Review

A Macroscopic Exploration of the Ideoscape on Exosomes for Bone Regeneration Using BERTopic and Knowledge Graphs

Maria Teresa Colangelo , Marco Meleti , Stefano Guizzardi , Carlo Galli *

Posted Date: 2 August 2024

doi: 10.20944/preprints202407.2583.v1

Keywords: exosomes; bone regeneration; BERTopic; knowledge graph; osteogenesis; tissue engineering; regenerative medicine

Review

A Macroscopic Exploration of the Ideoscape on Exosomes for Bone Regeneration Using BERTopic and Knowledge Graphs

Maria Teresa Colangelo [1], Marco Meleti [2], Stefano Guizzardi [1] and Carlo Galli [1,*]

[1] Histology and Embryology Laboratory, Department of Medicine and Surgery, University of Parma, Via Volturno 39, 43126 Parma, Italy
[2] Dental School, Department of Medicine and Dentistry, University of Parma, 43126 Parma, Italy
* Correspondence: carlo.galli@unipr.it

Abstract: Exosomes, nanoscale extracellular vesicles, are a hot area of investigation in the biomedical field, as they have been shown to play a crucial role in the physiology of numerous tissues, including bone. This study explores the landscape of exosome-based bone regeneration research using infometric techniques. By relying on BERTopic, an advanced topic modeling algorithm, we analyzed a comprehensive corpus of scientific literature to identify key themes and trends in exosome research in the bone regeneration field. We then extracted significant concepts from the abstracts of the corpus and used them to create knowledge graphs using GPT 3.5 turbo Large Language Model, to map the distribution of information within this domain. The resulting ideoscape highlights the key ideas that aggregate most of the recent research in the area, from mesenchymal stem cell-derived exosomes to engineered exosome-integrated biomaterials. This exploration provides a holistic view of the field and uncovers its emerging trends, providing useful insight to guide future research directions.

Keywords: exosomes; bone regeneration; BERTopic; knowledge graph; osteogenesis; tissue engineering; regenerative medicine

1. Introduction

Exosomes are a subtype of extracellular vesicles (EVs) released from eukaryotic and prokaryotic cells and defined as a spheroidal structure composed of a lipid bilayer containing a wide variety of cargo, depending on the donor cell type and their physiological conditions (Lange et al., 2024). These nanoscale vesicles usually range from 30 to 150 nanometers (average 100nm) in diameter and have an endosomal origin (Théry et al., 2006). They are formed through sequential invaginations of the plasma membrane that lead to the formation of multivesicular bodies (MVBs) containing intraluminal vesicles (ILVs) which are ultimately secreted via exocytosis as exosomes (Arya et al., 2024). The interaction of MVBs with other vesicles and intracellular organelles contributes to the diversity of exosome constituents, such as DNA, RNA, lipids, metabolites, and cytosolic and cell-surface proteins (Kalluri & LeBleu, 2020). By transferring these bioactive molecules between cells, that exosomes are natural endogenous nano-carrier that play a crucial role in near and long-distance intercellular communication in health and disease (Mathieu et al., 2019). Their intrinsic properties in regulating complex intracellular pathways have drawn significant interest and enhanced their potential utility in the field of early diagnosis and therapy of tissue diseases (Harrell et al., 2020; Kučuk et al., 2021; Mukherjee et al., 2022).

Bone tissue engineering (BTE) that uses stem cells has become a potential strategy for tissue defect regeneration and, in this regard, mesenchymal stromal/stem cells (MSCs) have been shown to be efficacious for the treatment of bone defects and diseases (Egido-Moreno et al., 2021; Xue et al., 2022). Interestingly, accumulating evidence indicates that the regenerative effect of MSCs is primarily

 doi:10.20944/preprints202407.2583.v1

to promote the activity of tissue-resident recipient cells through paracrine, such as exosomes (X. Wang & Thomsen, 2021). As a key element of cell-free therapy, exosome is becoming a promising tool of bone regeneration in recent decades (Tan et al., 2020; Torrecillas-Baena et al., 2023), because of its promoting osteogenesis and osteogenic differentiation function *in vivo* and *in vitro* by overcoming the disadvantages of their cellular counterparts due to their low immunogenicity, small size, and wide range of sources (Irfan et al., 2023). Furthermore, in order to enrich the exosomal cargo and increase exosome targeting efficiency, engineering exosomes and exosome-integrated biomaterials have been developed and shown to achieve ideal capabilities, such as specific bone targeting, better osteogenic and angiogenic properties and bone healing promotion (Meng et al., 2022). In light of these findings, exosomes have been shown to be a valid alternative to traditional cell-based therapies and an attractive strategy for therapeutic purposes in bone repair, regeneration and age-related defects.

To better understand the potential of exosomes for bone regeneration, it is essential to understand the current landscape of research in this fast-moving field. In this study, we take an computational approach to explore the knowledge structure surrounding exosome research focused on bone regeneration. A useful heuristic concept we use is the "macroscope," a metaphorical tool introduced by de Rosnay in the 1970s. (De Rosnay, 2014) Unlike a microscope, which zooms in on tiny details, or a telescope, which looks at distant objects, a macroscope provides a broad, integrative view of large-scale patterns and structures. Instead of focusing on the results individual reports have contributed to the field, we focus our attention on the themes that compose this research mosaic, and how they intersect.

The macroscope is advantageous to explore the ideoscape of this research field. The concept of "ideoscape" helps us visualize and analyze the flow and distribution of information within a scientific domain. Originally used in anthropology, "ideoscape" combines "idea" with the suffix "-scape," suggesting a landscape of information that can be explored (Appadurai, 1996). In anthropology, this concept helps understand how information moves within and between cultures, shaped by social practices, media, and technology. When applied to scientific research, we use this ideoscape concept to map out how knowledge is produced and disseminated, showing how certain studies or topics become influential and how new ideas spread.

To this purpose, we employ advanced topic modeling techniques, using transformer-based embeddings and the BERTopic algorithm (Grootendorst, 2022). Transformer-based embeddings, like those from BERT (Bidirectional Encoder Representations from Transformers), offer a deep understanding of text, which improves the accuracy of topic modeling (Q. Liu et al., 2020). BERTopic uses these embeddings to create dynamic topic representations, helping us identify evolving themes and trends in exosome research (Z. Wang et al., 2023). By integrating these methods, we systematically analyze a large collection of scientific articles on exosome research, giving us a comprehensive view of the field.

This macroscopic approach is particularly useful in today's globalized world, where scientific research is shaped by various factors, including funding, collaboration networks, technological advancements, and socio-political contexts (Börner & Record, 2017). Unlike a traditional narrative or systematic review, which are invaluable at understanding what research has elucidated on a given topic, mapping out knowledge production in a research field helps us understand its development, the forces at play, and identify gaps or opportunities for future research.

2. Materials and Methods

Data were analyzed on Google Colab Pro notebooks powered by Python 3.10.12 (Bassi, 2007) and running on T4 GPUs (Jia et al., 2019). The dataset was created with the Biopython library (Cock et al., 2009) through a query-driven exploration of MEDLINE using the Entrez.esearch function. The query utilized for this exploration was:

"Exosome"[MeSH] OR "Exosome*"[Title/Abstract] OR "Extracellular Vesicle*"[Title/Abstract] OR "Microvesicle*"[Title/Abstract]) AND ("Tissue Regeneration"[Title/Abstract] OR "Regenerative Medicine"[MeSH] OR "Regenerative Medicine"[Title/Abstract] OR "Tissue Repair"[Title/Abstract] OR "Regenerative Therapy"[Title/Abstract] OR "Wound Healing"[MeSH]*

 doi:10.20944/preprints202407.2583.v1

We iterated the web scraping process through Entrez.efetch spanning publication years from 1952 to 2025. We thus retrieved PubMed ID (PMID), title, publication year, authors, abstract, and Mesh keywords for all publications, and formatted them into a pandas dataframe (Mckinney, 2010).

We decided to classify articles based on their titles, as these are a summary of the article topic (Cook et al., 2007): a title is usually designed to communicate the gist of a report in few words, and we thus deemed it useful to characterize the article's topic (Hartley, 2007). The dataset was not pre-processed, unlike previous publications (Guizzardi et al., 2023), besides removing entries without titles. We did not remove stopwords—common grammatical words with no semantic meaning—to let BERTopic use them to create contextual embeddings (Saif et al., 2014).

Embeddings are dense vectors that semantically encode word meanings into high-dimensional space (Gutiérrez & Keith, 2019). Unlike previous algorithms (S. Wang et al., 2020), BERT understands context and produces different embeddings for the same word based on its context (Q. Liu et al., 2020; Vaswani et al., 2017), which is crucial for nuanced semantics in titles.

We used BERTopic to extract topics from publication titles in our dataset. BERTopic involves taking the embedding models, reducing their dimensionality, clustering them, and tagging the clusters with an algorithm known as cTF-IDF (Qaiser & Ali, 2018)(D. D. Xu & Wu, 2014). We selected Huggingface's 'all-mpnet-base-v2' model, pre-trained on over 1 billion token pairs. These multi-dimensional embeddings needed dimensionality reduction to maintain the data's topological structure. We used UMAP, which constructs a high-dimensional graph and optimizes it in lower-dimensional space (McInnes et al., 2018). Documents were then clustered with HDBSCAN to detect high-density semantic areas, forming clusters (McInnes et al., 2017). cTf-Idf was then applied to highlight key topic words within clusters.

Besides, BERTopic supports topic fine-tuning with additional representation models. We used KeyBERT and Maximal Marginal Relevance (MMR) models. KeyBERT uses the transformers library to extract keywords (Issa et al., 2023), using a Bag of Words approach (Y. Zhang et al., 2010). The similarity between document embeddings and keyword embeddings is then compared to select the most similar keywords. MMR selects keywords similarly to KeyBERT but it is designed to boosts their diversity to capture a broader spectrum of meaning (Bennani-Smires et al., 2018).

Default topic labels in BERTopic are just a string of its 4 main keywords. We used OpenAI's GPT-3.5 Turbo to generate more readable labels for the topics. A Large Language Model (LLM) like GPT-3.5 Turbo is capable to understand, interpret, and produce human language in a coherent and contextually appropriate manner (Thirunavukarasu et al., 2023). LLMs need a prompt from users, which serves as the starting point for the model to generate relevant text based on the information provided. We set the following prompt:

I have a topic that contains the following documents:

[DOCUMENTS]

The topic is described by the following keywords: [KEYWORDS]

Based on the information above, extract a short but highly descriptive topic label of at most 5 words. Make sure it is in the following format:

topic: <topic label>

Stopwords were removed after topic creation, using the sklearn Countvectorizer function (Akre et al., 2023).

Data were visualized using BERTopic's inbuilt functions, and the matplotlib (Hunter, 2007) and seaborn libraries (Waskom, 2021). In addition, the Datamapplot library was utilized for effective cluster visualization (McInnes, 2024).

The trend in research topics was analyzed by a linear regression model using the scipy Linregress library (Virtanen et al., 2020). The independent variable was the publication year in different time intervals, and the dependent variable was the number of papers within a given topic. The slope indicated how quickly the number of papers on a given topic rose over the time interval.

