From Nodes to Themes: A Social Network Analysis and Thematic Progress in the field of Biomedical Ontologies

Elaheh Hosseini*✉, Code Orkid: 0000-0003-0417-7199
Department of Information Science, Faculty of Education & Psychology, Alzahra University, Tehran, Iran,
E.hosseini@alzahra.ac.ir
Maral Alipour-Tehrani, Code Orkid: 0009-0000-9276-4718
Department of Information Science, Faculty of Education & Psychology, Alzahra University, Tehran, Iran, M-
Alipour@student.alzahra.ac.ir
Hadi Zare Marzouni, Code Orkid: 0000-0002-2482-7227
Qaen Faculty of Medical Sciences, Birjand University of Medical Science, Birjand, Iran, Zaremh931@bums.ac.ir

Abstract—The paper aimed to analyze the thematic evolution and various networks of intellectual structures in the field of biomedical ontologies during 2014-2023. This applied research used an analytical and descriptive method, co-word techniques, and social network analysis. A web-based interface of bibliometrix, Microsoft Excel, and VOSviewer software were used for descriptive bibliometric study, data analysis, and network structure visualization. The period from mid-2020 to mid-2021 presented an increased dissemination of significant and prominent keywords within the overlay network in the field. Five major topic clusters were identified based on a co-occurrence network. These clusters labeled 'gene ontology', 'biomedical informatics focusing on AI techniques', 'bioinformatics applications in biomarker discovery', 'protein interaction networks in Alzheimer's proteomics', and 'network-based molecular mechanism'. Basic clusters were 'gene ontology', 'bioinformatics', and 'gene expression'. Moreover, five clusters experienced significant developments between 2023 and 2024, namely 'bioinformatics', 'deep learning', 'machine learning', 'transcriptome', and 'network pharmacology'. These topics are the latest and hottest concepts in this field. Clusters, namely 'deep learning',' machine learning, and 'ontology' were recognized as niche and the most well-developed themes. The most mature and mainstream thematic clusters were namely 'transcriptome', 'prognosis', and 'rna-seq'. The most undeveloped and chaotic themes were 'network pharmacology' and 'molecular docking'.

Keywords— Social Network Analysis; Thematic Evolution; Biomedical Ontologies; Biblioshiny; Bioinformatics; Gene ontology.

I. introduction

Ontologies, composed of explicitly defined concepts organized in trees or acyclic graphs [1], are increasingly applied in biomedical research and healthcare systems for data processing and they are utilized along with controlled vocabularies that enable human and machine-readable semantic descriptions [2]. Ontology is an integral part of ongoing efforts aimed at advancing conceptualizations and knowledge representations for problem solving. They can be considered as a tool to enhance the computational representation of concepts and their relationships, similar to advanced semantic networks [3].

Infectious diseases and pandemics have been captured and represented using biomedical ontologies [4]. In other words, biomedical ontologies serve a vital function in connecting terms with more overarching concepts [5]. They serve as the key to weaving together information on distinct model entities obtained from varying conditions and stored in diverse databases [6].

A comprehensive understanding of historical patterns, ongoing advancements, and the evolution of topics can foster awareness among policymakers and funders, enabling them to make informed and data-driven decisions. This kind of study establish a deep understanding of collaborative research and knowledge structures, and scientific approaches over time to analyze trends, knowledge mapping, thematic growth and maturity of the field in various thematic frontiers through multidimensional analysis. Moreover, it equips semantic developers to adopt a well-informed approach amidst the rapid and dynamic technological growth, facilitating the creation of predictive semantic tools for the future. Additionally, stakeholders and researchers can be familiar with influential topics in disseminating information. In terms of practical implications, collaboration among semantic scholars, ontology developers, semantic designers, and policy-makers can be facilitated by utilizing the outcomes as a thematic policy map. This collaborative effort aims to prevent redundant research and enable well-informed decision-making.

This study aims to analyze and visualize different social networks and explore the thematic growth and topical evolution of research frontiers in the field of biomedical ontologies from 2014 to 2023. This research utilized different methods and mapping tools simultaneously, such as co-word analysis, scientific mapping, social network analysis (SNA), network structure

visualization. Therefore, Biblioshiny, and VOSViewer were used to provide an appropriate, illustrative, systematic, and inclusive approach to the field of biomedical ontologies. It can be considered as the originality and novelty of the paper.

Results of the study provide a holistic view of thematic landscape of the field. This multidimensional analysis transparent the interplay between social networks, co-occurrence structures, and thematic content.

The following questions are used to conduct the research:

Q1: How is the intellectual structure of biomedical ontologies analyzed in terms of network structure and overlay visualization based on the co-occurrence algorithm by VOSViewer?

Q2: How are the clusters of biomedical ontologies in terms of the thematic map?

Q3: How are the clusters of biomedical ontologies in terms of thematic evolution?

II.   Literature Review

A bibliometric approach can facilitate the analysis and visualization of knowledge development, and there are numerous indications that this method can be employed effectively to explore the advancement of a field within a particular domain of knowledge [7], for example in [8], researchers used the Web of Science database to extract ontology articles from 1986 to 2020 with the method of bibliometric analysis. The findings showed that semantic web and gene ontology are prominent research topics in ontology studies, so that semantic web is related to natural language processing, while gene ontology is related to bioinformatics.

