualize how

the diversity of research topics evolved within each discipline using the matplotlib [24] and seaborn libraries [25].

3. Results

3.1. General Characteristics of the Dataset

The dataset generated for the present study comprises a total of 412036 scientific articles published between 1994 and 2023 across six dental disciplines. The distribution of articles across the 6 disciplines is summarized in Table 1. The number of articles per discipline varies significantly, reflecting differing levels of research activity within each field.

Table 1. Number of Articles per Discipline (1994-2023).

	Discipline	Number of Articles
1	Orthodontics	51872
2	Prosthodontics	98852
3	Periodontics	93510
4	Implant Dentistry	42826
5	Oral Surgery	61719
6	Restorative Dentistry	63257

Taken together these fields possess, as expected, a vast research output, which reflects their central importance in dental practice and research, with noticeable differences, though. Prosthodontics emerged as the most prolific discipline, with 98852 articles, followed closely by Periodontics, with 93510. The remaining datasets were quite smaller, with restorative Dentistry and Oral Surgery a little above 60000 articles.

This disparity may suggest that these fields are more specialized sub-fields of other disciplines (such as might be the case with Implant Dentistry, which can be often conceived as a specific application of Oral Surgery) or that they have seen less research activity relative to the broader dental sciences, such as Orthodontics, because of their niche character. It must be remembered that we generated the datasets independently for each discipline, and did not exclude articles that were present in more than one disciplines. It can be assumed that certain topics can be ascribed to more than one discipline and to investigate this assumption we measured the overlap between datasets (Figure 1).
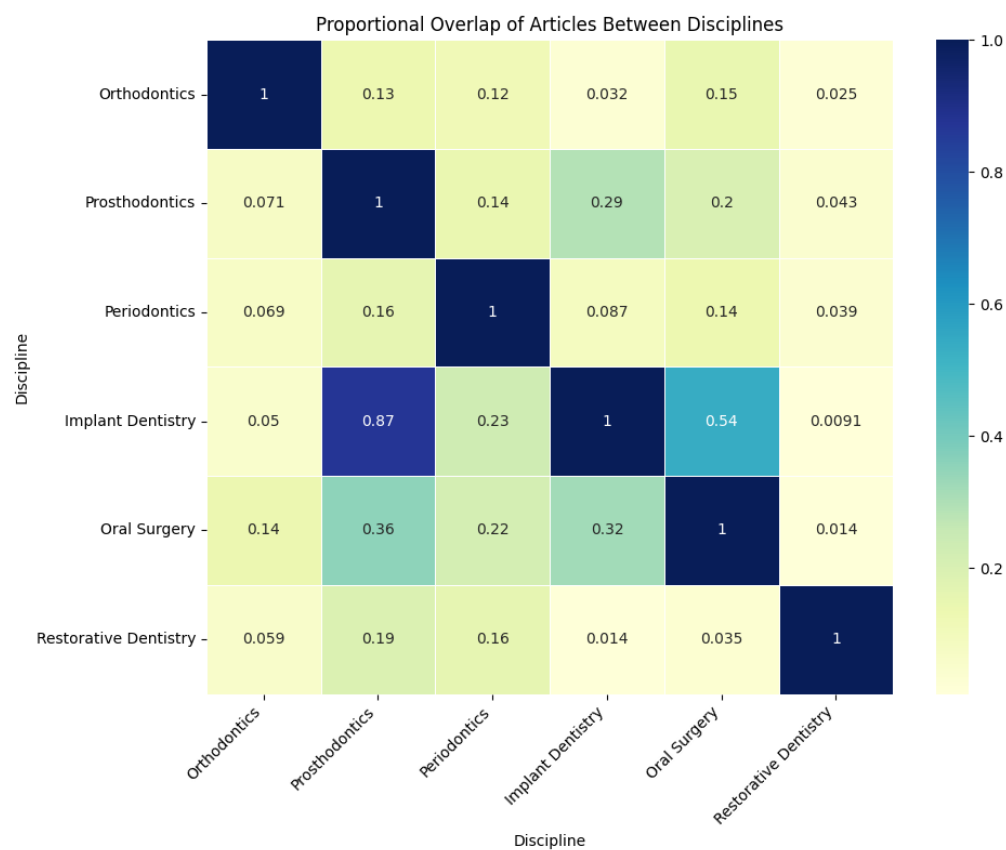
Figure 1. This heatmap visualizes the proportional overlap of articles between various dental disciplines.