To investigate the structure (or ideoscape) of individual topics, we resorted to knowledge graphs (Ji et al., 2022). To capture the semantic meaning of the abstracts, we employed GPT-3.5-turbo model, using the following prompt:

 doi:10.20944/preprints202407.2583.v1

You are an expert in extracting key concepts, keywords, and relationships from scientific text for creating a knowledge graph.

Please extract the key concepts and their relationships from the following text. Make sure to provide meaningful labels and categories for each concept and normalize similar keywords to the same form. Use a concise and clear format.

Here's the text:

\n\n{text}\n\n

Return the result in JSON format with 'nodes' and 'relationships' keys.

The results returned in JSON format containing 'nodes' and 'relationships' keys. Each concept was also categorized according to a predefined taxonomy, including categories such as Disease, Symptom, Treatment, Medicine, Process, Component, and others. We then utilized SpaCy for lemmatization and tokenization to normalize keywords and relationship labels extracted from the abstracts (Honnibal & Montani, 2017).

We established a connection to a Neo4j graph database (Miller, 2013) to store and visualize the relationships between the extracted concepts, with each article title represented as a 'Title' node, and relationships created between these nodes and their corresponding concept nodes. 'Year' nodes were also created to indicate the publication year of each article and connected to the 'Title' nodes. The free Neo4j aura online browser was used to execute Cypher queries and for creating and managing nodes and relationships in the Neo4j database.

3. Results and Discussion

3.1. Overview of the Dataset

As the purpose of the current investigation was to focus on research on the role of exosomes in tissue regeneration, and more precisely on bone regeneration, our resulting dataset was quite small (n=2868); the earliest articles did not date to before 1995 and the dataset mostly comprised recent or very recent research products, as shown in Figure 1.

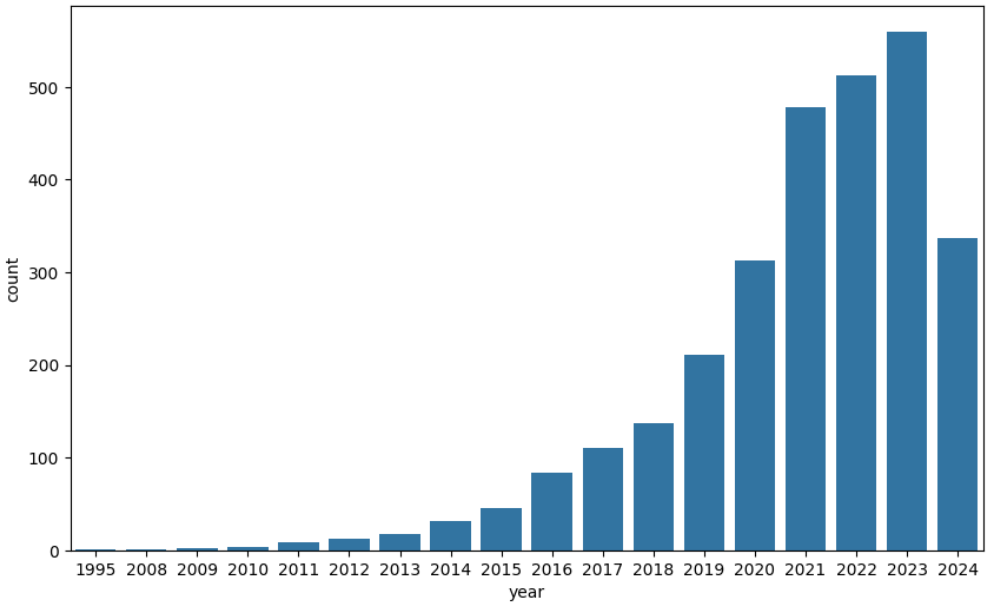

Figure 1. This barchart represents the distribution of publications in our dataset over the years.

As with most of the biomedical literature, the number of articles has been exponentially increasing in the last few years, and, in the case with exosomes in tissue regeneration, within the last decade. The number of published articles in 2024 is understandably lower than in 2023, at the time this manuscript is written.

 doi:10.20944/preprints202407.2583.v1

To get an overview of the knowledge structure of this science field, we focused our subsequent analysis on the titles of the articles in the dataset, and we ran the BERTopic algorithm on them. BERTopic segments a dataset of unstructured documents by clustering them based on their semantics. This is made possible by encoding every document, in this case every title, using transformer-based models, which generate sentence embeddings, i.e. a numerical representation of the text meaning. BERTopic can thus cluster these embeddings based on their similarity and, extracting representative keywords for each title cluster, can produce a description of its content. To obtain more readable topic labels and get a more 'human' feeling, we decided to use OpenAI's GPT to label them, in addition to BERTopic default meaning representations. GPT is a large language model (LLM), i.e. an artificial intelligence model capable to understand and generate language with a high degree of similarity to human language (Sarker, 2024). A brief comparison of the available GPT models up to the latest GPT 4o model showed no apparent difference, and given the remarkably lower usage cost, we opted for the GPT 3.5 turbo model.

This level of granularity in topic analysis can be fine-tuned arbitrarily. For our purposes, we set the minimum cluster size to 30, i.e. we excluded clusters with fewer than 30 articles. Articles that could not be classified into any other groups were allocated to a '-1' group.

BERTopic analysis, when executed with these parameters, revealed several key research areas within the academic literature on exosomes and a list of the identified topics can be found below (Table 1). For every topic group, we indicated the number of articles in it, its default name, which BERTopic creates by chaining together the 4 most relevant keywords that characterize it, the keywords extracted with the KeyBERT and MMR algorithms and the label obtained from the LLM.

Table 1. The table contains the list of topics identified by BERTopic, in order of size, in our exosome dataset.

TopicCount		Name	KeyBERT	MMR	LLM
-1	391	-1_stem_cells_derived_cell	['extracellular vesicles', 'derived exosomes', 'extracellular vesicle', 'exosomes', 'derived extracellular', 'stem cells', 'stem cell', 'exosome', 'vesicles', 'regenerative medicine']	['stem', 'cells', 'derived', 'cell', 'extracellular', 'vesicles', 'extracellular vesicles', 'mesenchymal', 'stem cells', 'stem cell']	['Stem Cell-Derived Vesicles in Regenerative Medicine']
0	443	0_bone_regeneration_stem_exosomes	['bone regeneration', 'vesicles bone', 'extracellular vesicles', 'marrow mesenchymal', 'extracellular vesicle', 'exosomes derived', 'stem cells', 'stem cell', 'tissue regeneration', 'cartilage regeneration']	['bone', 'regeneration', 'stem', 'exosomes', 'derived', 'osteoarthritis', 'cells', 'stem cells', 'mesenchymal', 'extracellular']	['Stem Cells for Bone Regeneration']
1	354	1_diabetic_adipose_wound_wound healing	['exosomes adipose', 'healing exosomes', 'exosomes promote', 'exosomes derived', 'vesicles adipose',	['diabetic', 'adipose', 'wound', 'wound healing', 'healing', 'derived', 'diabetic wound', 'adipose	['Exosomes for Diabetic Wound Healing']

 doi:10.20944/preprints202407.2583.v1

			'derived exosomes', 'stem cells', 'exosomes', 'adipose mesenchymal', 'adipose derived']	derived', 'stem', 'exosomes']	
2	239	2_extracellular_vesicles_extracellular vesicles_extracellular vesicle	['vesicle therapeutics', 'extracellular vesicles', 'extracellular vesicle', 'vesicles regenerative', 'vesicles extracellular', 'vesicles novel', 'vesicle mediated', 'vesicles tissue', 'vesicles new', 'vesicles cell']	['extracellular', 'vesicles', 'extracellular vesicles', 'extracellular vesicle', 'vesicle', 'tissue', 'regenerative', 'regenerative medicine', 'medicine', 'applications']	['Extracellular Vesicles in Regenerative Medicine']
3	207	3_exosomes_derived exosomes_cell_derived	['exosomes regenerative', 'exosomes derived', 'exosomes research', 'engineered exosomes', 'derived exosomes', 'exosomes new', 'cell exosomes', 'exosomes tissue', 'exosomes', 'role exosomes']	['exosomes', 'derived exosomes', 'cell', 'derived', 'stem', 'mesenchymal stem', 'mesenchymal', 'exosome', 'stem cell', 'cell derived']	['Mesenchymal Stem Cell Exosomes']
4	161	4_amniotic_lung_cell_cells	['stem cell', 'stem cells', 'extracellular vesicles', 'derived extracellular', 'derived exosomes', 'exosomes', 'regenerative medicine', 'covid 19', 'cell derived', 'cell extracellular']	['amniotic', 'lung', 'cell', 'cells', 'mesenchymal', 'stem', 'covid', '19', 'covid 19', 'extracellular vesicles']	['Mesenchymal stem cell therapies for COVID-19']
5	146	5_wound_wound healing_healing_skin	['vesicles skin', 'extracellular vesicles', 'extracellular vesicle', 'vesicles wound', 'vesicles derived', 'healing mesenchymal', 'healing cell', 'healing skin', 'skin regeneration', 'wound healing']	['wound', 'wound healing', 'healing', 'skin', 'extracellular', 'vesicles', 'extracellular vesicles', 'skin wound', 'cells', 'stem']	['Extracellular Vesicles in Skin Healing']
6	133	6_mesenchymal_stem_cells_medicine	['stem cells', 'stem cell', 'regenerative medicine', 'mesenchymal stem', 'cells regenerative', 'cell	['mesenchymal', 'stem', 'cells', 'medicine', 'stem cells', 'regenerative', 'regenerative medicine',	['Regenerative Medicine with Stem Cells']

 doi:10.20944/preprints202407.2583.v1

			therapy', 'tissue regeneration', 'secretome mesenchymal', 'mesenchymal stromal', 'human mesenchymal']	'mesenchymal stem', 'cell', 'stromal']	
7	128	7_wound_wound healing_exosomes_healing	['healing exosomes', 'exosomes wound', 'exosomes derived', 'exosomes promote', 'derived exosomes', 'exosomes', 'exosomes released', 'exosome', 'skin regeneration', 'stem cells']	['wound', 'wound healing', 'exosomes', 'healing', 'skin', 'cutaneous', 'derived', 'derived exosomes', 'cutaneous wound', 'exosomes derived']	['Exosome-enhanced cutaneous wound healing']
8	119	8_injury_spinal_spinal cord_cord injury	['treatment spinal', 'cells spinal', 'recovery spinal', 'derived exosomes', 'stem cell', 'stem cells', 'exosomes improve', 'exosomes', 'nerve regeneration', 'spinal cord']	['injury', 'spinal', 'spinal cord', 'cord injury', 'cord', 'brain', 'neural', 'stroke', 'stem', 'traumatic']	['Stem Cell Therapy Spinal Injury']
9	109	9_mesenchymal_extracellular vesicles_extracellular_vesicles	['extracellular vesicles', 'vesicles regenerative', 'extracellular vesicle', 'vesicles derived', 'vesicles extracellular', 'vesicles provide', 'vesicles mesenchymal', 'vesicles isolated', 'derived extracellular', 'vesicles tissue']	['mesenchymal', 'extracellular vesicles', 'extracellular', 'vesicles', 'derived extracellular', 'cell derived', 'mesenchymal stem', 'stromal', 'mesenchymal stromal', 'derived']	['Mesenchymal Stem Cell Extracellular Vesicles']
10	91	10_cardiac_heart_myocardial_cardiovascular	['cardiac regeneration', 'vesicles cardiac', 'cardiac repair', 'cardiac cell', 'vesicles cardiovascular', 'cardiomyocytes', 'cells cardiac', 'cardiac tissue', 'vesicles therapeutic', 'cardiac progenitor']	['cardiac', 'heart', 'myocardial', 'cardiovascular', 'repair', 'extracellular', 'vesicles', 'extracellular vesicles', 'cardiac repair', 'cardiac regeneration']	['Cardiac Repair with Extracellular Vesicles']
11	69	11_exosomes_myocardial_cardiac_exosome	['exosomes cardiac', 'exosomes cardiovascular', 'injury exosomes', 'exosomes	['exosomes', 'myocardial', 'cardiac', 'exosome', 'cardiovascular',	['Exosome Therapy in Cardiac Repair']