Also, some studies have been conducted in subfields related to knowledge graphs using bibliometric approaches or co-word analysis, such as, ontology, for example in [9], the study examines the outputs of global ontology research from 1900 to 2012 using bibliometric analysis methods, emphasizing its role in enhancing information understanding.

In [10] the overall goal is to depict the dynamic landscape and practical applications of ontologies in the evolving realm of information retrieval. Conducting co-occurrence analysis on the Web of Science database, this study examines the thematic content of ontology and its subfields. Practical methodologies, including taxonomy, co-words analysis, and SNA are utilized. By extracting 17,015 records of ontology essays spanning from 2000 to 2016, the research explores the thematic patterns and intra-relationships within ontology subfields. In [11] delves into the complexities of genome annotation using controlled vocabularies, with a specific focus on gene ontology. The text underscores the challenges associated with automating this annotation process, especially in disambiguating word meanings, as terms like *'development'* can carry diverse interpretations. The authors propose two strategies to tackle this issue: leveraging term co-occurrences and employing document clustering. Their method underwent evaluation on a dataset of 331 papers related to development and developmental biology, achieving a notable F-measure of 77%. The application of document clustering further enhances accuracy, reaching 82%, underscoring the practicality of this approach in automating genome annotation based on literature.

The main difference between the previous articles and the present study lies in several factors. According to the articles found in the literature review section, this field has not been investigated from such a perspective in the present research.

III.   Methodology

This applied study was conducted with a descriptive and analytical approach, scientometric techniques, co-word analysis, and SNA.

The research population included all the keywords extracted from all the documents about biomedical ontologies indexed in the WOS during 2014-2023. The following query as a researcher-made query was searched using an advanced search in the WOS Core Collection from Clarivate Analytics on 5th in December 2023. As a result, 9487 documents were retrieved. The researchers had authorized access to the WOS database through their institution. The WOS was chosen for its comprehensive coverage of interdisciplinary scope dating back to 1900 [12].

TS= (Biomedical ontolog* OR Biological ontolog* OR Biomedical Informatic* OR Biomedical Knowledge representation OR semantic biolog* OR semantic bioinformatics* OR biological semantic web OR biomedical semantic web)

Co-word analysis is a valuable technique for examining the connections among words in a document, offering valuable insights into the intellectual structure of a specific research field. This quantitative approach indicates the semantic relationships between words, frequently applied in scientometrics to present patterns of term co-occurrence in scientific research articles [13], [14]. An increased frequency of co-occurrence signifies a closer semantic relationship between the two words [15]. Through co-word analysis, emerging and well-established thematic clusters are depicted, offering valuable insights into potential directions for future research [16].

Bibliometric analysis offers a reliable method to categorize and quantitatively assess the bibliographic content within a particular scientific domain [17].

The R program and Biblioshiny, which is a web-based interface of bibliometrix, were used to produce a descriptive bibliometric study [18]. VOSviewer is a software tool designed to create various visual maps from network data, aiding in the recognition of connections among concepts within clusters [19]. SNA is a method used to study social structures and relationships between nodes [20] Some various measures and metrics can be used in SNA, such as centrality and density [21]. In the strategic diagram (SD), two axes are used to depict centrality and density. Specifically, the x-axis signifies centrality, while the y-axis corresponds to density. Consequently, the diagram is partitioned into four quadrants, each showcasing distinct levels of both centrality and density [22].

High centrality levels signify that a cluster holds a more significant position within the field. As depicted in Fig. 1., the first quadrant indicates mature clusters positioned at the field's center, characterized by high centrality and density. In contrast, the second quadrant features well-developed yet isolated clusters, exhibiting low centrality and high density. These clusters, while not central, have progressed well, indicating ongoing evolution. Conversely, clusters in the third quadrant exhibit low centrality and density, signaling emerging or declining themes with limited attention. They showcase a relatively fragmented and underdeveloped structure. Lastly, the fourth quadrant encompasses central clusters that are not fully developed, possessing high centrality but low density, indicating immaturity despite their central position within the field [22], [23].
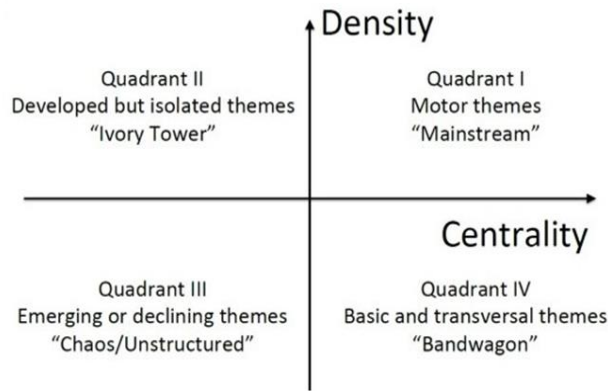
Fig. 1. Quadrants in a Strategic Diagram [22].

Creating thematic maps can be achieved using clustering algorithms, like the Walktrap clustering algorithm. This hierarchical clustering algorithm identifies communities or clusters within a network based on the premise that connected nodes are more likely to belong to the same cluster. The algorithm operates by progressively merging connected nodes until the desired level of clustering is attained. Thematic maps offer valuable insights into spatial patterns and trends within a specific theme or topic [24].

IV.   Results

Q1: How is the intellectual structure of biomedical ontologies analyzed in terms of network structure and overlay visualization based on the co-occurrence algorithm by VOSViewer?

The threshold value in the software to the test and error was considered to be ≥ 15 co-occurrence, which resulted in the formation of five main clusters of the 20301 keywords; of these, 310 meet the threshold, including 310 author keywords.