In most cases the datasets displayed little overlap, with the noticeable exception of Implant Dentistry and Prosthodontics (87%), in which case the Implant Dentistry dataset can be considered a specialized sub-set of Prosthodontics, and a 54% overlap between Implant Dentistry and Oral Surgery, which again is not completely surprising considering the nature of the clinical interventions that are involved with implant insertion.

3.2. Distribution of Articles Over Time

The diachronic distribution of articles may provide further insights into the evolution of research focus within dental sciences. Figure 2 illustrates the number of articles published per year within each discipline from 1994 to 2023.
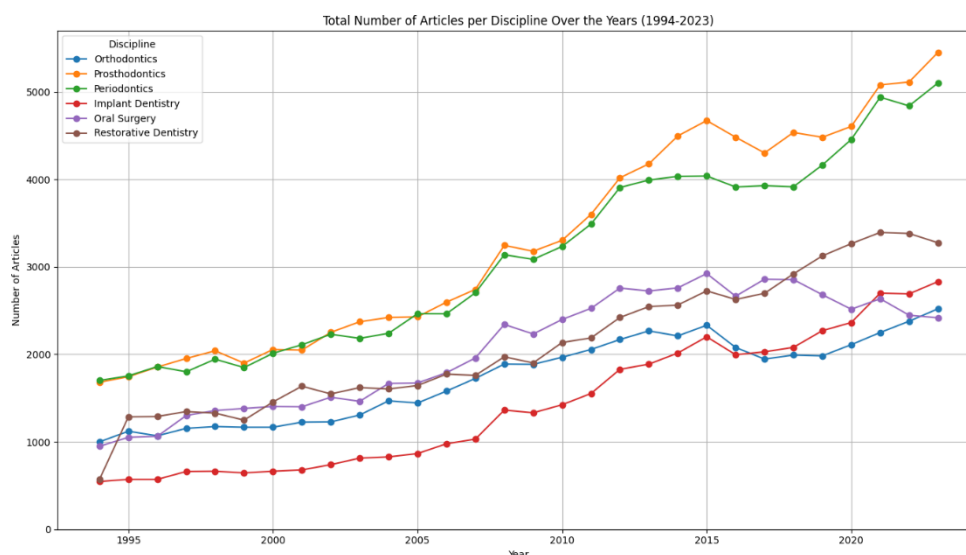
Figure 2. Distribution of Articles Over Time by Discipline (1994-2023).

From the data, it can be observed that there has been a general increase in the number of publications in every discipline over the study period, and that the number of new articles per year has increased for most disciplines, reflecting the growing body of research and the expanding scope of the dental sciences, a phenomenon that is widely seen across various disciplines, not only dental-related ones [26]. The growth, however, differed visibly across disciplines.

Prosthodontics and Periodontics grew with an -almost exponentially - increasing rate, at least until about 2015, when they both plateaued at about 4000 new articles/year, to start growing again only after a few years and recently exceed 5000 new publications per year (Figure 1). This suggests an increasing research focus on these areas, possibly driven by advancements in materials science, technology, or clinical techniques.

Other disciplines did not experience such a fast increase in the number of publications; in particular, Oral Surgery appears to have peaked around the mid 2010s, with about 3000 new articles/year and the number of new publications has remained consistent or has even slightly dropped since (Figure 2).

3.3. Identification of Research Topics

We then identified distinct research topics within each of the six dental disciplines with BERTopic. This topic-modeling process, which analyzed the titles of a little more than 400000 articles. Although abstracts could have been used to get an even more accurate overview of the themes of the different disciplines, we preferred to limit the analysis to titles. Our choice was supported by previous experience that showed us that titles can be sufficiently accurate in portraiting main area of a manuscript [22] and as they are considerably shorter than abstracts, this proved very advantageous to process hundreds of thouands of documents, as this operation can be very resource-consuming in such large datasets. BERTopic revealed a wide range of themes within each field, reflecting diverse research areas in these dental disciplines.

The number of distinct topics identified varied across the disciplines (Table 2), with restorative Dentistry having the highest number of topics (n=59) and Orthodontics and Prosthodontics both trailing behind with just 32 topics. The smallest dataset, Implant Dentistry, had also, not unexpectedly, the smallest number of topics, with just 22 topics identified. This could possibly reflect differences in the complexity and breadth of research within each field, but it must be remembered that the absolute number of topics in a dataset of documents heavily depends on the structure of the data and on the parameters used for topic modeling, such as the minimum cluster sized used in the HDBSCAN algorithm and the number of neighbors used by the UMAP dimensionality reduction process, i.e., how granular our analyzes intended to be, or additional parameters such as the cluster

selection epsilon, which merges topics that are similay beyond a set threshold. We tested a wide range of parameters through extensive grid searches and arbitrarily chose the algorithm's parameters that avoided hyperinflating the number of topics. Minimum cluster size=50 and number of neighbors=25 were chosen because they did not fragment the dataset across an excessive number of topics, and these parameters were used for all our analyses.