 doi:10.20944/preprints202407.2583.v1

			derived', 'cell exosomes', 'derived exosomes', 'exosomes', 'roles exosomes', 'exosomes secreted', 'injury exosome']	'myocardial infarction', 'infarction', 'derived', 'heart', 'ischemic']	
12	56	12_kidney_renal_kidney injury_acute kidney	['vesicles kidney', 'vesicles renal', 'extracellular vesicles', 'extracellular vesicle', 'injury extracellular', 'renal injury', 'cells kidney', 'cells renal', 'kidney regeneration', 'kidney injury']	['kidney', 'renal', 'kidney injury', 'acute kidney', 'injury', 'acute', 'extracellular', 'extracellular vesicles', 'vesicles', 'cells']	['Extracellular Vesicles in Kidney Injury']
13	53	13_corneal_corneal epithelial_retinal_epithelial	['exosomes corneal', 'vesicles corneal', 'corneal epithelial', 'cell exosomes', 'corneal cell', 'exosomes', 'derived exosomes', 'human corneal', 'treatment corneal', 'corneal wound']	['corneal', 'corneal epithelial', 'retinal', 'epithelial', 'corneal stromal', 'mesenchymal', 'cells', 'cell', 'mesenchymal stem', 'human corneal']	['Exosome-based Corneal Wound Healing']
14	47	14_tendon_cuff_rotator_rotator cuff	['exosomes tendon', 'tendon regeneration', 'tendon healing', 'promote tendon', 'tendon repair', 'improve tendon', 'enhance tendon', 'cuff tendon', 'cuff healing', 'tendon injury']	['tendon', 'cuff', 'rotator', 'rotator cuff', 'healing', 'bone', 'bone healing', 'tendon bone', 'tendon healing', 'achilles']	['Exosome Therapy for Rotator Cuff']
15	44	15_liver_fibrosis_hepatic_mesenchymal	['liver regeneration', 'hepatic fibrosis', 'liver fibrosis', 'secretome liver', 'ameliorate hepatic', 'cells liver', 'stem cells', 'stem cell', 'liver stem', 'mesenchymal stem']	['liver', 'fibrosis', 'hepatic', 'mesenchymal', 'liver fibrosis', 'mesenchymal stem', 'stem', 'cells', 'liver diseases', 'derived']	['Mesenchymal Stem Cells in Liver']
16	39	16_microvesicles_derived microvesicles_microvesicles derived_role microvesicles	['microvesicles regeneration', 'microvesicles derived', 'derived microvesicles', 'microvesicles	['microvesicles', 'derived microvesicles', 'microvesicles derived', 'role microvesicles', 'derived', 'tissue',	['Therapeutic Role of Stem Microvesicles']

 doi:10.20944/preprints202407.2583.v1

			mesenchymal', 'microvesicles cell', 'healing microvesicles', 'microvesicles tissue', 'extracellular microvesicles', 'microvesicle mediated', 'microvesicles intercellular']	'mesenchymal', 'stem', 'mesenchymal stem', 'role']	
17	39	17_micrornas_mir_exosomal_microrna	['exosomal micrornas', 'micrornas mesmirizing', 'micrornas', 'microrna stem', 'micrornas derived', 'microrna', 'micrornas novel', 'exosomal mir', 'micrornas perspective', 'micrornas vascular']	['micrornas', 'mir', 'exosomal', 'microrna', 'mirna', 'cells', 'mesenchymal', 'mesenchymal stem', 'cancer', 'exosomal micrornas']	['Exosomal MicroRNAs in Mesenchymal Cells']

The largest topic group, interestingly, is topic #0 Stem Cells for Bone Regeneration (n = 443), followed by topic #1 Exosomes for Diabetic Wound Healing (n = 354). Some topics are more general and reflect lexical differences rather than a specific focus, such as topic #2 Extracellular Vesicles in Regenerative Medicine. Other topics are more focused on the application of exosome research in specific fields, such as #5 Extracellular Vesicles in Skin Healing (n = 146), #8 Stem Cell Therapy for Spinal Injury (n = 119), and topic #10 Cardiac Repair with Extracellular Vesicles (n = 91). Unsurprisingly, these topics are not represented consistently over the span of the last 30 years. Figure 2A shows the number of topics over time in our dataset. The number of topics increased progressively since 1995 and reached the final level of 19 topics only in 2019, indicating that research has been expanding in quantity and in scope, with new areas of research being actively investigated and added to the field.

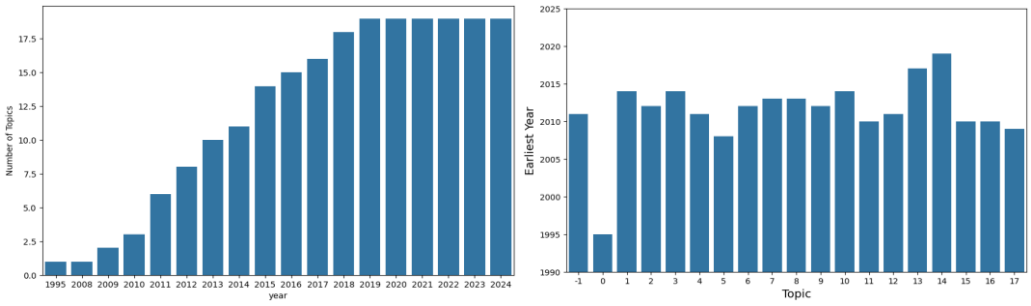

Figure 2. A) Barchart representing the number of topics in the dataset per year. B) Barchart representing the earliest publication per topic.

Figure 2B indicates that the oldest - and largest - topic is #0 Stem Cells for Bone Regeneration. Most other topics emerged during the 2010-2015 period, with the most recent being topic #14 Exosome Therapy for Rotator Cuff (n = 47).

To obtain a visual representation of how this research field is structured and understand the knowledge architecture of this research area, we reduced the embeddings to 2 dimensions, using the dimensionality reduction algorithm UMAP, so that each title could be represented as a dot in a

 doi:10.20944/preprints202407.2583.v1

cartesian semantic space. Dots that are closer together represent article titles with similar meanings, while dots that are farther apart indicate articles with more distant topics, even though the actual placement of the dots is random. Figure 3 displays a scatter plot for the 2868 titles in the exosome dataset. Grey dots represent unclassified titles scattered across the semantic space and classified by HDBSCAN as cluster -1. Interestingly, the left part of the semantic space is populated by articles on wound healing, such as skin wound healing and diabetic wound healing, which are understandably close to each other. Bone regeneration research is clustered in a different area, closer to mesenchymal stem cell research, rotator cuff research, and maybe surprisingly, corneal wound healing research.

3.2. Exosomes for Bone Regeneration

We then decided to focus our attention on bone regeneration alone, and isolated the papers belonging to topic #0, re-running BERTopic on this subset alone. BERTopic thus identified 7 topics (Table 2) within the bone subset. Noticeably, only 14 articles remained unclassified in the -1 cluster, while most of the articles of the dataset was allocated to a specific group.

The largest topic, #0 (n=126), was labelled quite generically "Mesenchymal Stem Cell Therapy" by GPT. Insights into its content can be provided by its keyword descriptors, which situate it quite firmly in the orthopedic area, with special focus on osteoarthritis:

['cells osteoarthritis', 'cartilage regeneration', 'exosomes derived', 'stem cells', 'stem cell', 'treatment osteoarthritis', 'extracellular vesicles', 'derived exosomes', 'osteoarthritis treatment', 'mesenchymal stem']

Visual inspection of the topic group revealed several papers devoted to this theme, e.g.:

"Mesenchymal stem cell-derived exosomes as a promising cell-free therapy for knee osteoarthritis."(Luo et al., 2024)

"Exosomes Reshape the Osteoarthritic Defect: Emerging Potential in Regenerative Medicine-A Review."(Sankaranarayanan et al., 2024)

"Exosomes Derived from Bone Marrow Mesenchymal Stem Cells Alleviate Rheumatoid Arthritis Symptoms via Shuttling Proteins."(L. Wang et al., 2024)

"Exosomes derived from MSC as drug system in osteoarthritis therapy."(Wen et al., 2024)

Topic #1, a similarly large cluster (n=103) appeared to mostly contain research on bone fractures and bone healing and LLM thus labelled it "Bone fracture healing exosomes". A closer inspection of the cluster revealed that it contained at least 2 main research lines, i.e. both articles investigating the physiological role of exosomes in bone repair, e.g.:

"*Proteomic analysis of exosomal proteins associated with bone healing speed in a rat tibial fracture model."*(Hong et al., 2024)

"*Cellular and Molecular Connections Between Bone Fracture Healing and Exosomes."*(Lv et al., 2023)

and research articles devoted to the biotechnological use of exosomes to improve bone healing, e.g.:

"*Role of nano-hydrogels coated exosomes in bone tissue repair."* (Pan et al., 2023)

"*Enhancing bone regeneration and immunomodulation via gelatin methacryloyl hydrogel-encapsulated exosomes from osteogenic pre-differentiated mesenchymal stem cells."*(Li et al., 2024)

Topic #2 Dental Stem Cell Regeneration (n=80) and topic #4 'Dental Pulp Stem Cell Exosomes (n=48) were mostly centered about dental and periodontal applications of exosomes. As their LLM label might not fully account for the difference between these two topic clusters, we inspected the keyword descriptors for topic #2:

['periodontal regeneration', 'vesicles periodontal', 'vesicles dental', 'periodontal tissue', 'pulp regeneration', 'regeneration dental', 'pulp periodontal', 'extracellular vesicles', 'extracellular vesicle', 'dental pulp']

which suggests that this topic contains articles focusing on periodontics applications of exosomes. This was confirmed through visual inspection of this group, which revealed academic articles in the periodontics area, e.g.:

"*Clinical Efficacy of Extracellular Vesicle Therapy in Periodontitis: Reduced Inflammation and Enhanced Regeneration."*(Puletic et al., 2024)

 doi:10.20944/preprints202407.2583.v1

"3D bioprinted small extracellular vesicles from periodontal cells enhance mesenchymal stromal cell function."(Han et al., 2024)

"Effects of periodontal cells-derived extracellular vesicles on mesenchymal stromal cell function."(Han et al., 2023)

On the contrary, the KeyBERT-derived descriptors for topic #4 were:

