Identifying Emerging Research Topics in Computer Science using Overlapping Community Detection on Graph Neural Network Predicted Graphs

Constantinos Djouvas, Christos Christodoulou, Antonis Charalampous
Dept. of Communication and Internet Studies
Cyprus University of Technology
Limassol, Cyprus
{costas.tziouvas, ci.xristodoulou, ag.charalampous}@cut.ac.cy
0000-0003-1215-7294, 0000-0002-4846-8353, 0000-0001-9665-9675

Nicolas Tsapatsoulis
Dept. of Marketing and Communication
Cyprus University of Technology)
Limassol, Cyprus
nicolas.tsapatsoulis@cut.ac.cy
0000-0002-6739-8602

2024 IEEE Conference on Pervasive and Intelligent Computing (PICom) | 979-8-3315-2274-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/PICOM64201.2024.00028

***Abstract*—Identifying emerging research topics is a challenging task. This study presents a novel methodology for the future prediction of emerging research topics, employing Graph Neural Networks (GNNs). The proposed framework is based on the construction of co-keyword graphs, serving as input for our innovative machine learning model designed to forecast keyword frequencies in forthcoming time periods. The proposed methodology is validated by applying it to forecast emerging topics and comparing the results against an observed year, 2022 in our case. Our findings underscore the potential of employing sophisticated computational methods to uncover emergent themes and offer valuable information for researchers, policymakers, and practitioners seeking to anticipate and navigate the dynamics of contemporary research.**

***Index Terms*—Machine Learning, Graph Neural Networks, Research Trends, Network Analysis, Community Detection**

I. Introduction

The identification of emerging research topics has attracted significant attention from governments, policy makers, and institutions, all with the aim of proactively predicting future trends and laying the foundation for the development and success of potential research fields [1]. To achieve this goal, various methodological approaches have been developed over the years to detect emerging research themes within the scientific literature. These methods often depend on a mixture of scientometric and bibliometric properties [2]–[5].

A common feature among most case studies is that the identification of topics is done retrospectively. This means that subjects are chosen based on the analysis of data up to the year they were collected, with few exceptions [1], [6]. This temporal limitation presents a significant challenge, as the potential of identified emergence topics can change considerably over time, rendering implemented policies or granted research projects obsolete prematurely.

To address the limitation discussed earlier, our research presents a novel methodology that utilizes recent breakthroughs in Machine Learning. By treating the problem as a time series forecasting task and utilizing Graph Neural Networks (GNNs), the proposed methodology has the potential to identify future emerging research topics. This approach is operationalized by constructing time-sliced co-keyword networks based on the keywords provided by authors for each publication. These networks are then fed into a Graph Neural Network (GNN), which predicts future network characteristics. By analyzing the predicted network using various network/graph analysis techniques, we can identify and prioritize emerging research topics, enabling policymakers to make well-informed decisions.

The rest of the paper is structured as follows. First, in sections II and III, we, respectively, introduced some of the key theoretical and some technical concepts associated with emerging topic identification. Then, in Section IV, we present our methodology, which encompasses four fundamental stages: data collection, network construction, model training, and emerging community detection along with the results derived by applying the methodology using different configurations. Finally, Section V, presents the results of the evaluation of the proposed methodology along with a real application of the proposed methodology for the forecasting of emerging research topics. Through this structured approach, our aim is to provide a comprehensive understanding of the theoretical underpinnings, methodological intricacies, and empirical outcomes associated with the identification of emerging research topics.

II. Background

In this section, we explore the core concepts and prior studies essential for contextualizing our research. Initially, we define the characteristics that represent an emerging topic. Following this, we present a concise summary of the fundamental principles that form the backbone of our methodological strategy. Finally, we examine the pertinent literature and related studies to highlight existing research and their relevance to our.

979-8-3315-2274-2/24/$31.00 ©2024 IEEE
DOI 10.1109/PICom64201.2024.00028

A. Defining emerging research topics

The term "emergence" is commonly used, but often lacks a precise definition [7]. Over time, various efforts have been made to define and operationalize the emergence through the use of specific *indicators* or *attributes*.

Based on a thorough review of scientific papers related to *emerging technologies*, Rotolo et al. [8] offered the following definition:

> [A] radically novel and relatively fast growing technology characterized by a certain degree of coherence persisting over time and with the potential to exert a considerable impact on the socio-economic domain(s) which is observed in terms of the composition of actors, institutions and patterns of interactions among those, along with the associated knowledge production processes. Its most prominent impact, however, lies in the future and so in the emergence phase is still somewhat uncertain and ambiguous.