Table 2. Number of Topics per Discipline.

	Discipline	Number of Topics
1	Orthodontics	32
2	Prosthodontics	32
3	Periodontics	49
4	Implant Dentistry	22
5	Oral Surgery	34
6	Restorative Dentistry	58

The investigation furthermore identified several key topics that dominated specific disciplines, and can be found in Table 3. For instance, in Restorative Dentistry, the topic with the highest number of articles, indicated by BERTopic as topic 0, the main theme so to speak, is "Childhood Caries Prevention Study", while dental implants are a relevant topic in Implant Dentistry but also, consistently with Figure 1, in Prosthodontics and Oral Surgery.

Table 3. Prominent Topics per Discipline.

	Discipline	Main Topic
1	Orthodontics	Orthodontic Treatment Evaluation Study
2	Prosthodontics	Dental Implant Clinical Study
3	Periodontics	Periodontal Disease Treatment Studies
4	Implant Dentistry	Dental Implant Management Insights
5	Oral Surgery	Dental Implant Surgical Evaluation
6	Restorative Dentistry	Childhood Caries Prevention Study

To better understand, however, how these research fields evolved in time, we needed a diachronic analysis of research topics. Topics are not, in fact, consistently represented across the years; some topics emerge later in time or subside, as the interest of the research community shifts elsewhere, e.g., because of changes in clinical priorities or the resolution of key research questions [4].

*3.4. Diachronic Analysis of Topics*

The number of topics across the disciplines we included in our dataset tended to increase over time in the 30-year period, although not homogeneously (Figure 3).
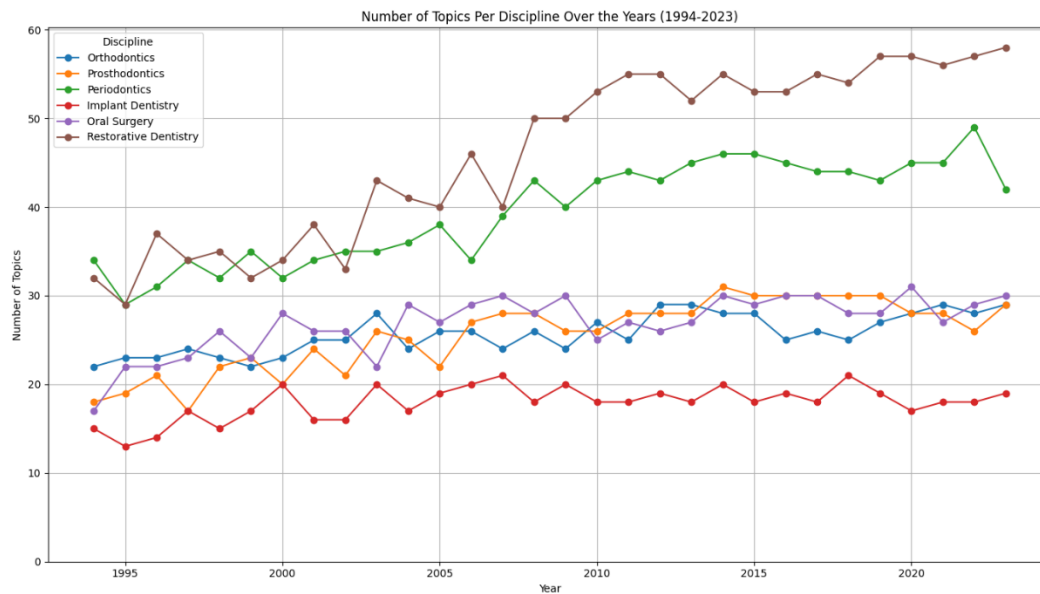
Figure 3. Line plots showing the number of topics identified over time for each discipline.