['exosomes dental', 'exosomes regenerative', 'exosomes oral', 'exosomes derived', 'cell exosomes', 'exosomes', 'cells exosomes', 'exosomes enhance', 'derived exosomes', 'regenerative endodontic']

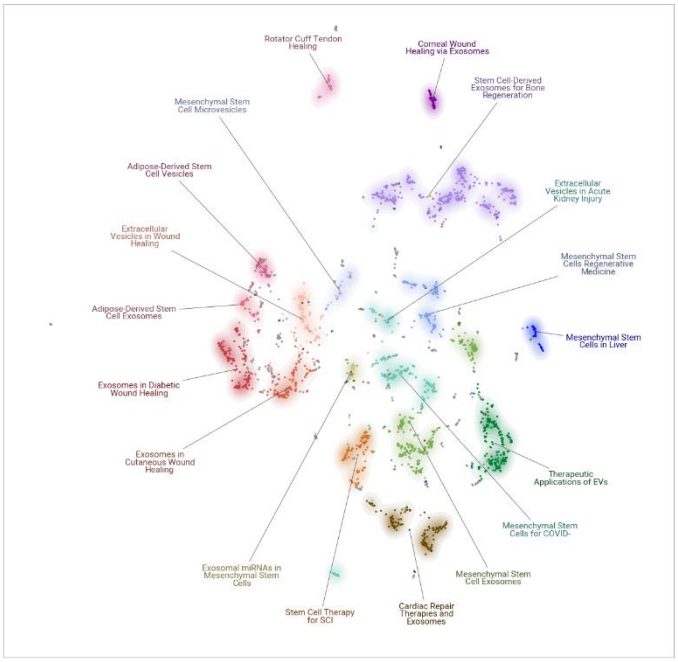

Figure 3. Scatterplot of the semantic distribution of the dataset of titles of scientific articles on exosomes and tissue regeneration. Titles are not homogeneously distributed, but rather form clusters that correspond to topics. Every topic is marked by a different color and is indicated by its LLM label.

Table 2. This is a table. Tables should be placed in the main text near to the first time they are cited.

TopicCount		Name	KeyBERT	MMR	LLM
-1	14	-1_bone_related_mesenchymal_containing mesenchymal	['exosomes osteoarthritis', 'bioprinted scaffolds', 'osteonecrosis jaw', 'uptake osteoblast', 'bone cartilage', 'regenerative treatments', 'osteonecrosis', 'lyosecretome bone', 'mesenchymal stem', 'related osteonecrosis']	['bone', 'related', 'mesenchymal', 'containing mesenchymal', 'medication', 'medication related', 'stromal lyosecretome', 'controlled release', 'mesenchymal stem', 'stem']	Bone Regeneration Controlled Release
0	126	0_osteoarthritis_mesenchymal_cartilage_extracellular	['cells osteoarthritis', 'cartilage	['osteoarthritis', 'mesenchymal', 'cartilage',	Mesenchymal Stem Cell Therapy

 doi:10.20944/preprints202407.2583.v1

			regeneration', 'extracellular', 'treatment', 'exosomes derived','stem', 'derived', 'stem cells', 'stem cell','mesenchymal stem', 'treatment 'exosomes', 'extracellular osteoarthritis', vesicles'] 'extracellular vesicles', 'derived exosomes', 'osteoarthritis treatment', 'mesenchymal stem']		
1	103	1_bone_exosomes_mesenchymal stem_derived	['exosomes bone', 'regeneration exosomes', 'exosomes promote', 'cell exosomes', 'exosomes', 'bone regeneration', 'derived exosomes','osteogenesis angiogenesis', 'exosomal mirna','marrow mesenchymal']	['bone', 'exosomes', 'mesenchymal stem', 'derived', 'fracture', 'mesenchymal', 'stem', 'cells', 'healing', 'stem cells']	Bone fracture healing exosomes
2	80	2_dental_stem_extracellular_vesicles	['periodontal regeneration', 'vesicles periodontal', 'vesicles dental', 'periodontal tissue', 'pulp regeneration', 'regeneration dental', 'pulp periodontal','extracellular vesicle','dental pulp']	['dental', 'stem', 'extracellular', 'vesicles', 'extracellular vesicles', 'cells', 'regeneration', 'stem cells', 'periodontal', 'cell']	Dental Stem Cell Regeneration
3	56	3_extracellular_vesicles_extracellular vesicles_bone	['bone regeneration', 'vesicles bone', 'hydroxyapatite scaffold', 'osteoblast derived', 'extracellular vesicles', 'bone'derived	['extracellular', 'vesicles', 'extracellular vesicles', 'bone', 'bone regeneration', 'regeneration', 'derived', extracellular',	Stem Cell-Derived Vesicles for Bone Regeneration

 doi:10.20944/preprints202407.2583.v1

			healing', 'extracellular'small extracellular', vesicle', 'regeneration'small'] extracellular', '3d hydroxyapatite', 'tissue regeneration']		
4	48	4_exosomes_derived_stem _pulp	['exosomes dental', 'exosomes regenerative', 'exosomes oral', 'exosomes derived', 'cell exosomes', 'exosomes', 'cells exosomes', 'exosomes enhance', 'derived exosomes', 'regenerative endodontic']	['exosomes', 'derived', 'stem', 'pulp', 'derived exosomes', 'cell', 'dental', 'exosomes derived', 'cells', 'dental pulp']	Dental Pulp Stem Cell Exosomes
5	16	5_disc_intervertebral disc_intervertebral_disc degeneration	['disc degeneration', 'exosomes ameliorate', 'exosomes derived', 'intervertebral disc', 'exosome transplantation', 'exosomes', 'cell exosomal', 'stem cells', 'exosomal mirnas']	['disc', 'intervertebral disc', 'intervertebral', 'disc degeneration', 'degeneration', 'stem', 'cells', 'stem cells', 'derived cells', 'intervertebral']	Exosome Therapy for Disc Degeneration

The MMR algorithm confirmed that this latter topic focuses on dental pulp rather than periodontal tissues as topic #2 (as suggested by the LLM label):

['exosomes', 'derived', 'stem', 'pulp', 'derived exosomes', 'cell', 'dental', 'exosomes derived', 'cells', 'dental pulp']

which justifies the distinction between these two groups. Inspection of this cluster confirmed the descriptors, highlighting the focus on the dental pulp, as origin or therapeutic target of the exosomes, e.g.:

"*Exosomes as Promising Therapeutic Tools for Regenerative Endodontic Therapy.*"(Kong et al., 2024)

"*Human dental pulp stem cell-derived exosomes decorated titanium scaffolds for promoting bone regeneration.*"(S. Zhang et al., 2024)

"*Exosomes derived from odontogenic stem cells: Its role in the dentin-pulp complex.*"(Zou et al., 2023)

Topic #3 Stem Cell-Derived Vesicles for Bone Regeneration seems to substantially overlap topic #1, but its descriptors reveal an additional and possibly distinctive focus, i.e. a special attention to biomaterial scaffolds, which was absent in topic #0:

['bone regeneration', 'vesicles bone', 'hydroxyapatite scaffold', 'osteoblast derived', 'extracellular vesicles', 'bone healing', 'extracellular vesicle', 'regeneration extracellular', '3d hydroxyapatite', 'tissue regeneration']

 doi:10.20944/preprints202407.2583.v1

To better understand the composition of this group, we manually screened its titles and identified articles that focused on materials, scaffolds and the use of exosomes for therapy, rather than the study of their physiological role, e.g.:

"*Functionalized 3D Hydroxyapatite Scaffold by Fusion Peptides-Mediated Small Extracellular Vesicles of Stem Cells for Bone Tissue Regeneration.*"(Ma et al., 2024)

"*Transplanted MSCs promote alveolar bone repair via hypoxia-induced extracellular vesicle secretion.*"(Y. Liu et al., 2024)

"*Bioengineering extracellular vesicles: smart nanomaterials for bone regeneration.*"(Man et al., 2023)

"*Biomimetic synthesis and optimization of extracellular vesicles for bone regeneration.*"(Song et al., 2023)

The last topic is small (n=16) but very specific, as it is labelled "Exosome Therapy for Disc Degeneration" and contained very consistently themed articles on this specific niche area, e.g.:

"*Nanoscale Treatment of Intervertebral Disc Degeneration: Mesenchymal Stem Cell Exosome Transplantation.*"(Hu et al., 2023)

"*MSC-Derived Exosomes Ameliorate Intervertebral Disc Degeneration By Regulating the Keap1/Nrf2 Axis.*"(G. Xu et al., 2023)

"*Exosomes: A promising therapeutic strategy for intervertebral disc degeneration.*"(C. Wang et al., 2022)

Figure 4 represents the semantic distribution of these topics, which confirms that topic #5 is located at a certain distance from the other bone topics, possibly due to the use of a very specific language that positions it quite far apart from the rest of the orthopedic literature. On the contrary, topic #1 Bone fracture healing exosomes is visibly juxtaposed to topic #3 Stem Cell-Derived Vesicles for Bone Regeneration, confirming that there is some continuity between the two topic groups.

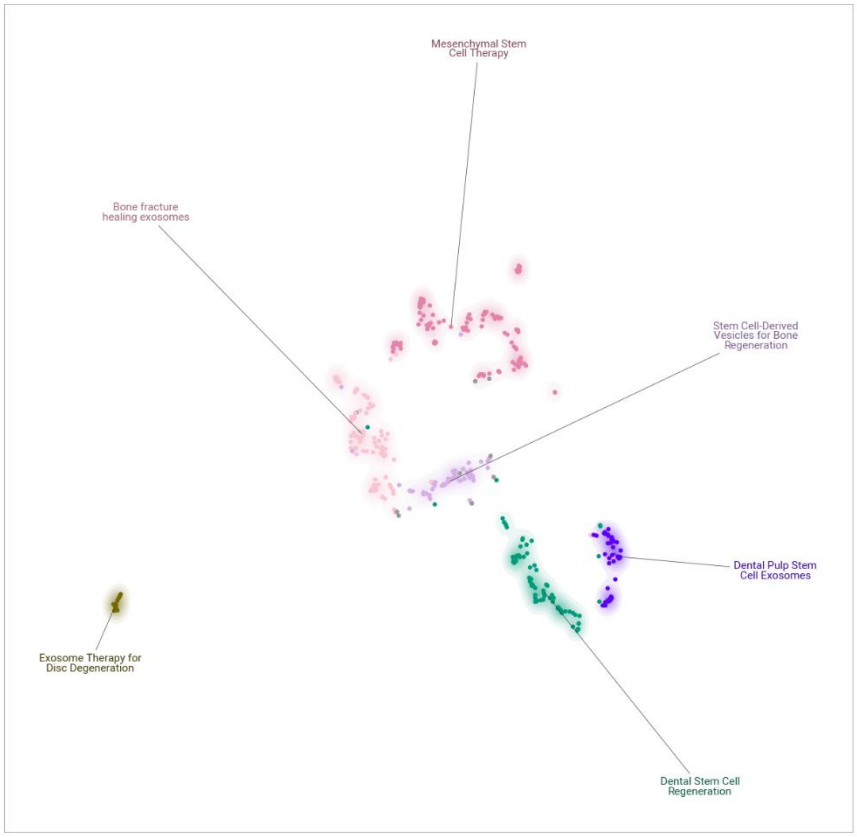

Figure 4. Scatterplot of the semantic distribution of the dataset of titles of scientific articles on exosomes and bone regeneration. Titles are not homogeneously distributed, but rather form clusters that correspond to topics. Every topic is marked by a different color and is indicated by its LLM label.

 doi:10.20944/preprints202407.2583.v1

3.3. Evolution of the Field

To better understand how the bone research field evolved over time, we plotted the growth of the topics as number of articles in the topic group over time (Figure 5). The plot clearly shows that, just like the rest of the dataset, this particular subset of literature went through a phase of expansion around the year 2017, with a rapid increase in the number of publications. While most of the topics flourished in the last 5 years, with a drop in the last part of the chart that is attributable only to the fact that the number of articles published in 2024 is only partial, it seems that Topic #0 of this subset peaked around the year 2020, while Topic #1 Bone fracture healing exosomes has taken over as the fastest growing topic.