This definition encompasses five key factors: radical novelty, swift growth, coherence, significant impact, and uncertainty. Although there is considerable overlap between emerging topics and emerging technologies, Wang et al. [9] refined this definition to better distinguish between the two, characterizing an emerging research topic as a "radically novel and relatively fast-growing research topic with coherence and significant scientific impact." For our research, we adopt this definition, highlighting *novelty*, *growth*, and *coherence*.

B. Geometric Deep Learning and Graph Neural Networks

Geometric deep learning seeks to extend deep neural networks to non-Euclidean spaces [10]. Graph Neural Networks (GNNs), a subset of this field, are designed to process graph-structured data and have attracted significant attention due to their versatility and the powerful representational capacity of graphs [11]. GNN models have been successfully applied across various domains, including knowledge graphs [12], fake news detection [13], physics simulations [14], antibacterial discovery [15], traffic prediction [16], and recommendation systems [17].

C. Community Detection

Identifying clusters of tightly interconnected nodes, known as community detection in graphs, is essential across various fields such as social network analysis, biology, and computer science [18]. The aim is to partition a graph into distinct communities or clusters based on their connection patterns. A range of algorithms tackles this task, from traditional ones like modularity optimization to more contemporary methods like spectral clustering and deep learning-based techniques. Grasping how nodes form communities provides insights into complex systems, network robustness, information dissemination, and pinpointing key or influential nodes. Our focus is on methods allowing nodes to belong to multiple communities. Specifically, we employed an implementation of the Speaker-Listener Label Propagation Algorithm (SLPA) [19], which facilitates the detection of overlapping communities.

III. Related work

Many studies that aim to identify emerging research themes in large datasets of academic publications usually follow a three-step procedural approach.

Initially, the creation of a scientific publications dataset occurs via the indexing of journals in systems like Web of Science (WoS) or SCOPUS. Concurrently, a preliminary data reduction process may be employed, concentrating on the elite tier of publications, typically represented by the top 1% [20], [21].

Next, a method for clustering the collected data into recognizable groups is devised based on a specified relational framework. Common categorization strategies include citation-based, text-based, and hybrid techniques. In text-based approaches, networks are created using word or phrase co-occurrences within specific document sections, such as the title [22]–[24]. For citation-based approaches, networks are constructed through direct citation links, where one paper cites another [25]–[27], or co-citation links, where two papers are cited together in a third document [2], [28]. Some research, such as [29], combines direct citation and co-citation relationships within their analytical models. Hybrid techniques that blend text-based and citation-based relationships to cluster publications are also mentioned [3], [20].

The concluding phase entails the formulation of emergence indicators, subsequently evaluated in relation to established definitions and characteristics. Given the diverse attributes highlighted in various studies, an array of indicators can be generated to examine these characteristics. These indicators are generally assessed through computational or conventional methods, or a blend of both [6].

Small et al. [28] targeted two primary attributes of emergence: novelty and growth. They created two models utilizing direct citation and co-citation analyses to quantify these attributes. Emerging topics were identified by merging these models through an "Emergence Potential" function, which rewards clusters demonstrating both attributes. This approach was implemented on the Scopus database from 1996 to 2010, and the Emergence Potential (EP) was computed to determine the top 25 emerging topics for each year from 2007 to 2010. The research concluded that the integration of direct citation clustering and co-citation threading is effective in detecting emerging topics, providing a comprehensive global outlook in contrast to more localized methods.

On the other hand, Huang et al. [30] adopted an alternate method by utilizing four metrics—novelty, growth, coherence, and impact—to detect potential emerging topics within a dynamic co-word network. This technique combined a link prediction model with machine learning methods. Time-sliced co-word networks were created, enabling the use of a back-propagation neural network to forecast the network's evolution by predicting connections between previously unlinked nodes.

This methodology was applied to an information science case study, showcasing its effectiveness. The same four indicators, along with 'relevance,' were utilized by [31] in their research on emerging topics in energy storage. Data processing involved a direct citation network, with clusters identified using the Louvain algorithm [32].