Generally speaking, the more numerous the articles in a dataset, the more numerous the topics (Figure A1) and thus Implant Dentistry always possessed fewer topics than the remaining disciplines and maintained a fairly consistent number of them over the years (Figure 3). However, Restorative Dentistry and Periodontics, which is about 50% larger than the previous one, had similar numbers of topics until the early 2000s, around 35, when the number of topics in Restorative Dentistry started to grow at a faster rate and exeeded the number of topics in the Periodontics dataset, creating a gulf between the two that has remained stable. Topics in the remaining disciplines increased at a slower pace, from about 20 topics in the 90s to the current levels of 30 topics (Figure 3). The number of topics can be partially considered a measure of diversity in a research field, but it does not reveal much about how actively researched these topics are. To quantitatively assess how the attention of researchers is partiotoned across topics within each discipline over time, we calculated Shannon's entropy based on the distribution of the topics identified by the BERTopic model. As previously mentioned, Shannon's entropy is a measure of uncertainty or diversity; in this context, it quantifies how evenly research efforts are distributed across different topics within a discipline. The analysis of entropy over time, as shown in Figure 4, highlights the changes in topic diversity within each discipline from 1994 to 2023.
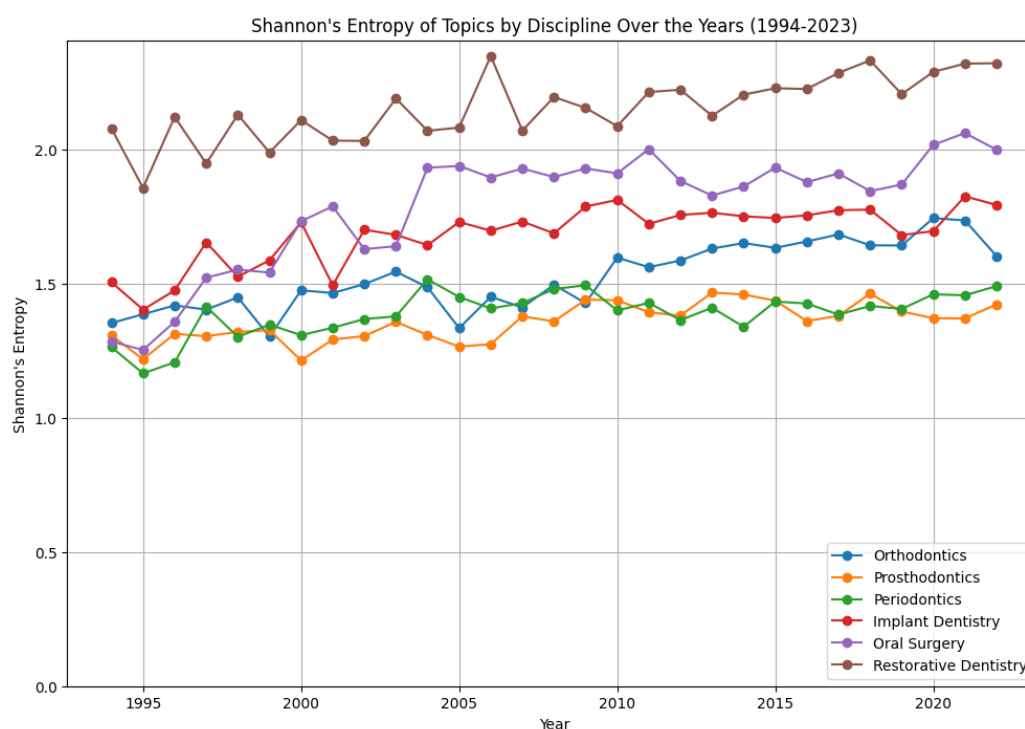
Figure 4. Line plots showing the the entropy level identified over time for each discipline.

This figure reveals that for most disciplines entropy has remained fairly consistent over the study period, especially for disciplines such as Periodontics and Prosthodontics (Figure 4). This is consistent with a stable number of topics in Prosthodontics; however, vis a vis the increase in topics in Periodontics, this observation can be interpreted as a sign of the dominance of few major topics that have attracted the most part of the new publications. Moreover, it should also be noted that Periodontics, which has a very high number of topics has also the lowest level of Entropy, together with Prosthodontics, confirming the idea that Periodontics revolves around a main core of Topics. One main exception to the overall trend in entropy can be noted: a robust increase in entropy for Oral Surgery at least until the early 2000s, which could be interpreted as an increase in diversification of research by actively pursuing a wide array of topics.

4. Discussion

The results of this study provide interesting insights into the diversity and evolution of research within dental sciences over the past three decades. By employing topic modeling techniques and shannon's entropy, we quantified and compared the diversity of research topics across six distinct dental disciplines.