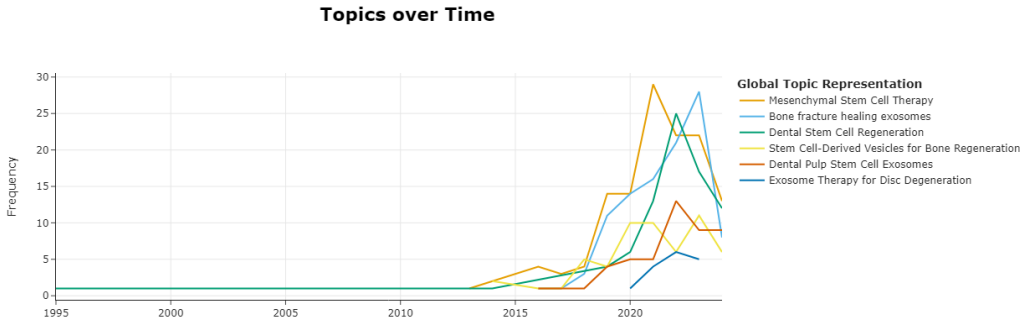

Figure 5. Line chart tracking the number of papers published by year in the bone regeneration topic subset. Each topic is indicated by a solid line of different color.

To gain further insight into the publication trends in these topics, we performed a linear regression over the number of papers published in the different clusters. Linear regression finds the equation of a line that minimizes the distance between the points on the line and the points in the dataset. The slope of the fitted line indicates the publication trend (Table 3), and it can be considered a crude indicator of how fast the number of publications is changing over time. Though most topics are growing (Table 3), with the only possible exception of topic #3 Stem Cell-Derived Vesicles for Bone Regeneration which has been maintaining similar levels of publications in the analyzed period, with a slope of 1, two topics stand out as growth, i.e. topic #2 Dental Stem Cell Regeneration and topic #1 Bone fracture healing exosome, with slopes exceeding 4. As Figure 6 shows, while Topic #0 Mesenchymal Stem Cell Therapy and Topic #1 Bone fracture healing exosomes are fast growing research fields, Topic #2 Dental Stem Cell Regeneration is the fastest growing topic, with a slope of 4.5 over the 2019-2023 period.

Table 3. List of the bone regeneration topics in the 2019-2023 period sorted by slope of the linear regression of the number of publications/year.

Topic	Trend (slope)
Dental Stem Cell Regeneration	4.5
Bone fracture healing exosome	4.1
Mesenchymal Stem Cell Therapy	2.4
Dental Pulp Stem Cell Exosomes	1.8
Exosome Therapy for Disc Degeneration	1.4
Stem Cell-Derived Vesicles for Bone Regeneration	1.0

 doi:10.20944/preprints202407.2583.v1

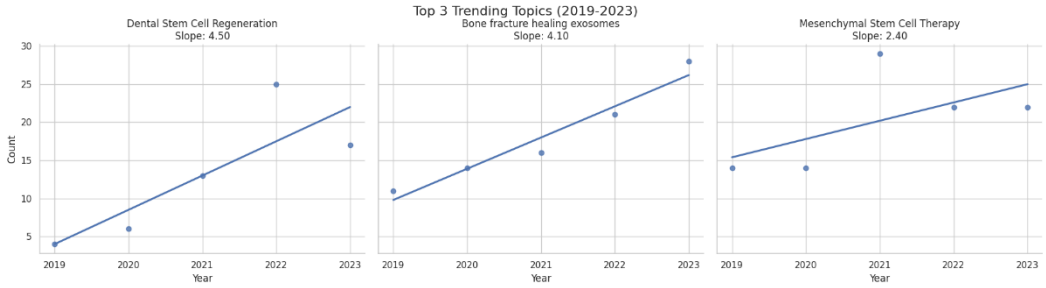

Figure 6. Trend analysis for the three highest-ranking topics in the bone regeneration subset. Linear regression was used to fit the number of publications per year in the 2019-2023 period and infer the publication trend from its slope.

3.4. Knowledge Structure of Exosome Research: The Ideoscape

To get a better understanding of how ideas and themes are shared in this fast growing field, we decided to limit our observation to topic #2 Dental Stem Cell Regeneration, which has been undergoing the fastest growth. We thus created a knowledge graph out of the abstracts of the articles belonging to this cluster, extracting meaningful keywords and relationships between concepts through GPT 3.5 turbo. We chose to use abstracts instead of titles because we assumed that abstracts briefly pan out the procedures, the research approaches and the main findings of individual research, unlike titles that are often only indicative of the general topic or area of research - and are thus very indicated for topic modeling. Unlike what we did with BERTopic, we did not extract key concepts through keyword-extraction algorithms, e.g. KeyBERT, but decided to rely completely on the LLM, with a specific prompt designed to get the main concepts and findings out of the report abstracts. This allowed us to build a network of ideas, which we consider a simplified representation of the ideoscape of this field, which was then investigated with Neo4j, a platform specifically designed to visualize and query knowledge graphs. Neo4j is an interactive tool to visualize graphs, which allows users to search through the graph using a query language called Cypher, similarly to what is done with the SQL language for relational databases (Francis et al., 2018).

A knowledge graph is a structured representation of knowledge that captures entities, their attributes, and the relationships between them, facilitating both human readability and the adoption of computational procedures to extract insights, or knowledge, from the data through queries (Hogan et al., 2022). It comprises nodes, representing entities or concepts, and edges, depicting the relationships between these nodes. Both nodes and edges can have multi-layered attributes that provide additional context. Such a graph is therefore a complex representation of concepts and how these concepts are related to each other (Peng et al., 2023).

GPT was capable of not only extracting keywords and assign them a label but also cluster keywords by categories, such as Anatomy, Cell models, Procedures, or Therapies, providing a further layer of insight and structure into these key concepts. We only reported selected screenshots of Neo4j graphs, or of meaningful subgraphs, although these are only a partial representation of the whole graph, and lack their distinguishing interactivity.

Figure 7 shows a graph of the nodes categorized as anatomical sites (orange circles) that were investigated by at least 2 research articles (large green circles), i.e. that were connected to at least 2 article nodes by a “Contains” edge.

Anatomy keywords that were connected through only 1 edge to an article node are not represented in this plot, because our purpose was to see how ideas could be shared between papers. So, only a few Anatomy nodes are displayed, out of a total of 84 concept nodes for this category. Large Yellow circles are Year nodes, i.e. nodes that represent the publication year of an article and are connected to articles published in that year through a “Published_in” edge.

The graph in Figure 7 shows how periodontal tissue has catalyzed most of the recent attention, with several articles gravitating around the “periodontal tissue” keyword (in agreement with the central interest for Periodontal disease as shown in graph S1), especially in the 2022-2024 period. This

 doi:10.20944/preprints202407.2583.v1

is in agreement with a very recent systematic review on this topic, which illustrated the many applications of exosomes for this area (T. Wang et al., 2024). Noticeably, in the case of the paper "Periodontal Ligament Stem Cells: Current Knowledge and Future Perspectives." (Trubiani et al., 2019) the LLM felt a closer connection to the Dental tissue descriptor node (Figure 7), possibly because the report was a review about the multi-lineage potential of stem cells from the periodontium. Interestingly, the paper was published in 2019, possibly before the periodontal tissue had imposed itself.

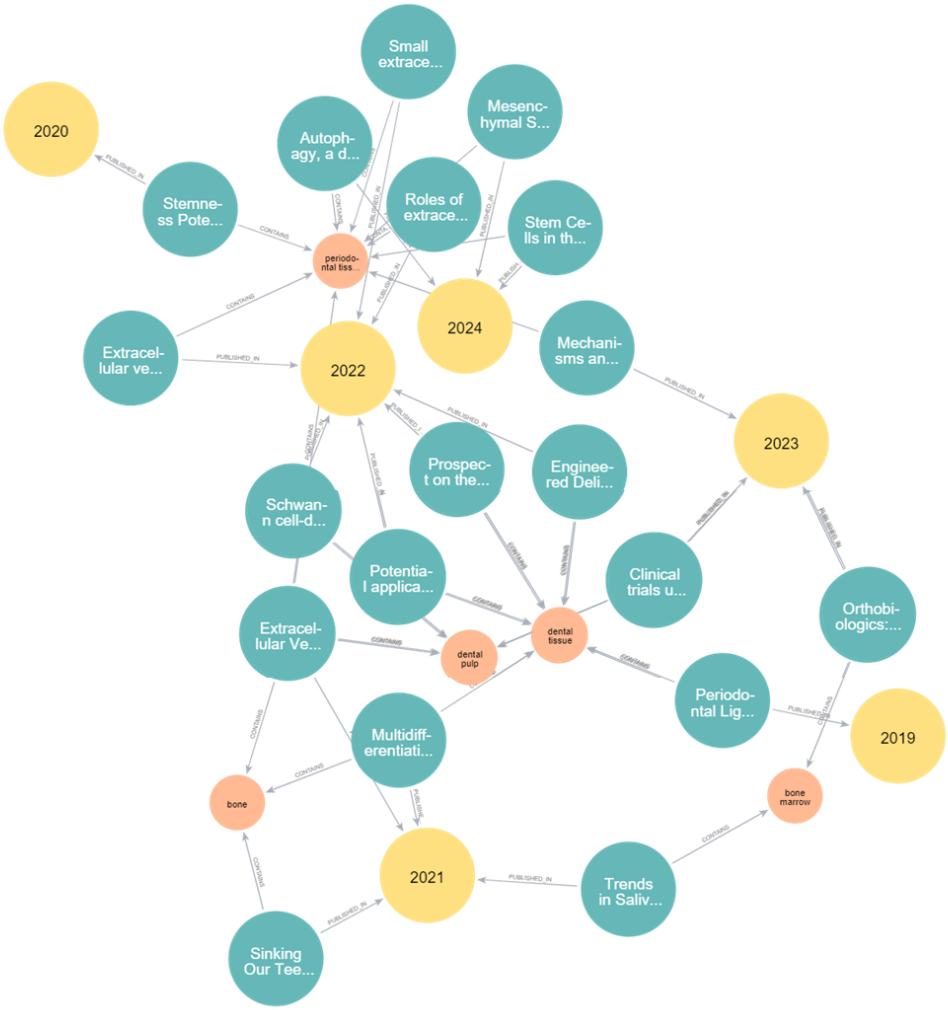

Figure 7. Neo4j Knowledge graph of Anatomy nodes (orange) connected to at least two article nodes (in green. A part of their title is written on the node). Yellow nodes indicate publication year.