Finally, a methodology comparable to ours was adopted in [33], where Temporal Graph Neural Networks were used to forecast future keyword (i.e., node) frequencies in segmented co-word networks. In our study, we build on this approach by focusing on the prediction of network connections [34], allowing for a more nuanced and fluid analysis of the data.

IV. Methodology

The proposed methodology consists of four major stages: data collection, network construction, model training and prediction, and emerging community detection.

A. Data collection

For our data collection process, we utilized the Scopus Search API [35] to extract Computer Science (COMP) publications from the Scopus database from 2014 to 2021. After collecting the data, we focused on identifying the most influential papers. To accomplish this, we applied a filtering technique using the citation counts of the papers to select only the top 1% cited. By adopting this approach, we aimed to capture a subset of publications that had received substantial attention and recognition within the research community. Lastly, we used the Scopus Abstract Retrieval API to collect the titles and author keywords of the papers that were cited within each publication in our dataset. This additional step allowed us to augment the information available for each paper, allowing us to later apply a network community detection algorithm to identify emerging topics in Computer Science.

B. Network Construction

The construction of the network begins with the creation of a graph that formulates the correlation among the keywords of the papers. Each node (vertex) in the graph corresponds to a keyword, and two nodes are connected if they co-exist in a paper. This initial graph has no edge or node attributes; it only maintains the connections of keywords. Unlike similar approaches that rely on terms found in the title or abstract, the advantage of this approach is that these author-defined keywords already possess an inherent importance. This avoids potential shortcomings associated with such algorithms, which might otherwise overlook key terms.

Then yearly clones of this initial graph are created, augmented with one node attribute and one edge attribute. More precisely, nodes are added a frequency attribute counting the times the keyword appeared during the particular year, and edges are added a co-occurrence attribute counting the times the two keywords co-existed in all the papers of the dataset. These graphs are denoted as $G = \{G_{t(1)}, G_{t(2)}, ..., G_{t(n)}\}$, where the nodes of a network $G_{t(k)}$ are the keywords of all publications, while the edges are weighted by the frequency of terms co-occurrence. In case that a particular keyword does not appear in a year, its frequency and its corresponding edge co-occurrence are set to zero. We can formally define this graph as follows:

$$G = (V, E) \tag{1}$$

where V and E represent the sets of vertices and edges, respectively. Furthermore, we can define E_i as follows:

$$E_i = (v_j, v_k, w_i) \tag{2}$$

That is, E_i connects the vertices (nodes) *j* and *k* and its weight (i.e., the co-occurrence of the two keywords) it w_i. Finally, it is important to mention that the graphs described above are created using the NetworkX [36] Python library.

C. PyG Data Construction

As already mentioned, the main challenge of this work is to accurately predict future values of the co-occurrence attribute of graph edges. To do so, we used the PyTorch Geometric (PyG) [37] Python framework to train Graph Neural Network (GNN) capable of handling the nature of our data. Thus, an important step of the process is the transformation of the graphs created in the previous step into PyG data structures.

The conversion of NetworkX graphs to PyG data objects involves several steps to ensure compatibility with GNN operations. The node features and edge indices are tensorized, and additional mappings from the node indices back to their original identifiers are maintained to facilitate interpretability and reconstruction of the original graph.

In the following, we present the structure of the PyG data object:

$$D_i = (x, edge\_index, edge\_attr, node\_names) \tag{3}$$

where:

$$x = \begin{bmatrix} \mathrm{f}_1, & \mathrm{f}_2, & \ldots, & \mathrm{f}_n \end{bmatrix} \tag{4}$$

where f_i is the frequency of vertex i.

$$\text{edge\_index} = \begin{bmatrix} \text{source}_1, & \text{source}_2, & \ldots, & \text{source}_m \\ \text{target}_1, & \text{target}_2, & \ldots, & \text{target}_m \end{bmatrix} \tag{5}$$

where each pair $(source_i, target_i)$ represents the two vertices that edge i is connecting.

$$\text{edge\_attr} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} \tag{6}$$

where w_i is the weight of edge i (i.e., the co-occurrence of the two keywords it is connecting).

Finally, $node\_names$ is the list of the corresponding keywords of each vertex.

Having the above structure, the goal is to predict future values of the $edge\_attr$; At the same time, it carries all necessary information that allows recreating of the original graph.