Fig. 2. depicts the network structure in the field of biomedical ontologies, including the keywords visualized by VOSviewer 1.6.19. The network consisted of five main clusters in different colors by using the algorithms and analyses of VOSviewer of total link strength of 12923, and 6325 total links.
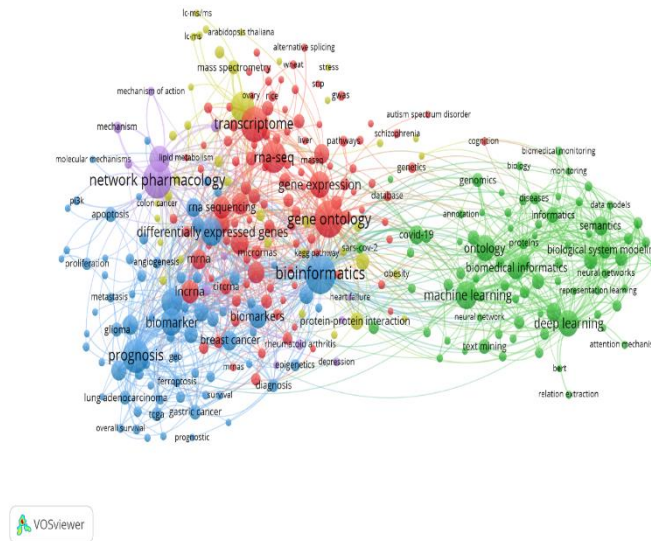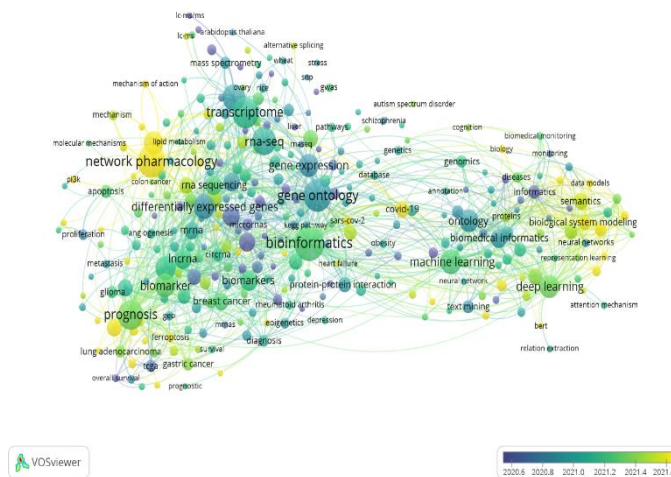

Fig. 2. Network visualization in the field of biomedical ontologies by Authors' keywords by VOSviewer software.

Fig. 3. illustrates the overlay visualization of the network in this field. The colors of this map were determined by their weight in the network. Blue has the lowest score, green indicates the average score, and yellow has the highest score. It means that movement from blue to yellow indicates more importance and weight due to the greater score and significance of the keyword in the network [19]. The year between the middle of 2020 till middle of 2021 includes more prominent and major keywords in the overlay network in the field. Table I. represents characteristics of the clusters based on keywords' co-occurrence.

FIG. 3. OVERLAY VISUALIZATION IN THE FIELD OF BIOMEDICAL ONTOLOGY BY AUTHORS' KEYWORDS BY VOSVIEWER SOFTWARE.

TABLE I. "CHARACTERISTICS OF THE CLUSTERS BASED ON KEYWORDS' CO-OCCURRENCE ANALYZED BY VOSVIEWER"

The first cluster: 106 keywords/ total co-occurrences: 4749/ links: 4270/ total link strength:7589	
gene ontology	*Top Five Keywords in the Cluster*
	gene ontology, transcriptome, rna-seq, gene expression, lncrna.
The second cluster: 78 keywords/ total co-occurrences: 3207/ links: 2985/ total link strength :7073	
biomedical informatics focusing on AI techniques	*Top Five Keywords in the Cluster*
	machine learning, deep learning, ontology, biomedical informatics, natural language processing, semantics.
The third cluster:75 keywords/ total co-occurrences:4013/ links: 3672/ total link strength:7800	
Bioinformatics Applications in Biomarker Discovery	*Top Five Keywords in the Cluster*
	Bioinformatics, prognosis, differentially expressed genes, biomarker, bioinformatics analysis
The fourth cluster:38keywords/ total co-occurrences:1238/ links: 1264/ total link strength :2022	
Protein Interaction Networks in Alzheimer's Proteomics	*Top Five Keywords in the Cluster*
	Proteomics, alzheimer's disease, protein-protein interaction, mass spectrometry, itraq.
The fifth cluster:18keywords/ total co-occurrences: 842/ links: 459/ total link strength:1362	
Network-based Mechanistic Insights	*Top Five Keywords in the Cluster*
	network pharmacology molecular docking traditional chinese medicine mechanism molecular mechanism

Q2: HOW ARE THE CLUSTERS OF BIOMEDICAL ONTOLOGIES IN TERMS OF THE THEMATIC MAP?

FIG. 4. SHOWS THE THEMATIC MAP OF AUTHORS' KEYWORDS INCLUDING 250 WORDS, 3 LABELS, AND 5 MIN CLUSTER FREQUENCY (PER THOUSAND DOCS), AND IS BASED ON THE WALKTRAP CLUSTERING ALGORITHM.