The dataset was arbitrarily generated through Pubmed searches to collect a sizable corpus of publications in six core areas of dental practice and research. The queries were generic enough to embrace a large portion of publications in each discipline, without getting into details of specific clinical or scientific issues within each discipline. Although other areas could have been – and they actually were - devised and searched, e.g., TMJ disorders or endodontics, some preliminary attempts did not yield a comparable number of publications, and we refrained from expanding our corpus to maintain a certain degree of comparability between disciplines. For the same reason we abstained from expanding our corpus to earlier dates than 1994, although by doing so, we likely missed some pivotal events in the development of dentistry, such as the introduction of bonding in Restorative Dentistry [27], or the development of root-form osseointegrated fixtures in Implant Dentistry [13], which presumably greatly impacted the trends of research in those areas at the time.

Our results confirm that a growing number of publications is published each year, i.e., that the rate of publication is accelerating, as it has been previously reported [28], although different dental

disciplines understandably move at different speeds. In particular, Periodontics and Prosthodontics have been experiencing a faster growth than the other disciplines we considered. The growth of some disciplines, such as Oral Surgery, might even be starting to slow down, as the number of new publications per year has remained constant in the last 10 years. Several factors may compound in determining the number of publications in a discipline, including new discoveries (or new challenges), new needs that arise and must be addressed, or even changes in the way a certain issue is culturally conceived, interpreted or categorized. Determining why these disciplines are behaving in a specific way is outside of the scope of the present report, which aimed at measuring one aspect of the research infoscape in dentistry, i.e., its diversity. Diversity is a polysemous term, which has acquired a novel set of meanings, including that, currently prevalent in social sciences, of inclusivity of people in events, activities, jobs etc. beyond racial, gender, or religious divides [29,30]. However, this word has a rich history in the natural sciences, where it has been long used to refer to wide range of different life forms within ecosystems [31]. Our present works has focused on diversity in research, assessed through the main product of scientific activity, i.e., its dissemination through published papers [32]. In this context, diversity can be conceived as the presence, within a certain scientific field, of multiple areas of investigations, which are actively pursued by the scientific community. Dentistry was chosen, out of all the life science fields, because of the specific domain knowledge of the authors, and because it has quite defined boundaries, as compared to e.g., medicine, or biology, so it was easier to identify some major areas of investigation, which correspond to the main clinical branches of dental practice.

To understand more of what kind of research is conducted in these fields, however, we resorted to Topic modeling. Topic modeling is a field interested in processing unlabelled texts and automatically understanding their theme, their topic (their "about") using a wide range of techniques [33,34]. Recently, neural networks have allowed the creation of high performing algorithms to cluster documents based on their semantics, and BERTopic in particular has repeatedly exceeded the benchmark performance of previous algorithms on a wide range of tasks [18,35]. Clearly, BERTopic is far from being omniscient and still requires human input to work properly. At its core there is a transformer architecture that takes sentences, translates them into numerical sequences, known as embeddings, then clusters them based on their similarity and finds appropriate labels to characterize these clusters; to do that, it relies on a series of representation models, which range from keyword descriptors to small sentences, as BERTopic can also accommodate LLMs such as GPT [23], which we used to quickly characterize the main topics in Table 3. Depending on how its parameters are set, BERTopic can identify fewer and larger topics, or smaller and very numerous topics, splitting up the dataset into sometimes tiny clusters. Two critical parameters to this effect are the number of neighbors that the UMAP dimensionality reduction algorithm uses to process embeddings prior to clustering, and the minimum size of clusters that the HDBSCAN algorithm uses. By changing their values, the number of topics in a discipline could go down to 2-3, or could skyrocket to several thousands for a large dataset such as Prosthodontics or Periodontics (data not shown). The choice of the correct set of parameters is mostly arbitrary, according to the need of the investigators, and there are no safe and tested recipes to get the optimal – not to mention "true" – number of topics in a corpus. We picked the settings that yielded a sizable number of topics, around fifty or less, while keeping them manageable, and avoiding micro-topics of just a few articles and dubious meaning. Interestingly, although fields with more articles also (unsurprisingly) had more topics, the diachronic change in topic number did not directly reflect the number of publications. So, at the beginning of the period we investigated, although the Prosthodontics and Periodontics datasets included about 50% more articles than Restorative Dentistry, the number of topics of the latter was progressively becoming higher than Periodontics and almost twice as large than Prosthodontics. This would suggest that Periodontics and Prosthodontics are fields whose literature corpora have actively expanded over time faster than Restorative Dentistry, but maintaining a similar or lower number of topics as compared to Restorative Dentistry. We could envisage a situation, in Restorative Dentistry, where the front of research knowledge expands more homogeneously along a number of different trajectories. This would correspond to higher levels of Entropy, which correlate with a more

homogeneous distribution of articles across the topics. On the contrary, when it comes to Periodontics and Prosthodontics, our data suggest that these fields attract a vast number of new publications every year but these are more focused, i.e., these disciplines contain fewer topics (as in the case with Prosthodontics) or at least they have some main research directions which contain most of the publications and which are prominent over the rest. This could be indicative of a field that has matured, with researchers concentrating on specific, well-established topics rather than exploring new areas. On the other hand, disciplines like Oral Surgery have shown an increase in entropy, at least over a sustained period of time, reflecting a diversification of research topics. This may correspond to an expansion of this field, driven by innovations in implant materials (which are in part contained in this dataset), techniques (such as the use of laser in oral surgery [36]), and therapeutic strategies [37].