D. Model Architecture

The architecture of the predictive model is based on a Graph Convolutional Network (GCN), designed to integrate the attributes of the nodes and edges effectively. The model comprises two primary GCN layers followed by an output layer that combines features for the final prediction. Formally, this is defined as follows:

$$X^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}X^{(l)}W^{(l)}\right) \tag{7}$$

$\tilde{A} = A + I$ represents the adjacency matrix A with added self-loops, $\tilde{D}$ is the degree matrix of $\tilde{A}$, $X^{(l)}$ denotes the node features at layer l, $W^{(l)}$ is the weight matrix for layer l, and σ is a non-linear activation function, specifically the ReLU function.

The GCN model is designed to first update the node representations via GCN layers and then concatenate these updated representations with log-scaled edge attributes for each edge. This concatenated vector forms the input to a linear layer that predicts the edge attribute, providing a holistic view of the graph's topology and edge-specific characteristics. Thus, Edge Features are defined as follows:

$$\text{Edge Features} = \text{Concat}\left(X_{\text{row}}, X_{\text{col}}, \log(1 + \text{edge\_attr})\right) \tag{8}$$

where X_{row} and X_{col} are the embeddings of the nodes connected by each edge, and $\log(1 + \text{edge\_attr})$ represents the logarithmic scaled edge attributes, enhancing the model's sensitivity to variations in edge weights.

E. Model Training

Model training involves minimizing the Mean Squared Error (MSE) between the predicted and true values of the edge attributes. The Adam optimizer is used for its efficient computation and adaptive learning rate capabilities. The model is trained iteratively, with each epoch involving a forward pass and a backward pass to adjust the model weights. The loss function $\mathcal{L}(\theta)$ is the following:

$$\mathcal{L}(\theta) = \frac{1}{|E|}\sum_{(i,j)\in E}\left(f(x_i, x_j; \theta) - y_{ij}\right)^2 \tag{9}$$

where $|E|$ is the number of edges, $f(x_i, x_j; \theta)$ is the prediction of the model for the edge between nodes i and j, and y_{ij} is the actual edge attribute.

Training is conducted over several epochs, with the model's performance on the test set evaluated at the end of each epoch to monitor progress and adjust training parameters as necessary.

F. Evaluation

The effectiveness of the model is quantified using the test Mean Squared Error (MSE), providing a measure of how accurately the model predicts new edge attributes. Additionally, predictions and actual values are extracted to assess the model's practical performance in predicting keyword co-occurrence.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{10}$$

where n is the number of test samples, $\hat{y}_i$ is the predicted edge attribute, and y_i is the actual edge attribute. Lower MSE values indicate better model performance, highlighting the predictive accuracy and reliability of the model.

$$\mathcal{L}(\theta) = \frac{1}{|E|}\sum_{(v_i,v_j)\in E}\left(f(v_i, v_j; \theta) - w_{ij}\right)^2 \tag{11}$$

where $\mathcal{L}(\theta)$ is the loss function, E is the set of edges, $f(v_i, v_j; \theta)$ is the predicted weight of the model and w_{ij} is the true weight of the data.

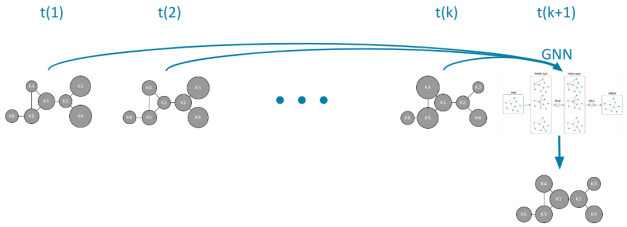


Fig. 1. Machine learning approach for graph based time-series forecasting

G. Emerging Community Detection

Since our model predicts the weight of each edge, that is, the co-occurrence of two keywords, we still have to define a process through which we can identify emerging research topics. To do so, we utilize the different graph snapshots that we have constructed using the observed and predicted edge attributes. This process unfolds in two main stages, shown in Figure 2 and Figure 3, respectively.