THE THEMATIC MAP INCLUDES FOUR QUADRANTS CONTAINING DIFFERENT DEGREES OF DENSITY AND CENTRALITY. IN OTHER WORDS, DENSITY AND CENTRALITY ARE INDICATORS OF THE SNA APPROACH. THE Y-AXIS STANDS FOR DENSITY, WHICH CAN MEASURE THE INTERNAL STRENGTH OF A CLUSTER. MOREOVER, THE X-AXIS INDICATES THE IMPORTANCE OF THE TOPIC AND REPRESENTS THE CENTRALITY. HIGH CENTRALITY (HIGH RELEVANCE DEGREE) MEANS THAT THE CLUSTER HAS A MORE IMPORTANT POSITION IN THE FIELD. AS IT IS EXPLAINED IN FIG. 1. AND SHOWN IN FIG. 4., QUADRANT II IMPLIES LOW CENTRALITY

AND HIGH DENSITY, WHICH REPRESENT NICHE THEMES. THEREFORE, CLUSTERS, NAMELY 'MACHINE LEARNING',' DEEP LEARNING', AND 'ONTOLOGY' ARE LOCATED IN THIS QUADRANT. THEY ARE NOT CENTRAL, ISOLATED BUT WELL-DEVELOPED THEMES IN THE FIELD OF BIOMEDICAL ONTOLOGIES.

MOREOVER, THEMES LOCATED IN QUADRANT IV INDICATE BASIC THEMES DUE TO HIGH CENTRALITY AND LOW DENSITY [23]. THEREFORE,' GENE ONTOLOGY', 'BIOINFORMATICS', AND 'GENE EXPRESSION' REPRESENT CENTRAL BUT NOT DEVELOPED CLUSTERS. QUADRANT I SIGNIFIES HIGH CENTRALITY AND HIGH DENSITY, WHICH REPRESENT MOTOR THEMES. THEREFORE, 'TRANSCRIPTOME', 'PROGNOSIS', AND 'RNA-SEQ' REPRESENT CENTRAL AND DEVELOPED CLUSTERS. QUADRANT III PRESENTS EMERGING OR DECLINING THEMES, NAMELY 'NETWORK PHARMACOLOGY' AND 'MOLECULAR DOCKING'.
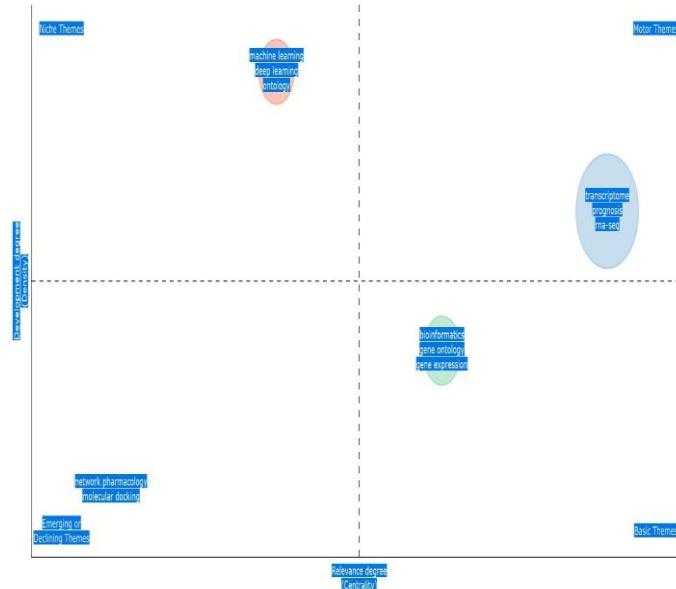

Fig. 4. Thematic Map of authors' keywords.

Q3: How are the clusters of biomedical ontologies in terms of thematic evolution?

An analysis of the evolution of the themes considering their keywords and evolution across time is shown in Fig. 5. It is depicted with authors' keywords, cutting point 2 (*cutting year 1 2019, cutting year 2 2021),* and the Walktrap clustering algorithm. It includes periods 2019-2020, 2021-2022, and 2023-2024 as the Sankey diagram.
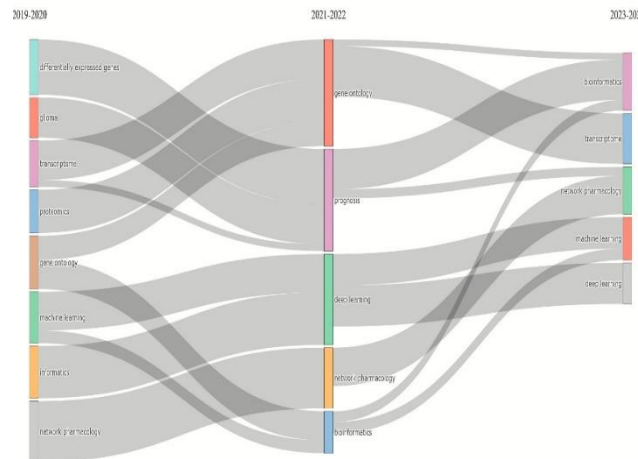

Fig. 5. Sankey Diagram of thematic Evolution of the research field of biomedical ontologies (2019-2024).