The divergence in entropy trends across disciplines raises some important questions about the direction of future research. Fields with increasing entropy, such as Oral Surgery but also, more limitedly Restorative Dentistry and Orthodontics, appear to be in a phase of exploration, where the topics expand isometrically, so to speak, and research questions are being continually investigated, although the number of new papers in a year is not increasing, i.e., the literature grows linearly. This could be interpreted as a sign of a dynamic and evolving field, which nevertheless is still being pursued by a 'niche' community, which is not expanding on par with other specialties (e.g., Periodontics). However, although a field like Periodontics is growing, as for number of publications, and number of topics, some main research axes (e.g., Periodontal Disease Treatment Studies, Smoking and periodontal disease, Periodontitis and preterm birth relation, to mention a few of its largest topics) are still robustly leading the research landscape [38].

Our findings suggest that dental disciplines are growing following different dynamics, and thus indirectly indicate the importance of the balance in a research agenda, to promote both depth and breadth. While specialization is necessary for advancing knowledge in specific areas, it is equally important to encourage exploratory research that can lead to the discovery of new topics and subfields. Policymakers, funding agencies, and academic institutions could consider similar approaches for data analysis when developing research priorities and funding strategies.

While this study may provide some valuable insights, it also has limitations that must be acknowledged. We relied on very generic bibliographic data from PubMed, using generic keywords. The results are not considered to be exhaustive of a specific field and may not fully capture the entirety of research within these disciplines [39]. Additionally, the selection of disciplines and the exclusion of others, such as TMJ disorders or Endodontics, were based on the availability of a comparable volume of publications, and richer datasets might allow to further expand this analysis and make it more extensive. Future studies could broaden the analysis to include these and other related fields to provide a more comprehensive overview of dental research.

5. Conclusions

This study provides a comprehensive analysis of the evolution and diversity of research within six core dental disciplines over nearly three decades. Using topic modeling (BERTopic), we quantified and compared the distribution of research topics; we measured the presence of predominant topics through Shannon's entropy over the years, revealing distinct trends across the disciplines. While some fields, such as Periodontics and Prosthodontics have experienced a strong growth in terms of article number, other fields have prompted the creation of more research topics, and the distribution of articles across these topics is more homogeneous. These insights emphasize these research fields in Dentistry differ for growth dynamics in the pursuit of their research lines. Such an approach could be useful to identify areas within the field of dental sciences that may need to improve innovation, to address the complex and evolving challenges they face.

Author Contributions: Conceptualization, C.G. and S.G.; methodology, C.G. and M.T.C.; software, C.G.; formal analysis, C.G. and M.T.C.; writing—original draft preparation, C.G. and M.T.C.; writing—review and editing, C.G. and S.G.; All authors have read and agreed to the published version of the manuscript.
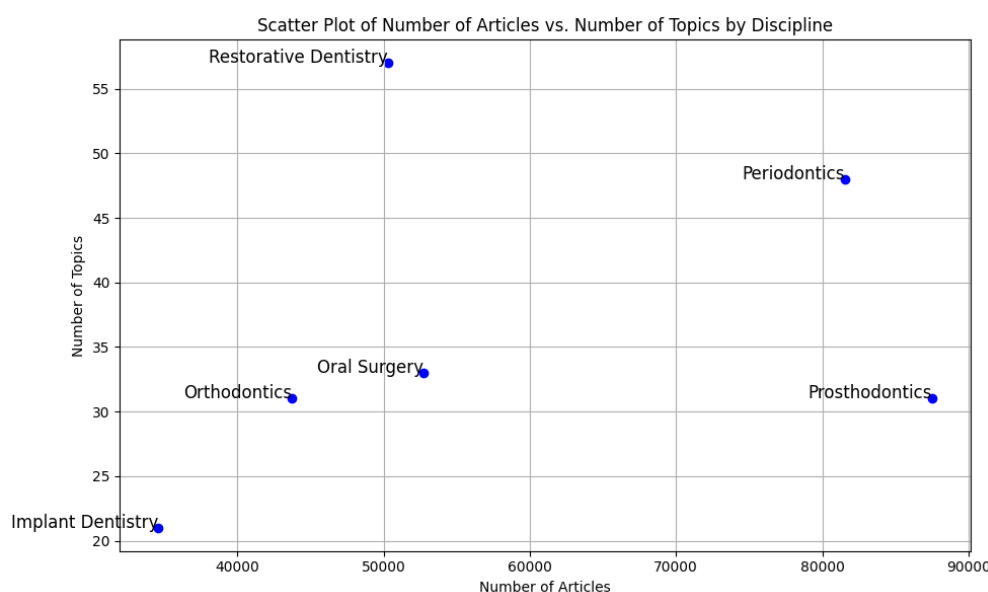
Institutional Review Board Statement: Not applicable.