As Stem cells were indeed a very popular keyword for this dataset, which recurred frequently in the descriptors extracted by BERTopic, we decided to analyze the most common stem cell models and topics, investigating the "Cells" nodes. Again, we only focused on those nodes that were connected to at least two article nodes (Figure 8, brown circles), to highlight idea interactions, or at least focal points of ideas. The main cell models that attracted most of the publications in the subsets were few (Figure 8), although the dataset contained 105 different Cells nodes.

 doi:10.20944/preprints202407.2583.v1

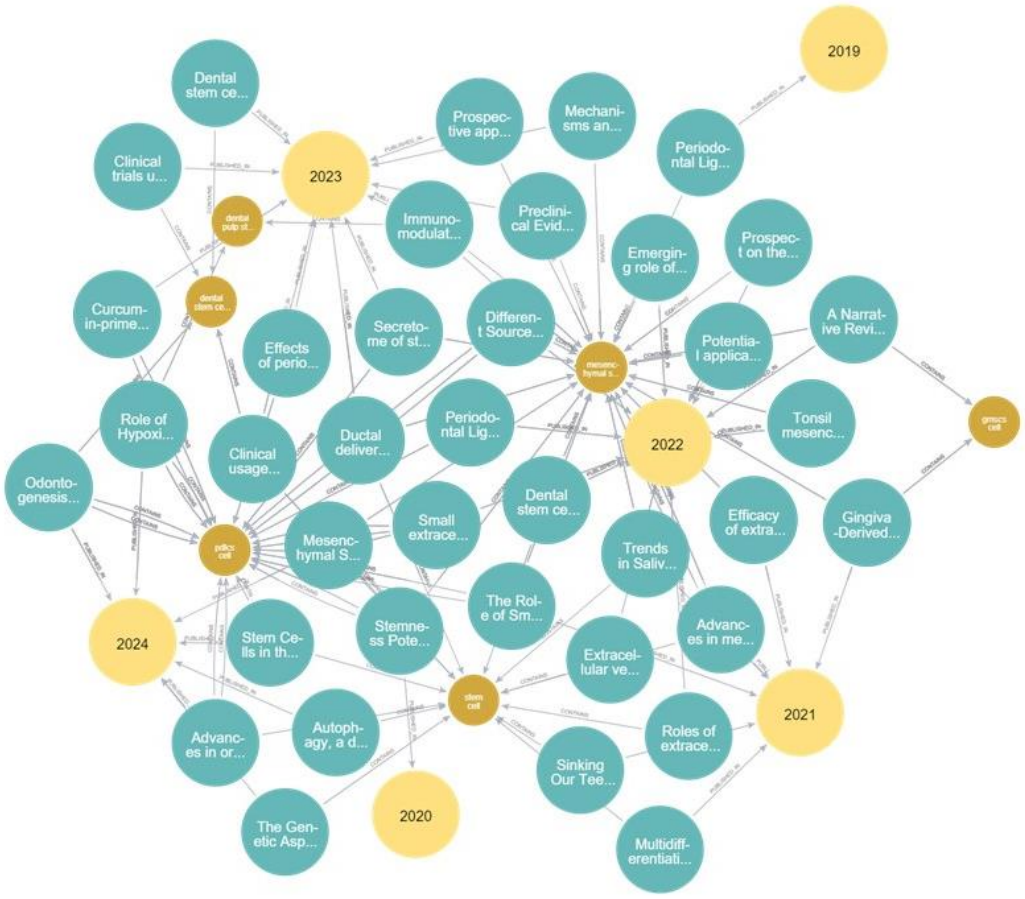

Figure 8. Neo4j Knowledge graph of Cells nodes (khaki) connected to at least two article nodes (in green. A part of their title is written on the node). Yellow nodes indicate publication year.

Mesenchymal stem cells were - possibly unsurprisingly so - the focus of a vast portion of the literature of this group. This could be explained both because mesenchymal stem cells are a flexible model in tissue regeneration and have a vast corpus of evidence (Ju et al., 2024) but also because deep periodontal tissues are of mesenchymal nature, and most approaches for their regeneration are thus naturally centered on the use of this cell model (Zeng et al., 2021).

Periodontal ligament stem cells were another common concept in this subset of the literature, and many research reports focused on them, albeit they have long represented a quite elusive cell population (Isamorad et al., 2023). On the contrary, dental stem cells and pulp stem cells were investigated by a smaller niche of the literature, while a dozen of articles gravitated around the more general "Stem cell" buzzword, while usually concomitantly connected also to a more specific cell node (Figure 8).

We then focused on the "Process" nodes that were investigated in this literature set, which is an interesting category per se given its polysemous nature (Figure 9). A total of 129 different Process nodes were identified by GPT, although those that are shared by at least 2 article nodes are much fewer. Interestingly, we could observe a pole of keywords that mostly concerned cellular processes, such as cell migration, proliferation and differentiation, and an opposing pole about Clinical application, biogenesis and tissue repair, with a strong Tissue regeneration concept node that coordinated almost the whole dataset, providing a key description to this group of articles (Figure 9). Overall, the LLM thus recognized two semantic areas for these processes: natural processes that are affected by exosomes or affect exosomes (i.e. the cell mechanisms) on one end and thematic areas of application of exosomes for therapy purposes.

 doi:10.20944/preprints202407.2583.v1

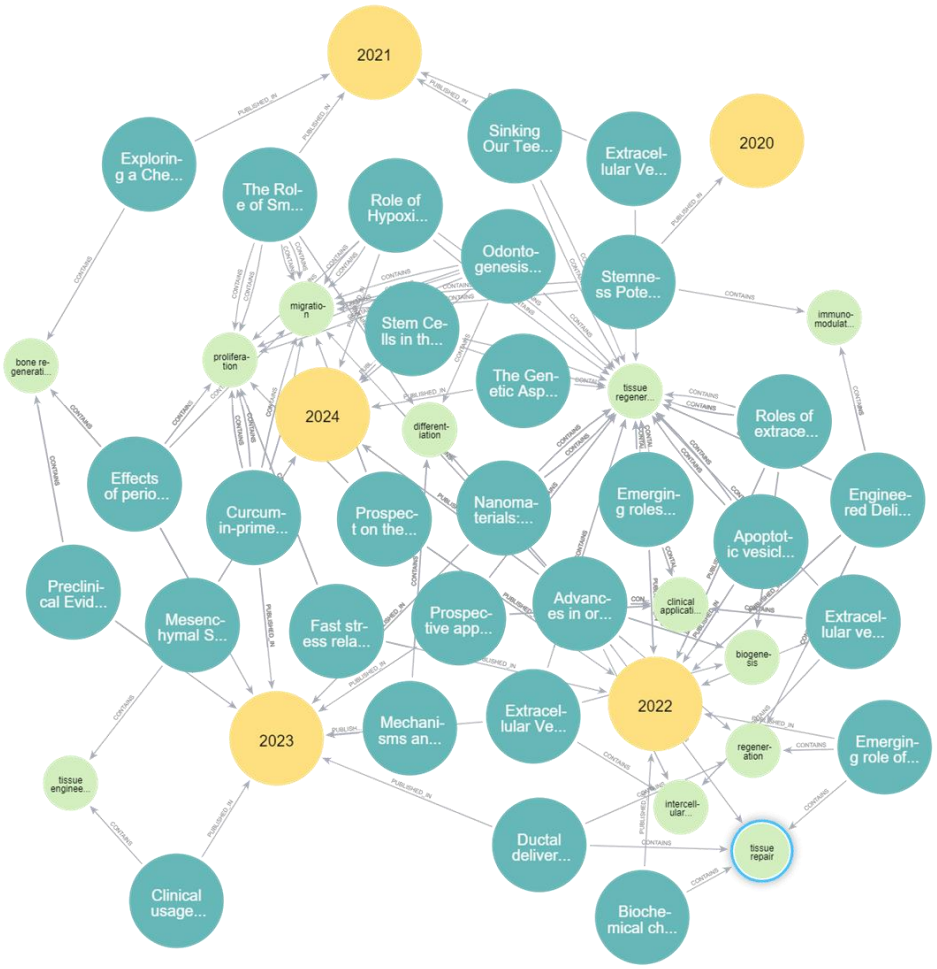

Figure 9. Neo4j Knowledge graph of Process nodes (light green) connected to at least two article nodes (darker green. A part of their title is written on the node). Yellow nodes indicate publication year.

GPT was unable to assign a specific category to a set of nodes, and these nodes were hence generically labelled "Unknown"; this large group (n=788) actually contains most of the generic keywords for the literature dataset (Figure 10), e.g. 'Chemokine', or 'In vitro research'. However, most of these concept nodes are very specific, e.g. 'beagle modem organism' and are associated to only 1 article node. The concept nodes that are shared across a number of articles are much more limited in number (Figure 10, light blue). It is easy to observe how "Tissue Regeneration" and "Extracellular Vescicles" are the main attractors for the more recent papers, as highlighted by the high number of connections that they have with articles in the dataset, which obviously is expected because of the nature of the dataset itself.

As expected, "Exosomes" and "Secretome" (albeit to a lesser extent) are also popular buzzwords, although they tend to cluster together with other concepts that are more typical of pre-clinical settings, such as "Proliferation", "Differentiation", "Angiogensis" or "Intercellular Communication" (Figure 10).

 doi:10.20944/preprints202407.2583.v1

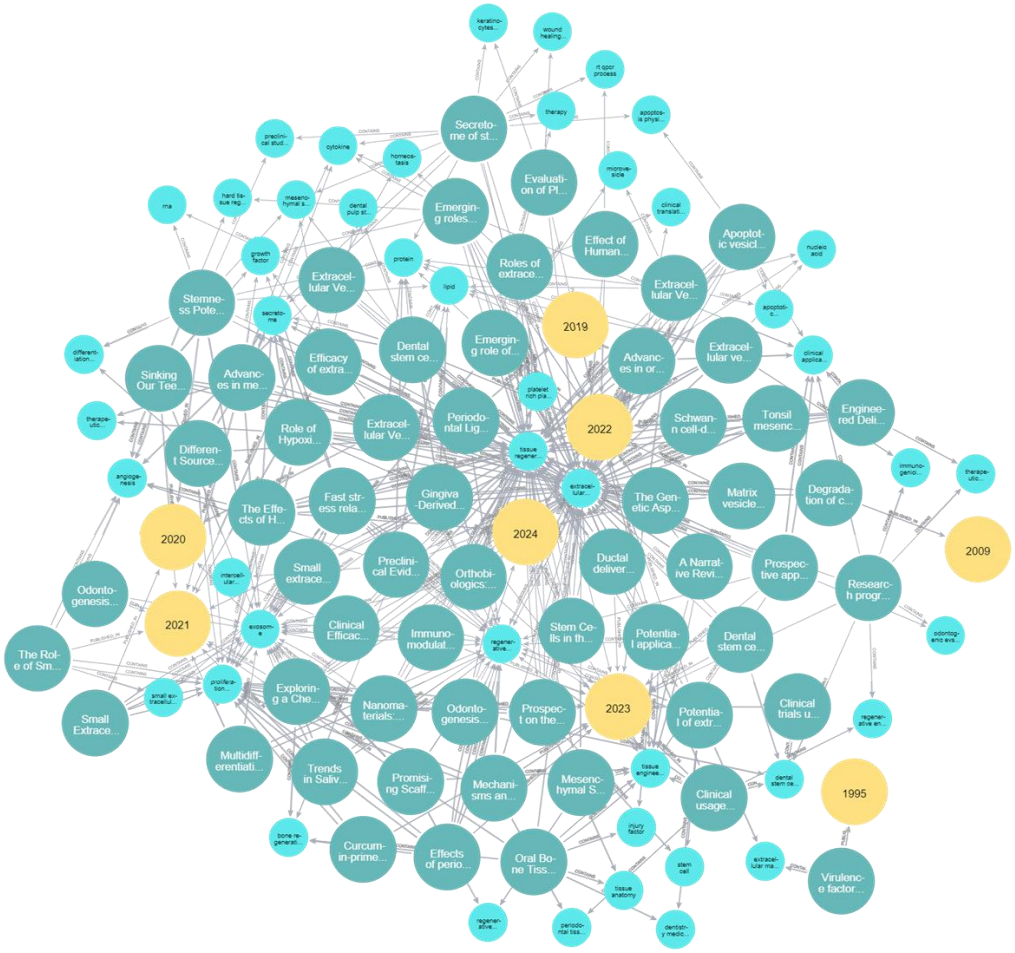

Figure 10. Neo4j Knowledge graph of generic "Unknown" category nodes (turquoise) connected to at least two article nodes (in green. A part of their title is written on the node). Yellow nodes indicate publication year.