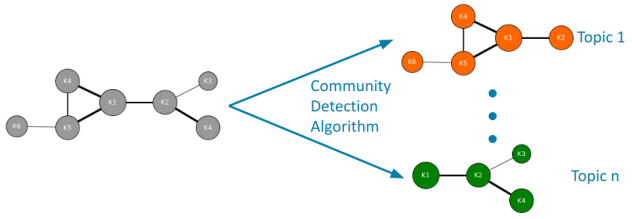


Fig. 2. Community detection

Initially, we apply a community detection algorithm to the predicted graph to map out the topics. Then we define an "emergence score" function, similar to [28], which is calculated for each extracted topic. The research topics with the highest scores are selected as emerging research topics. As already mentioned, for our community detection algorithm, we utilize the algorithm introduced in [19], a community detection method that iteratively updates the nodes' labels based on interactions with their neighbors, allowing identification of overlapping communities. By treating each node as both a

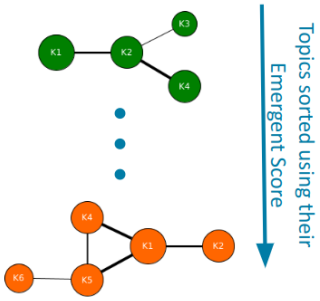


Fig. 3. Sorting of topic communities based on emergence score

speaker and a listener, SLPA effectively captures the complex structure of networks and reveals multiple community memberships for nodes.

Regarding the calculation of the "emergence score" of the N^{th} topic, we calculate it using the following formula:

$$ES_N = \frac{\sum_{t=k}^{k-3} G_t}{\sum_{t=k}^{k-3} (\mathbf{1}|\mathbf{G_t} \neq \mathbf{0})} \tag{12}$$

The above formula covers the "radically novel and relatively fast-growing research" aspect of Wang at all. [9] definition. G_t denotes the relative growth of the topic, that is, how much the topic at time $t(k+1)$ has grown (or shrinked) against the topic at time $t(k)$. Furthermore, the formula uses the last three graphs for estimating the growth, penalizing growths of the past, giving higher weight to fresh topics.

Finally, the growth between two consecutive graphs is calculated as a normalized pairwise difference of the weights (co-occurrences) of the corresponding edges. In other words, it captures how much a community grows (or shrinks). Formally, this is defined as follows:

$$G_t = \frac{\sum_{n \in N_k} F_{n_t}^{k+1} - F_{n_t}^{k}}{|N_k|} \tag{13}$$

where:

N_K are the nodes of topic K,

$F_{n_t}^{k}$ is the frequency of node n at t(k), i.e., last observed frequency,

$F_{n_t}^{k+1}$ is the frequency of node n at t(k+1), i.e., predicted frequency.

V. Evaluation & Results

This section elaborates on the implementation of the methodologies described in Section IV and presents the experimental results obtained from our predictive model. Utilizing the PyTorch Geometric framework (PyG), our architecture incorporates advanced techniques suited for temporal graph analysis, crucial for interpreting dynamic keyword co-occurrences in scientific publications.

Our model was extensively trained in 300 epochs, starting with an aggressive learning rate of 0.01, adjusted as necessary to optimize performance. The primary focus during training was the minimization of the Root Mean Squared Error (RMSE), a critical metric to gauge predictive accuracy in models producing continuous output.

Throughout the experimental phase, different configurations were tested by varying the amount of historical data used as input. Specifically, for predicting keyword co-occurrences over the next one year, seven years of historical data were utilized; for two years, six years of data were used; and for three years, five years of data were employed. This strategy was pivotal in determining the most effective data segmentation approach, which involved dividing the historical input into six-month chunks. Such segmentation significantly enhanced the model's capacity to identify subtle patterns within the data, thereby strengthening its learning and generalization framework.

The results are summarized as follows:

TABLE I
RMSE for Predicting Future Keyword co-occurrences

Prediction Duration	RMSE
Predicting next one year	0.11167
Predicting next two years	0.12793
Predicting next three years	0.18611

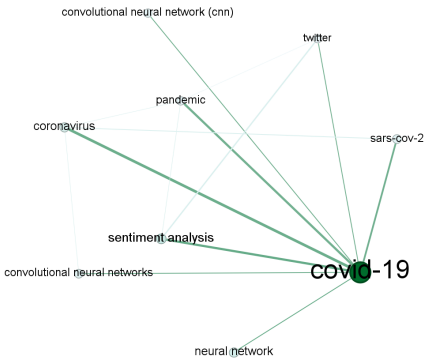
The lowest RMSE achieved, 0.11167, for one-year predictions underscores the model's high precision and reliability. In contrast, the increasing RMSE for two and three-year predictions highlights the challenges associated with longer-term forecasts in dynamic datasets. These outcomes not only demonstrate the model's robustness in handling complex predictive tasks but also validate the chosen architectural and data processing methodologies.