Some trends like *'differentially expressed genes', 'glioma', 'transcriptome', 'gene ontology', 'machine learning', 'informatics 'and 'network pharmacology'* were revealed as thematic growth in 2019-2020. An evolution was found to '*gene ontology, 'prognosis', 'network pharmacology, 'bioinformatics' and 'deep learning'* in 2021-20220. Moreover, five main thematic evolutions during 2023-2024 were identified as '*bioinformatics', 'transcriptome', 'network pharmacology', 'machine learning', and 'deep learning'*. According to it, the research in this field presents dramatic cohesion seeing that the detected themes are categorized within a discovered cluster that comes from the thematic map (Fig. 4.) like 'machine learning', 'deep learning', as niche themes and 'gene ontology' and 'bioinformatics' as basic and traversal theme. Besides, Table I. shows the flow of thematic evolution within words and occurrences.

V. DISCUSSION

This paper highlights outcomes and showcases findings from a topic-driven bibliometric analysis of academic papers on the subject of biomedical ontologies published from 2014-2023.

The mainstream topics of biomedical ontologies were located in Quadrant I and were defined as a motor clusters, namely the *'transcriptome', 'prognosis', and 'rna-seq'*. The concepts are considered the focal point of the field's topics. They are the most robust and mature clusters that are located in a central position in the field. The clusters contained the most extensive themes within this field.

Topics such as *'machine learning', 'deep learning', and 'ontology'* are located in quadrant II as niche themes. These themes are considered as ivory tower, not axial but well-developed, prominent, and isolated clusters. This kind of isolation is due to switching, maturation, and developing into higher-level subjects based on some trends or research frontiers. They may not be the most studied topics in the field of biomedical ontologies, but they have managed to evolve and advance distinctly. Their well-developed nature signifies that scholars find value in exploring these topics deeply [25]. Their placement in quadrant II highlights their distinctiveness and specialization within the field of biomedical ontologies.

Therefore, the professionals can concentrate on particular research areas such as developing feature ontologies, model interoperability ontologies, and develop explainable AI ontologies. They can make a special contribution to a comprehensive understanding of the diverse aspects of the field of biomedical ontologies.

Quadrant IV shows the basic, underdeveloped, immature, central, and transversal themes, e.g., *'gene ontology', 'bioinformatics', and 'gene expression'*. These themes will be extended more in the future as revolutionary ideas aiming to consider paradigms. They are considered foundational concepts within the domain. Their placement (quadrant IV) highlights their foundational nature and pivotal role in the field. These fundamental themes provide the conceptual infrastructure for fruitful knowledge representation. As a result, scholars in the field of biomedical ontologies can uncover potential topics for improvements and advancements through these basic themes.

Themes that are located in the second quadrant of the strategic diagram, namely *'network pharmacology' and 'molecular docking'* have strong internal relationships and a good level of maturity in this field. they not axial, but developing. The placement indicates that they are emerging topics and are known as chaos, marginal and unstructured themes.

The rise of certain trends, such as 'deep learning,' 'machine learning,' and 'ontology' (Fig. 5.) can be attributed to the abundance of extensive data and an increasing demand for methods and utilities to extract information and knowledge. Semantic developers and practitioners emphasized the importance of knowledge representation in a formal and machine-understandable manner, driven by these techniques. [26], [27], [28].

The evolution of *'gene ontology'* during 2019-2020 and 2021-2022 reflects the growth of gene-related knowledge focusing on gene ontology as a widely accepted semantic tool and a domain ontology for knowledge representation [29] in the fields of bioinformatics and biomedicine [30]. Gene Ontology (GO) is a widely used and very detailed ontology [48] and the most comprehensive in the field of bioinformatics that describes the functions of genes and gene products across different species [31].

The emphasis on *'protein'* in the cluster 4 in Table I. indicates a specific concentration on the modeling and integration of knowledge related to proteins, such as protein-protein interactions (PPIs). PPIs play a crucial role in biological processes and serve as a primary mechanism for functions and signaling in biological systems. The outcomes of certain machine learning approaches, including promising and informative results from deep learning techniques (sequence-based), were observed in the realm of PPIs [32].

Five major topic clusters were identified based on a co-occurrence network. These clusters labeled *'gene ontology', 'biomedical informatics focusing on AI techniques', 'bioinformatics applications in biomarker discovery', 'protein interaction networks in Alzheimer's proteomics', and 'network-based molecular mechanism'*. AI techniques play a critical role in advancing biomedical informatics by enhancing predictive modeling, data integration, and mining to discover patterns, trends, and associations that may not be apparent through traditional statistic methods [33].

Moreover, bioinformatics plays a significant role in biomarker discovery by leveraging computational tools, algorithms, and omics datasets to analyze biological data and uncover potential biomarkers [34]. Understanding the molecular mechanisms underlying Alzheimer's disease is possible through mapping protein interaction networks. Studying the interactions between proteins in Alzheimer's proteomics is helpful in identifying potential therapeutic targets and biomarkers [35]. Network-based molecular mechanisms try to analyze the intricate relationships between molecules and their functions in biological processes. On this way, researchers can uncover key regulatory pathways, identify critical nodes or hubs, and make ability to integrate data from multiple sources, such as genomics and proteomics [36]. As [49] Suggested that metadata managers and curators of molecular biology databases can take advantage of librarians' approach to develop systematic conceptualizations of authority metadata and task-based relationships.

VI. CONCLUSION

The study aimed to summarize recent progress and outline potential paths for future investigations in the field of biomedical ontologies. As expected, the interdisciplinary nature of biomedical ontologies positions it to continually introduce innovative approaches and applications in the close future. It is concluded that diverse findings obtained through various methods indicate substantial consistency within the realm of biomedical ontologies. This suggests that the identified themes and trends converge towards a shared direction or boundaries.