Data Availability Statement: Data are available upon request.

Appendix A

Figure A1. Scatterplot showing the relation between number of articles and number of topics in the dataset per discipline.

References

1. Wang, X.; Guo, J.; Gu, D.; Yang, Y.; Yang, X.; Zhu, K. Tracking Knowledge Evolution, Hotspots and Future Directions of Emerging Technologies in Cancers Research: A Bibliometrics Review. *J. Cancer* **2019**, *10*, 2643.
2. Mantikayan, J.; Abdulgani, M. Factors Affecting Faculty Research Productivity: Conclusions from a Critical Review of the Literature. *JPAIR Multidiscip. Res.* **2018**, *31*, 1–21.
3. Schulman, K.A.; Rubenstein, L.E.; Chesley, F.D.; Eisenberg, J.M. The Roles of Race and Socioeconomic Factors in Health Services Research. *Health Serv. Res.* **1995**, *30*, 179–195.
4. Guizzardi, S.; Colangelo, M.T.; Mirandola, P.; Galli, C. Modeling New Trends in Bone Regeneration, Using the BERTopic Approach. *Regen. Med.* **2023**, *18*, 719–734.
5. Kherwa, P.; Bansal, P. Topic Modeling: A Comprehensive Review. *ICST Trans. Scalable Inf. Syst.* **2018**, *0*, 159623. https://doi.org/10.4108/eai.13-7-2018.159623.
6. Grootendorst, M. BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure. *arXiv preprint arXiv:2203.05794* **2022**.
7. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018,** arXiv:1810.04805.
8. Godden, J.W.; Bajorath, J. Analysis of Chemical Information Content Using Shannon Entropy. *Rev. Comput. Chem.* **2007**, *23*, 263–289.
9. Abdelrazek, A.; Eid, Y.; Gawish, E.; Medhat, W.; Hassan, A. Topic Modeling Algorithms and Applications: A Survey. *Inf. Syst.* **2023**, *112*, 102131.
10. Koltcov, S.; Ignatenko, V.; Koltsova, O. Estimating Topic Modeling Performance with Sharma–Mittal Entropy. *Entropy* **2019**, *21*, 660. https://doi.org/10.3390/e21070660.
11. Chen, L.; Zhang, H.; Jose, J.M.; Yu, H.; Moshfeghi, Y.; Triantafillou, P. Topic Detection and Tracking on Heterogeneous Information. *J. Intell. Inf. Syst.* **2018**, *51*, 115–137. https://doi.org/10.1007/s10844-017-0487-y.
12. Pulgar, R.; Jiménez-Fernández, I.; Jiménez-Contreras, E.; Torres-Salinas, D.; Lucena-Martín, C. Trends in World Dental Research: An Overview of the Last Three Decades Using the Web of Science. *Clin. Oral. Investig.* **2013**, *17*, 1773–1783. https://doi.org/10.1007/s00784-012-0862-6.
13. Buser, D.; Sennerby, L.; De Bruyn, H. Modern Implant Dentistry Based on Osseointegration: 50 Years of Progress, Current Trends and Open Questions. *Periodontol. 2000* **2017**, *73*, 7–21.