4. Conclusions

The field of exosomes and extracellular vesicles for tissue regeneration is a fast-expanding and multi-faceted area of research, which has grown exponentially in the last few years. Instead of diving into the details of the biology or uses of exosomes, we adopted in the present report a macroscope approach, to try to understand the field in its complexity, at an overview level of its topics and trends. We believe that this approach, although not exhaustive can complement traditional reviews and prove very useful to assess how the field is developing. Our investigation has showed that Exosomes, their biology and their potential applications in tissue regeneration have captivated researchers from several science fields, with a net prevalence of stem cell studies and tissue healing both in skin, in the musculoskeletal system, and various organs, such as kidneys or the eye (Figure 3). When we focused our attention on bone regeneration, the two main theme area of orthopedics and dentistry were well represented, and topically well distinguishable into specific subsets. Dentistry (and periodontics in particular) and bone fractures appeared as the two booming research fields, with growth trends exceeding the remaining bone regeneration topics by at least twofolds (Table 3). We then attempted to investigate the main ideas that associated subsets of papers in the field, and conveniently focused on the topic that has been growing the most, i.e. topic #2 Dental Stem cell regeneration. To do that we extracted the main concepts from the abstracts of the articles in this subsets and the relations associating them, to create a knowledge graph, which represents the ideoscape of the field, i.e. the shared ideas around which most papers in the dataset gravitate. Although the results we report are necessarily only partial, and the results of specific queries we conducted, our data outline how articles in the last 3 years tend to cluster around concepts such as periodontal stem cells (Figure 8),

 doi:10.20944/preprints202407.2583.v1

investigating their cellular processes, such as proliferation, migration, and differentiation, associated to their clinical use (Figure 9), together with more general cell models such as mesenchymal stem cells (Figure 8).

Although the ideoscape of exosomes in bone regeneration was used through a macroscope approach for the present report, the methodology we are proposing can be effectively employed to search very specific queries, and investigate the field to get specific answers, and thus represents a flexible tool, which is mostly automated and require minimal handwork, to scan and get insights from any field in the literature.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Figure S1: Neo4j Disease nodes.

Author Contributions: Conceptualization, C.G. and S.G.; methodology, C.G.; formal analysis, C.G. and M.T.C.; data curation, S.G. and M.M.; writing—original draft preparation, M.T.C.; writing—review and editing, M.M.; supervision, S.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are available on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

Akre, P., Malu, R., Jha, A., Tekade, Y., & Bisen, W. (2023). Sentiment Analysis using Opinion Mining on Customer Review. *International Journal of Engineering and Management Research*, *13*(4), 41–44.

Appadurai, A. (1996). *Modernity at large: Cultural dimensions of globalization* (Vol. 1). U of Minnesota Press.

Arya, S. B., Collie, S. P., & Parent, C. A. (2024). The ins-and-outs of exosome biogenesis, secretion, and internalization. *Trends in Cell Biology*, *34*(2), 90–108.

Bassi, S. (2007). A primer on python for life science researchers. *PLoS Computational Biology*, *3*(11), e199.

Bennani-Smires, K., Musat, C., Hossmann, A., Baeriswyl, M., & Jaggi, M. (2018). Simple unsupervised keyphrase extraction using sentence embeddings. *ArXiv Preprint ArXiv:1801.04470*.

Börner, K., & Record, E. (2017). Macroscopes for making sense of science. In *Proceedings of the practice and experience in advanced research computing 2017 on sustainability, success and impact* (pp. 1–2).

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. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, *25*(11), 1422–1423.

Cook, D. A., Beckman, T. J., & Bordage, G. (2007). A systematic review of titles and abstracts of experimental studies in medical education: many informative elements missing. *Medical Education*, *41*(11), 1074–1081.

De Rosnay, J. (2014). *Le macroscope. Vers une vision globale*. Média Diffusion.

Egido-Moreno, S., Valls-Roca-Umbert, J., Céspedes-Sánchez, J. M., López-López, J., & Velasco-Ortega, E. (2021). Clinical efficacy of mesenchymal stem cells in bone regeneration in oral implantology. Systematic review and meta-analysis. *International Journal of Environmental Research and Public Health*, *18*(3), 894.

Francis, N., Green, A., Guagliardo, P., Libkin, L., Lindaaker, T., Marsault, V., Plantikow, S., Rydberg, M., Selmer, P., & Taylor, A. (2018). Cypher: An evolving query language for property graphs. *Proceedings of the 2018 International Conference on Management of Data*, 1433–1445.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *ArXiv Preprint ArXiv:2203.05794*.

Guizzardi, S., Colangelo, M. T., Mirandola, P., & Galli, C. (2023). Modeling new trends in bone regeneration, using the BERTopic approach. *Regenerative Medicine*, *18*(9), 719–734.

Gutiérrez, L., & Keith, B. (2019). A systematic literature review on word embeddings. *Trends and Applications in Software Engineering: Proceedings of the 7th International Conference on Software Process Improvement (CIMPS 2018) 7*, 132–141.

Han, P., Johnson, N., Abdal-Hay, A., Moran, C. S., Salomon, C., & Ivanovski, S. (2023). Effects of periodontal cells-derived extracellular vesicles on mesenchymal stromal cell function. *Journal of Periodontal Research*, *58*(6), 1188–1200. https://doi.org/10.1111/jre.13171

Han, P., Raveendran, N., Liu, C., Basu, S., Jiao, K., Johnson, N., Moran, C. S., & Ivanovski, S. (2024). 3D bioprinted small extracellular vesicles from periodontal cells enhance mesenchymal stromal cell function. *Biomaterials Advances*, *158*, 213770. https://doi.org/10.1016/j.bioadv.2024.213770

Harrell, C. R., Jovicic, N., Djonov, V., & Volarevic, V. (2020). Therapeutic use of mesenchymal stem cell-derived exosomes: from basic science to clinics. *Pharmaceutics*, *12*(5), 474.

Hartley, J. (2007). Planning that title: Practices and preferences for titles with colons in academic articles. *Library & Information Science Research*, *29*(4), 553–568.

Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G. De, Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., Ngomo, A.-C. N., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2022). Knowledge Graphs. *ACM Computing Surveys*, *54*(4), 1–37. https://doi.org/10.1145/3447772

Hong, H., Lin, C., Fang, M., Liu, J., Hsu, H.-C., Chang, C.-J., & Wang, H. (2024). Proteomic analysis of exosomal proteins associated with bone healing speed in a rat tibial fracture model. *Biomedical Chromatography : BMC*, *38*(5), e5846. https://doi.org/10.1002/bmc.5846

Honnibal, M., & Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *To Appear*, *7*(1), 411–420.

Hu, Y.-C., Zhang, X.-B., Lin, M.-Q., Zhou, H.-Y., Cong, M.-X., Chen, X.-Y., Zhang, R.-H., Yu, D.-C., Gao, X.-D., & Guo, T.-W. (2023). Nanoscale Treatment of Intervertebral Disc Degeneration: Mesenchymal Stem Cell Exosome Transplantation. *Current Stem Cell Research & Therapy*, *18*(2), 163–173. https://doi.org/10.2174/1574888X17666220422093103

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3). https://doi.org/10.1109/MCSE.2007.55

Irfan, D., Ahmad, I., Patra, I., Margiana, R., Rasulova, M. T., Sivaraman, R., Kandeel, M., Mohammad, H. J., Al-Qaim, Z. H., Jawad, M. A., Mustafa, Y. F., & Ansari, M. J. (2023). Stem cell-derived exosomes in bone healing: focusing on their role in angiogenesis. *Cytotherapy*, *25*(4), 353–361. https://doi.org/10.1016/j.jcyt.2022.08.008

Isamorad, F., ouhestani F, Aghandeh P, & Motamedian SR. (2023). Application of Periodontal Ligament Stem Cells in Periodontal Regeneration: A Systematic Review. *J Appl Biotechnol Rep*, *10*(3), 1055–1068.

Issa, B., Jasser, M. B., Chua, H. N., & Hamzah, M. (2023). A Comparative Study on Embedding Models for Keyword Extraction Using KeyBERT Method. *2023 IEEE 13th International Conference on System Engineering and Technology (ICSET)*, 40–45.

Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2022). A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(2), 494–514. https://doi.org/10.1109/TNNLS.2021.3070843

Jia, Z., Maggioni, M., Smith, J., & Scarpazza, D. P. (2019). Dissecting the NVidia Turing T4 GPU via microbenchmarking. *ArXiv Preprint ArXiv:1903.07486*.

Ju, R., Gao, X., Zhang, C., Tang, W., Tian, W., & He, M. (2024). Exogenous MSCs based tissue regeneration: a review of immuno-protection strategies from biomaterial scaffolds. *Journal of Materials Chemistry B*.

Kalluri, R., & LeBleu, V. S. (2020). The biology, function, and biomedical applications of exosomes. *Science*, *367*(6478). https://doi.org/10.1126/science.aau6977

Kong, Q., Wang, Y., Jiang, N., Wang, Y., Wang, R., Hu, X., Mao, J., & Shi, X. (2024). Exosomes as Promising Therapeutic Tools for Regenerative Endodontic Therapy. *Biomolecules*, *14*(3). https://doi.org/10.3390/biom14030330

Kučuk, N., Primožič, M., Knez, Ž., & Leitgeb, M. (2021). Exosomes engineering and their roles as therapy delivery tools, therapeutic targets, and biomarkers. *International Journal of Molecular Sciences*, *22*(17), 9543.