It is inferred that five clusters experienced significant developments between 2023 and 2024, namely *'bioinformatics', 'deep learning,' 'machine learning,' 'transcriptome,' and 'network pharmacology'*. These topics are the latest and hottest concepts in this field. The interlink between transcriptomics and proteomics with biomedical ontologies is crucial for organizing, annotating, standardizing the description of genes, proteins, and their functions, which can be used to annotate and integrate transcriptomic

and proteomic data, and integrating large biological data sets (omics data) generated by these fields, enabling researchers to better understand the underlying biology of diseases and identify potential targets for therapeutic intervention [37].

Machine learning and deep learning are powerful tools for digitizing metadata [46], generating metadata [47], analyzing and making predictions in transcriptome data analysis to help investigate gene expression, function and structure of data contained within biomedical ontologies. By leveraging these techniques, researchers and clinicians can gain new insights into disease mechanisms, improve diagnostic accuracy, specific biological processes and develop more effective treatments, and disease progression [38]. Prediction of gene expression, prediction and classification of splicing, prediction of transcription factor binding sites, auxiliary diagnosis using transcriptomics, prediction protein-protein interactions, and prediction nucleic acid-binding protein are applications of these techniques in transcriptome and proteomics [39], [40]. Moreover, by linking network pharmacology data to biomedical ontologies, researchers can leverage the rich semantics encoded in the ontologies to discover new relationships and patterns in their data. In addition, these ontologies can identify potential drug targets based on their biological functions or to predict drug-disease associations [41], [42]. Additionally, bioinformatics relies on ontologies and linked data approach to annotate data, guide data mining, and improve knowledge representation and discovery [43], [43], [50].

Five major topic clusters were identified based on a co-occurrence network. The Walktrap algorithm was utilized to identify 11 clusters within the network structure. Clusters, namely *'deep learning',' machine learning, and 'ontology'* were recognized as niche and the most well-developed themes. Additionally, basic clusters were *'gene ontology', 'bioinformatics', and 'gene expression'*. The most mature and mainstream thematic clusters were namely *'transcriptome', 'prognosis', and 'rna-seq'*. The most undeveloped and chaotic themes were *'network pharmacology' and 'molecular docking'*. They are emerging topics and are known as marginal and unstructured themes. Generally, the integration of molecular data with biomedical ontologies enhances the discoverability, interoperability, and reproducibility in the field of biomedicine [44], [45]. It is concluded that we need the development of advanced computational tools and algorithms for data mining, knowledge discovery, and predictive modeling in areas such as genomics and proteomics.

The study is subject to certain limitations, notably the constraint of utilizing data solely from the WOS Core Collection in the Clarivate Analytics database. Moreover, the investigation is confined to a specific time span, covering the years 2014 to 2023.

Results showed that the period from mid-2020 to mid-2021 presented an increased prevalence of significant and prominent keywords within the overlay network in the field.

Hence, it is advisable that upcoming research endeavors consider analyzing data from scholarly citation databases like Scopus and Dimensions over a longer period. Exploring topic trends through alternative topic modeling algorithms like Latent Dirichlet Allocation (LDA) and structure topic modelling (STM) is also recommended to unveil latent patterns. Furthermore, proposing the use of deep neural networks to predict topic trends, modeling past research, and forecasting future subjects is highly suggested.

Future studies should focus on analyzing topic clusters with an interdisciplinary approach in subfields of biomedical ontologies such as PPI, transcriptomics, proteomics, bioinformatics, and semantic bioinformatics.

REFERENCES

[1] A. Groß, C. Pruski, and E. Rahm, "Evolution of biomedical ontologies and mappings: Overview of recent approaches," *Computational and Structural Biotechnology Journal*, vol. 14, pp. 333–340, 2016.

[2] B. M. Konopka, "Biomedical Ontologies—A Review," *Biocybernetics and Biomedical Engineering*, vol. 35, no. 2, pp. 75-86, 2015.

[3] V. Maojo, J. Crespo, M. García-Remesal, D. de la Iglesia, D. Perez-Rey, and C. Kulikowski, "Biomedical ontologies: toward scientific debate," *Methods of Information in Medicine*, vol. 50, no. 3, pp. 203–216, 2011.

[4] L. Bayoudhi, N. Sassi, and W. Jaziri, "An Overview of Biomedical Ontologies for Pandemics and Infectious Diseases Representation," *Procedia Computer Science*, vol. 192, pp. 4249-4258, 2021.

[5] R. W. Filice and C. E. Kahn Jr, "Biomedical Ontologies to Guide AI Development in Radiology," *Journal of Digital Imaging*, vol. 34, no. 6, pp. 1331–1341, 2021.

[6] O. Bodenreider, J. A. Mitchell, and A. T. McCray, "Biomedical ontologies," *Pacific Symposium on Biocomputing*, pp. 76–78, 2005.

[7] K. Börner, C. Chen, and K. W. Boyack, "Visualizing knowledge domains," *Annu Rev Inf Sci Technol*, vol. 37, no. 1, pp. 179-255, 2003.

[8] A. Wu and Y. Ye, "Bibliometric Analysis on Bibliometric-based Ontology Research," *Science & Technology Libraries*, vol. 40, no. 4, pp. 435-453, 2021.