14. Bassi, S. A Primer on Python for Life Science Researchers. *PLoS Comput. Biol.* **2007**, *3*, e199.
15. Jia, Z.; Maggioni, M.; Smith, J.; Scarpazza, D.P. Dissecting the NVidia Turing T4 GPU via Microbenchmarking. *arXiv preprint arXiv:1903.07486* **2019**.
16. Cock, P.J.A.; Antao, T.; Chang, J.T.; Chapman, B.A.; Cox, C.J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B. Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423.
17. Mckinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the Proceedings of the 9th Python in Science Conference; van der Walt, S., Millman, J., Eds.; 2010; pp. 51–56.
18. Wang, Z.; Chen, J.; Chen, J.; Chen, H. Identifying Interdisciplinary Topics and Their Evolution Based on BERTopic. *Scientometrics* **2023**, 1–26.
19. Reimers, N.; Gurevych, I. Sentence-Bert: Sentence Embeddings Using Siamese Bert-Networks. *arXiv preprint arXiv:1908.10084* **2019**.
20. McInnes, L.; Healy, J.; Melville, J. Umap: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv preprint arXiv:1802.03426* **2018**.
21. McInnes, L.; Healy, J.; Astels, S. Hdbscan: Hierarchical Density Based Clustering. *J. Open Source Softw.* **2017**, *2*, 205.
22. Galli, C.; Donos, N.; Calciolari, E. Performance of 4 Pre-Trained Sentence Transformer Models in the Semantic Query of a Systematic Review Dataset on Peri-Implantitis. *Information* **2024**, *15*, 68.
23. Gue, C.C.Y.; Rahim, N.D.A.; Rojas-Carabali, W.; Agrawal, R.; Rk, P.; Abisheganaden, J.; Yip, W.F. Evaluating the OpenAI's GPT-3.5 Turbo's Performance in Extracting Information from Scientific Articles on Diabetic Retinopathy. *Syst. Rev.* **2024**, *13*, 135.
24. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*. https://doi.org/10.1109/MCSE.2007.55.
25. Waskom, M. Seaborn: Statistical Data Visualization. *J. Open Source Softw.* **2021**, *6*. https://doi.org/10.21105/joss.03021.
26. Landhuis, E. Scientific Literature: Information Overload. *Nature* **2016**, *535*, 457–458.
27. Singh, H.; Kaur, M.; Dhillon, J.S.; Mann, J.S.; Kumar, A. Evolution of Restorative Dentistry from Past to Present. *Indian. J. Dent. Sci.* **2017**, *9*, 38–43.
28. Rawat, S.; Meena, S. Publish or Perish: Where Are We Heading? *J. Res. Med. Sci.* **2014**, *19*, 87–89.
29. Dinesen, P.T.; Schaeffer, M.; Sønderskov, K.M. Ethnic Diversity and Social Trust: A Narrative and Meta-Analytical Review. *Annu. Rev. Political Sci.* **2020**, *23*, 441–465.
30. Budescu, D. V; Budescu, M. How to Measure Diversity When You Must. *Psychol. Methods* **2012**, *17*, 215.
31. Peet, R.K. The Measurement of Species Diversity. *Annu. Rev. Ecol. Syst.* **1974**, 285–307.
32. Hyland, K. Academic Publishing: Issues and Challenges in the Construction of Knowledge. **2016**.
33. Churchill, R.; Singh, L. The Evolution of Topic Modeling. *ACM Comput. Surv.* **2022**, *54*, 1–35.
34. Vayansky, I.; Kumar, S.A.P. A Review of Topic Modeling Methods. *Inf. Syst.* **2020**, *94*, 101582. https://doi.org/10.1016/j.is.2020.101582.
35. Gan, L.; Yang, T.; Huang, Y.; Yang, B.; Luo, Y.Y.; Richard, L.W.C.; Guo, D. Experimental Comparison of Three Topic Modeling Methods with LDA, Top2Vec and BERTopic. In Proceedings of the International Symposium on Artificial Intelligence and Robotics; Springer, 2023; pp. 376–391.
36. Noba, C.; Mello-Moura, A.C.V.; Gimenez, T.; Tedesco, T.K.; Moura-Netto, C. Laser for Bone Healing after Oral Surgery: Systematic Review. *Lasers Med. Sci.* **2018**, *33*, 667–674.
37. Lee, K.C.; Chuang, S.-K. History of Innovations in Oral and Maxillofacial Surgery. *Front. Oral. Maxillofac. Med.* **2022**, *4*, 6. https://doi.org/10.21037/fomm-21-24.
38. Alqahtani, H.M.; Haq, I.UI.; Alrubayan, M.; Alammari, F.; Alotaibi, F.; Al Khammash, A. A Bibliometric Analysis of the Top 100 Cited Articles in Regenerative Periodontics Surgery: Insights and Trends. *J. Int. Soc. Prev. Community Dent.* **2024**, *14*, 167–179. https://doi.org/10.4103/jispcd.jispcd_8_24.
39. Khare, R.; Leaman, R.; Lu, Z. Accessing Biomedical Literature in the Current Information Landscape. *Biomed. Lit. Min.* **2014**, 11–31.