Lange, M., Babczyk, P., & Tobiasch, E. (2024). Exosomes: A New Hope for Angiogenesis-Mediated Bone Regeneration. *International Journal of Molecular Sciences*, *25*(10), 5204. https://doi.org/10.3390/ijms25105204

Li, X., Si, Y., Liang, J., Li, M., Wang, Z., Qin, Y., & Sun, L. (2024). Enhancing bone regeneration and immunomodulation via gelatin methacryloyl hydrogel-encapsulated exosomes from osteogenic pre-differentiated mesenchymal stem cells. *Journal of Colloid and Interface Science*, *672*, 179–199. https://doi.org/10.1016/j.jcis.2024.05.209

Liu, Q., Kusner, M. J., & Blunsom, P. (2020). A survey on contextual embeddings. *ArXiv Preprint ArXiv:2003.07278*.

Liu, Y., Zhang, Z., Ma, C., Song, J., Hu, J., & Liu, Y. (2024). Transplanted MSCs promote alveolar bone repair via hypoxia-induced extracellular vesicle secretion. *Oral Diseases*. https://doi.org/10.1111/odi.14982

Luo, D., Zhu, H., Li, S., Wang, Z., & Xiao, J. (2024). Mesenchymal stem cell-derived exosomes as a promising cell-free therapy for knee osteoarthritis. *Frontiers in Bioengineering and Biotechnology*, *12*, 1309946. https://doi.org/10.3389/fbioe.2024.1309946

Lv, S., Wang, G., Dai, L., Wang, T., & Wang, F. (2023). Cellular and Molecular Connections Between Bone Fracture Healing and Exosomes. *Physiological Research*, *72*(5), 565–574. https://doi.org/10.33549/physiolres.935143

Ma, S., Ma, B., Yang, Y., Mu, Y., Wei, P., Yu, X., Zhao, B., Zou, Z., Liu, Z., Wang, M., & Deng, J. (2024). Functionalized 3D Hydroxyapatite Scaffold by Fusion Peptides-Mediated Small Extracellular Vesicles of Stem Cells for Bone Tissue Regeneration. *ACS Applied Materials & Interfaces*, *16*(3), 3064–3081. https://doi.org/10.1021/acsami.3c13273

Man, K., Eisenstein, N. M., Hoey, D. A., & Cox, S. C. (2023). Bioengineering extracellular vesicles: smart nanomaterials for bone regeneration. *Journal of Nanobiotechnology*, *21*(1), 137. https://doi.org/10.1186/s12951-023-01895-2

Mathieu, M., Martin-Jaular, L., Lavieu, G., & Théry, C. (2019). Specificities of secretion and uptake of exosomes and other extracellular vesicles for cell-to-cell communication. *Nature Cell Biology*, *21*(1), 9–17. https://doi.org/10.1038/s41556-018-0250-9

McInnes, L. (2024, May 24). *DataMapPlot*. Https://Github.Com/TutteInstitute/Datamapplot/Blob/Main/Doc/Demo.Ipynb.

McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *J. Open Source Softw.*, *2*(11), 205.

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *ArXiv Preprint ArXiv:1802.03426*.

Mckinney, W. (2010). Data Structures for Statistical Computing in Python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 51–56).

Meng, F., Xue, X., Yin, Z., Gao, F., Wang, X., & Geng, Z. (2022). Research Progress of Exosomes in Bone Diseases: Mechanism, Diagnosis and Therapy. *Frontiers in Bioengineering and Biotechnology*, *10*. https://doi.org/10.3389/fbioe.2022.866627

Miller, J. J. (2013). Graph database applications and concepts with Neo4j. *Proceedings of the Southern Association for Information Systems Conference, Atlanta, GA, USA*, *2324*(36), 141–147.

Mukherjee, A., Bisht, B., Dutta, S., & Paul, M. K. (2022). Current advances in the use of exosomes, liposomes, and bioengineered hybrid nanovesicles in cancer detection and therapy. *Acta Pharmacologica Sinica*, *43*(11), 2759–2776.

Pan, Y., Li, Y., Dong, W., Jiang, B., Yu, Y., & Chen, Y. (2023). Role of nano-hydrogels coated exosomes in bone tissue repair. *Frontiers in Bioengineering and Biotechnology*, *11*, 1167012. https://doi.org/10.3389/fbioe.2023.1167012

Peng, C., Xia, F., Naseriparsa, M., & Osborne, F. (2023). Knowledge Graphs: Opportunities and Challenges. *Artificial Intelligence Review*, *56*(11), 13071–13102. https://doi.org/10.1007/s10462-023-10465-9

Puletic, M., Velikic, G., Maric, D. M., Supic, G., Maric, D. L., Radovic, N., Avramov, S., & Vojvodic, D. (2024). Clinical Efficacy of Extracellular Vesicle Therapy in Periodontitis: Reduced Inflammation and Enhanced Regeneration. *International Journal of Molecular Sciences*, *25*(11). https://doi.org/10.3390/ijms25115753

Qaiser, S., & Ali, R. (2018). Text mining: use of TF-IDF to examine the relevance of words to documents. *International Journal of Computer Applications*, *181*(1), 25–29.

Saif, H., Fernandez, M., He, Y., & Alani, H. (2014). *On stopwords, filtering and data sparsity for sentiment analysis of twitter*.

Sankaranarayanan, J., Lee, S. C., Kim, H. K., Kang, J. Y., Kuppa, S. S., & Seon, J. K. (2024). Exosomes Reshape the Osteoarthritic Defect: Emerging Potential in Regenerative Medicine-A Review. *International Journal of Stem Cells*. https://doi.org/10.15283/ijsc23108

Sarker, I. H. (2024). LLM potentiality and awareness: a position paper from the perspective of trustworthy and responsible AI modeling. *Discover Artificial Intelligence*, *4*(1), 40.

Song, X., Xu, L., & Zhang, W. (2023). Biomimetic synthesis and optimization of extracellular vesicles for bone regeneration. *Journal of Controlled Release : Official Journal of the Controlled Release Society*, *355*, 18–41. https://doi.org/10.1016/j.jconrel.2023.01.057

Tan, S. H. S., Wong, J. R. Y., Sim, S. J. Y., Tjio, C. K. E., Wong, K. L., Chew, J. R. J., Hui, J. H. P., & Toh, W. S. (2020). Mesenchymal stem cell exosomes in bone regenerative strategies—a systematic review of preclinical studies. *Materials Today Bio*, *7*, 100067.

Théry, C., Amigorena, S., Raposo, G., & Clayton, A. (2006). Isolation and characterization of exosomes from cell culture supernatants and biological fluids. *Current Protocols in Cell Biology*, *30*(1), 3–22.

Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. *Nature Medicine*, *29*(8), 1930–1940.

Torrecillas-Baena, B., Pulido-Escribano, V., Dorado, G., Gálvez-Moreno, M. Á., Camacho-Cardenosa, M., & Casado-Díaz, A. (2023). Clinical potential of mesenchymal stem cell-derived exosomes in bone regeneration. *Journal of Clinical Medicine*, *12*(13), 4385.

Trubiani, O., Pizzicannella, J., Caputi, S., Marchisio, M., Mazzon, E., Paganelli, R., Paganelli, A., & Diomede, F. (2019). Periodontal Ligament Stem Cells: Current Knowledge and Future Perspectives. *Stem Cells and Development*, *28*(15), 995–1003. https://doi.org/10.1089/scd.2019.0025

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., & Bright, J. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272.

Wang, C., Guo, S., Gu, Q., Wang, X., Long, L., Xiao, C., Xie, M., Shen, H., & Li, S. (2022). Exosomes: A promising therapeutic strategy for intervertebral disc degeneration. *Experimental Gerontology*, *163*, 111806. https://doi.org/10.1016/j.exger.2022.111806

Wang, L., Li, F., Wang, L., Wu, B., Du, M., Xing, H., & Pan, S. (2024). Exosomes Derived from Bone Marrow Mesenchymal Stem Cells Alleviate Rheumatoid Arthritis Symptoms via Shuttling Proteins. *Journal of Proteome Research*, *23*(4), 1298–1312. https://doi.org/10.1021/acs.jproteome.3c00697

Wang, S., Zhou, W., & Jiang, C. (2020). A survey of word embeddings based on deep learning. *Computing*, *102*, 717–740.

Wang, T., Zhou, Y., Zhang, W., Xue, Y., Xiao, Z., Zhou, Y., & Peng, X. (2024). Exosomes and exosome composite scaffolds in periodontal tissue engineering. *Frontiers in Bioengineering and Biotechnology*, *11*. https://doi.org/10.3389/fbioe.2023.1287714

Wang, X., & Thomsen, P. (2021). Mesenchymal stem cell–derived small extracellular vesicles and bone regeneration. *Basic & Clinical Pharmacology & Toxicology*, *128*(1), 18–36.

Wang, Z., Chen, J., Chen, J., & Chen, H. (2023). Identifying interdisciplinary topics and their evolution based on BERTopic. *Scientometrics*, 1–26.

Waskom, M. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, *6*(60). https://doi.org/10.21105/joss.03021

Wen, S., Huang, X., Ma, J., Zhao, G., Ma, T., Chen, K., Huang, G., Chen, J., Shi, J., & Wang, S. (2024). Exosomes derived from MSC as drug system in osteoarthritis therapy. *Frontiers in Bioengineering and Biotechnology*, *12*, 1331218. https://doi.org/10.3389/fbioe.2024.1331218

Xu, D. D., & Wu, S. B. (2014). An improved TFIDF algorithm in text classification. *Applied Mechanics and Materials*, *651*, 2258–2261.

Xu, G., Lu, X., Liu, S., Zhang, Y., Xu, S., Ma, X., Xia, X., Lu, F., Zou, F., Wang, H., Song, J., & Jiang, J. (2023). MSC-Derived Exosomes Ameliorate Intervertebral Disc Degeneration By Regulating the Keap1/Nrf2 Axis. *Stem Cell Reviews and Reports*, *19*(7), 2465–2480. https://doi.org/10.1007/s12015-023-10570-w

Xue, N., Ding, X., Huang, R., Jiang, R., Huang, H., Pan, X., Min, W., Chen, J., Duan, J.-A., & Liu, P. (2022). Bone tissue engineering in the treatment of bone defects. *Pharmaceuticals*, *15*(7), 879.

Zeng, W.-Y., Ning, Y., & Huang, X. (2021). Advanced technologies in periodontal tissue regeneration based on stem cells: Current status and future perspectives. *Journal of Dental Sciences*, *16*(1), 501–507.

Zhang, S., Wang, S., Chen, J., Cui, Y., Lu, X., Xiong, S., Yue, C., & Yang, B. (2024). Human dental pulp stem cell-derived exosomes decorated titanium scaffolds for promoting bone regeneration. *Colloids and Surfaces. B, Biointerfaces*, *235*, 113775. https://doi.org/10.1016/j.colsurfb.2024.113775

Zhang, Y., Jin, R., & Zhou, Z.-H. (2010). Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, *1*, 43–52.

Zou, J., Xia, H., Jiang, Q., Su, Z., Wen, S., Liang, Z., Ouyang, Y., Liu, J., Zhang, Z., Chen, D., Yang, L., & Guo, L. (2023). Exosomes derived from odontogenic stem cells: Its role in the dentin-pulp complex. *Regenerative Therapy*, *24*, 135–146. https://doi.org/10.1016/j.reth.2023.05.008