Having created a machine learning model that accurately predicts future states of the co-occurrences graph, the next step of the methodology is the detection of possibly overlapping communities, each one representing a topic in Computer Science. Sorting them using the "emergence score" refined in Section IV-G will reveal the most emerging research topics in Computer Science.

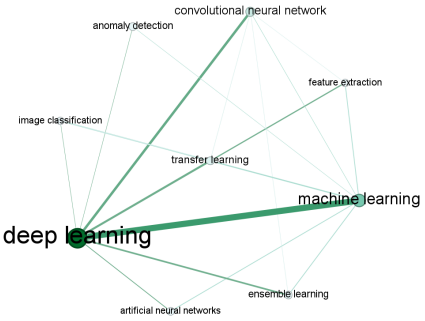
The co-occurrences graph was created using the NetworkX [36] library on which we applied the SLPA [19] implementation of the CDLIB [38] library for community detection. Then for each community, its emergence score is calulated; Figure 4 shows the top six topics created using the methodology described in Section IV.

VI. Conclusions and Future Work

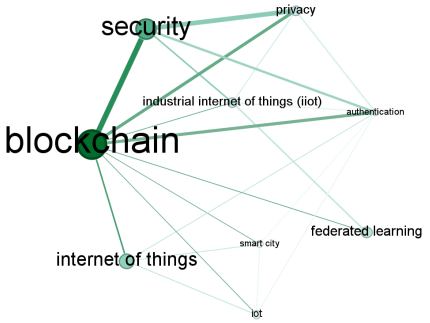
This research introduces an innovative method for detecting emerging research trends using Graph Neural Networks (GNNs). Through the combination of cutting-edge techniques in network analysis and neural network modeling, we have developed a promising framework that can predict emerging trends in academic research. Our results highlight the potential of using advanced computational methods to reveal developing research themes, providing valuable information to researchers, policy makers, and practitioners looking to foresee and manage the evolving landscape of research.



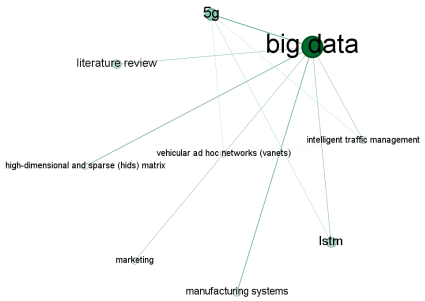
(a) Topic 1



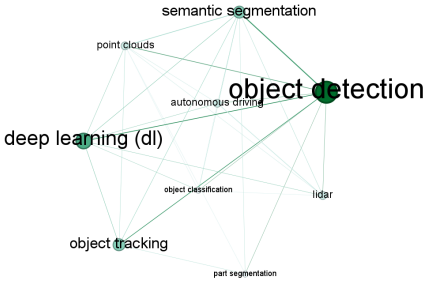
(b) Topic 2



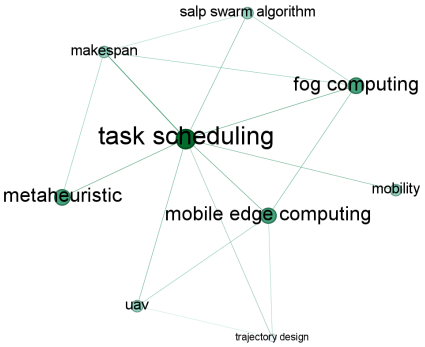
(c) Topic 3



(d) Topic 4



(e) Topic 5



(f) Topic 6

Fig. 4. The six most prominent topics represented as graphs

The experimental results confirm the robustness of the model architecture and training approach, showcasing its potential as a tool to improve the understanding of trends and patterns in scientific keyword usage. Future efforts will focus on expanding the applicability of the model to broader datasets and exploring additional predictive features to further improve its accuracy and effectiveness.

ACKNOWLEDGMENT

The work presented in the current study was carried out within the framework of the IBA-SwafS-Support-2-2020 project 'EUt- EXTRAS: European University of Technology - Experimentation to Transform Research Activities and Steering" under the contract H2020-IBA-SwafS-Support-2-2020/101035812.

REFERENCES

[1] S. Xu, L. Hao, X. An, G. Yang, and F. Wang, "Emerging research topics detection with multiple machine learning models," *Journal of Informetrics*, vol. 13, no. 4, p. 100983, 2019.