[9] Q. Zhu, X. Kong, S. Hong, J. Li, and Z. He, "Global ontology research progress: a bibliometric analysis," *Aslib Journal of Information*, 2015.

[10] N. Hormozi Nezhad and A. Koochak, "The survey of co-occurrence analysis method in the structural depiction of the scientific domain of ontology," *Qualitative and Quantitative Methods in Libraries (QQML)*, vol. 6, no. 3, pp. 495-507, 2019.

[11] B. Andreopoulos, D. Alexopoulou, and M. Schroeder, "Word Sense Disambiguation in biomedical ontologies with term co-occurrence analysis and document clustering," *International Journal of Data Mining and Bioinformatics*, vol. 2, no. 3, pp. 193-215, 2008.

[12] Clarivate, "Web of Science: Direct Links: Home," https://clarivate.libguides.com/directlinks, accessed April 5, 2023.

[13] X. Chen, J. Li, X. Sun, and D. Wu, "Early identification of intellectual structure based on co-word analysis from research grants," *Scientometrics*, vol. 121, pp. 349-369, 2019.

[14] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, and F. Herrera, "Science mapping software tools: Review, analysis, and cooperative study among tools," *J Am Soc Inf Sci Technol*, vol. 62, no. 7, pp. 1382-1402, 2011.

[15] J. Cho, "Intellectual structure of the institutional repository field: A co-word analysis," *J Inf Sci*, vol. 40, no. 3, pp. 386-397, 2014.

[16] R. Mokhtarpour and A. A. Khasseh, "Twenty-six years of LIS research focus and hot spots, 1990–2016: A co-word analysis," *J Inf Sci*, vol. 47, no. 6, pp. 794-808, 2021.

[17] M. K. Lazarides, I. Z. Lazaridou, and N. Papanas, "Bibliometric analysis: Bridging informatics with science," *Int J Lower Extrem Wounds*, 15347346231153538, 2023.

[18] M. Aria and C. Cuccurullo, "bibliometrix: An R-tool for comprehensive science mapping analysis," *J Informetrics*, vol. 11, no. 4, pp. 959-975, 2017.

[19] N. J. Van Eck and L. Waltman, "VOSviewer manual," 2020.

[20] R. Kumari, J. Y. Jeong, B. H. Lee, K. N. Choi, and K. Choi, "Topic modeling and social network analysis of publications and patents in humanoid robot technology," *J Inf Sci*, vol. 47, no. 5, pp. 658-676, 2021.

[21] A. Isfandyari-Moghaddam, M. K. Saberi, S. Tahmasebi-Limoni, S. Mohammadian, and F. Naderbeigi, "Global scientific collaboration: A social network analysis and data mining of the co-authorship networks," *Journal of Information Science*, vol. 49, no. 4, 2021.

[22] C. P. Hu, J. M. Hu, S. L. Deng, and Y. Liu, "A co-word analysis of library and information science in China," *Scientometrics*, vol. 97, no. 2, pp. 369-382, 2013.

[23] A. Khasseh, F. Soheili, H. Moghaddam, and A. Mousavi Chelak, "Intellectual structure of knowledge in iMetrics: A co-word analysis," *Information Processing & Management*, vol. 53, pp. 705-720, 2017.

[24] M. Brusco, D. Steinley, and A. L. Watts, "A comparison of spectral clustering and the walktrap algorithm for community detection in network psychometrics," *Psychological Methods*, 2022.

[25] E. Hosseini, S. Jafari, K. Taghizadeh Milani, S. A. Miran Fashandi, and T. Nikolaidis, "Thematic Clusters in the field of Gas Turbine Thermal Management: A Co-word Analysis during a Century," *International Journal of Information Science and Management (IJISM)*, vol. 21, no. 1, pp. 39-74, 2023.

[26] X. Shu and Y. Ye, "Knowledge Discovery: Methods from data mining and machine learning," *Social Science Research*, vol. 110, p. 102817, 2023.

[27] J. Hirschberg and C. D. Manning, "Advances in natural language processing," *Science*, vol. 349, no. 6245, pp. 261-266, 2015.

[28] M. V. Mäntylä, D. Graziotin, and M. Kuutila, "The evolution of sentiment analysis—A review of research topics, venues, and top cited papers," *Computer Science Review*, vol. 27, pp. 16-32, 2018.

[29] A. Sattar, E. S. M. Surin, M. N. Ahmad, M. Ahmad, and A. K. Mahmood, "Comparative analysis of methodologies for domain ontology development: A systematic review," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 5, 2020.

[30] M. Pomaznoy, B. Ha, and B. Peters, "GOnet: a tool for interactive Gene Ontology analysis," *BMC Bioinformatics*, vol. 19, no. 1, pp. 1-8, 2018.

[31] Gene Ontology Consortium, "The gene ontology resource: 20 years and still GOing strong," *Nucleic acids research*, vol. 47, no. D1, pp. D330-D338, 2019.

[32] F. Yang, K. Fan, D. Song, and H. Lin, "Graph-based prediction of protein-protein interactions with attributed signed graph embedding," *BMC Bioinformatics*, vol. 21, no. 1, pp. 1-16, 2020.

[33] D. K. Nayak, P. Mishra, P. Das, A. R. Jamader, and B. Acharya, "Application of Deep Learning in Biomedical Informatics and Healthcare," Smart Healthcare Analytics: State of the Art, *Intelligent Systems Reference Library*, vol. 213, pp. 113-132, 2022.

[34] L. Li and Z. P. Liu, "Biomarker discovery from high-throughput data by connected network-constrained support vector machine," *Expert Systems with Applications*, vol. 226, p. 120179, 2023.

[35] Q. Zhang, C. Ma, M. Gearing, P. G. Wang, L. S. Chin, and L. Li, "Integrated proteomics and network analysis identifies protein hubs and network alterations in Alzheimer's disease," *Acta Neuropathologica Communications*, vol. 6, pp. 1-19, 2018.

[36] M. S. Alam, M. M. Rahaman, A. Sultana, G. Wang, and M. N. H. Mollah, "Statistics and network-based approaches to identify molecular mechanisms that drive the progression of breast cancer," *Computers in Biology and Medicine*, vol. 145, p. 105508, 2022.

[37] C. Manzoni, D. A. Kia, J. Vandrovcova, J. Hardy, N. W. Wood, P. A. Lewis, and R. Ferrari, "Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences," *Briefings in bioinformatics*, vol. 19, no. 2, pp. 286-302, 2018.

[38] K. Huang, C. Xiao, L. M. Glass, C. W. Critchlow, G. Gibson, and J. Sun, "Machine learning applications for therapeutic tasks with genomics data," *Patterns*, vol. 2, no. 10, 2021.

[39] R. Li, L. Li, Y. Xu, and J. Yang, "Machine learning meets omics: applications and perspectives," *Briefings in Bioinformatics*, vol. 23, no. 1, p. bbab460, 2022.

[40] L. Cao, P. Liu, J. Chen, and L. Deng, "Prediction of Transcription Factor Binding Sites Using a Combined Deep Learning Approach," *Frontiers in Oncology*, vol. 12, p. 893520, 2022.

[41] M. Amith, Z. He, J. Bian, J. A. Lossio-Ventura, and C. Tao, "Assessing the practice of biomedical ontology evaluation: Gaps and opportunities," *Journal of biomedical informatics*, vol. 80, pp. 1-13, 2018.

[42] P. Xuan, Y. Song, T. Zhang, and L. Jia, "Prediction of potential drug–disease associations through deep integration of diversity and projections of various drug features". International Journal of Molecular Sciences, vol. 20,p 4102, 2019.

[43] A. Schuurman and A. Leszczynski, "Ontologies for bioinformatics," Bioinformatics and biology insights, vol. 2, pp. BBI-S451, 2008.

[44] J. Vamathevan, R. Apweiler, and E. Birney, "Biomolecular data resources: Bioinformatics infrastructure for biomedical data science," Annual Review of Biomedical Data Science, vol. 2, pp. 199-222, 2019.

[45] H. Mi and P. D. Thomas, "Ontologies and standards in bioscience research: for machine or for human," Frontiers in Physiology, vol. 2, p. 5, 2011.

[46] D. Smith-Glaviana, W.N. Ng, C. Miller, and J. Spencer, "Digitizing Metadata of a University Fashion Collection's Holdings Using OCR and Costume Core," Journal of Library Metadata, vol. 24, no. 2, pp. 57–86, 2024.

[47] S. M. Luke, M.J. Han, and T. Stratton, "Changes in Digital Collections and Their Metadata: A Longitudinal Study of UIUC Digital Library," Journal of Library Metadata, vol. 24, no. 3, pp. 147–164, 2024.

[48] O. Brandt, H. Gauza, J. Kaltenbach, M. E. Müller, G. Schneider, and C. Zinn, "A Minimal Metadata Schema and Its Tool to Improve the Searchableness of Research Data in Bioinformatics," Journal of Library Metadata, vol. 24, no. 3, pp. 165–188, 2024.

[49] S. Wu, B. Stvilia and D.J. Lee, Authority control for scientific data: The case of molecular biology: In Functional Future for Bibliographic Control . Routledge, pp. 9-30, 2017.

[50] B. Bushman, D. Anderson, and G. Fu, "Transforming the Medical Subject Headings into Linked Data: Creating the Authorized Version of MeSH in RDF," Journal of Library Metadata, vol. 15, no. 3–4, pp. 157–176, 2015.

[51]

Elaheh Hosseini   is an Assistant Professor in the Department of Information Science at Alzahra University in Tehran, Iran, which is part of the Faculty of Education & Psychology. Her research covers a wide range of topics, including knowledge graphs, linked data, the semantic web, knowledge retrieval, scientometrics, social network analysis, trend analysis, and artificial intelligence (AI). She is particularly interested in how AI intersects with knowledge representation and reasoning.   **ORCID: https://orcid.org/0000-0003-0417-7199**

Maral Alipour Tehrani is a senior student studying Information Science. She is very interested in how artificial intelligence (AI) can be used to improve services, especially in digital libraries, to make them more user-friendly. She also loves exploring AI tools and how they can help in research, in various context such as bioinformatics, social science, and knowledge organization. Her goal is to see how these tools can make studying and research easier and more effective. **ORCID: https://orcid.org/0009-0000-9276-4718**

Hadi Zare Marzouni is a faculty member at the Qaen Faculty of Medical Sciences, affiliated with Birjand University of Medical Sciences in Iran. His research interests include gene expression, bioinformatics, RNA sequencing, and transcriptomics. He is particularly interested in exploring the intersections between artificial intelligence (AI) and bioinformatics, such as using AI methods to analyze genetic data, predict gene expression, and interpret transcriptomic data. **ORCID: https://orcid.org/0000-0002-2482-**7227