[2] K. W. Boyack, R. Klavans, H. Small, and L. Ungar, "Characterizing the emergence of two nanotechnology topics using a contemporaneous global micro-model of science," *Journal of Engineering and Technology Management*, vol. 32, pp. 147–159, 2014.

[3] W. Glänzel and B. Thijs, "Using 'core documents' for detecting and labelling new emerging topics," *Scientometrics*, vol. 91, no. 2, pp. 399–416, 2012.

[4] H. Guo, S. Weingart, and K. Börner, "Mixed-indicators model for identifying emerging research areas," *Scientometrics*, vol. 89, no. 1, pp. 421–435, 2011.

[5] H. Partaourides, E. Kouzaridi, N. Tsapatsoulis, and C. Djouvas, "Thematic modeling of un sustainable development goals: A comparative meta-based approach," in *IFIP International Conference on Artificial Intelligence Applications and Innovations*. Springer, 2023, pp. 557–568.

[6] C. W. Cai, M. K. Linnenluecke, M. Marrone, and A. K. Singh, "Machine learning and expert judgement: analyzing emerging topics in accounting and finance research in the asia–pacific," *Abacus*, vol. 55, no. 4, pp. 709–733, 2019.

[7] S. Cozzens, S. Gatchair, J. Kang, K.-S. Kim, H. J. Lee, G. Ordóñez, and A. Porter, "Emerging technologies: quantitative identification and measurement," *Technology Analysis & Strategic Management*, vol. 22, no. 3, pp. 361–376, 2010.

[8] D. Rotolo, D. Hicks, and B. R. Martin, "What is an emerging technology?" *Research policy*, vol. 44, no. 10, pp. 1827–1843, 2015.

[9] Q. Wang, "A bibliometric model for identifying emerging research topics," *Journal of the association for information science and technology*, vol. 69, no. 2, pp. 290–304, 2018.

[10] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: going beyond euclidean data," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, 2017.

[11] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI open*, vol. 1, pp. 57–81, 2020.

[12] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *The semantic web: 15th international conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, proceedings 15*. Springer, 2018, pp. 593–607.

[13] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake news detection on social media using geometric deep learning," *arXiv preprint arXiv:1902.06673*, 2019.

[14] A. Sanchez-Gonzalez, J. Godwin, T. Pfaff, R. Ying, J. Leskovec, and P. Battaglia, "Learning to simulate complex physics with graph networks," in *International conference on machine learning*. PMLR, 2020, pp. 8459–8468.

[15] J. M. Stokes, K. Yang, K. Swanson, W. Jin, A. Cubillos-Ruiz, N. M. Donghia, C. R. MacNair, S. French, L. A. Carfrae, Z. Bloom-Ackermann *et al.*, "A deep learning approach to antibiotic discovery," *Cell*, vol. 180, no. 4, pp. 688–702, 2020.

[16] M. Lv, Z. Hong, L. Chen, T. Chen, T. Zhu, and S. Ji, "Temporal multi-graph convolutional network for traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3337–3348, 2020.

[17] C. Eksombatchai, P. Jindal, J. Z. Liu, Y. Liu, R. Sharma, C. Sugnet, M. Ulrich, and J. Leskovec, "Pixie: A system for recommending 3+ billion items to 200+ million users in real-time," in *Proceedings of the 2018 world wide web conference*, 2018, pp. 1775–1784.

[18] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[19] J. Xie, B. K. Szymanski, and X. Liu, "Slpa: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process," in *2011 ieee 11th international conference on data mining workshops*. IEEE, 2011, pp. 344–349.

[20] H. Partaourides, E. Kouzaridi, N. Tsapatsoulis, and C. Djouvas, "On the identification of influential topics in the social sciences using citation analysis," in *2023 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*. IEEE, 2023, pp. 0845–0851.

[21] N. Tsapatsoulis, H. Partaourides, C. Christodoulou, and C. Djouvas, "Quo vadis computer science? the topics of the influential papers during the period 2014-2021," in *2022 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*. IEEE, 2022, pp. 1–8.

[22] R. Ohniwa, A. Hibino, and K. Takeyasu, "Trends in research foci in life science fields over the last 30 years monitored by emerging topics," *Scientometrics*, vol. 85, no. 1, pp. 111–127, 2010.

[23] T. Furukawa, K. Mori, K. Arino, K. Hayashi, and N. Shirakawa, "Identifying the evolutionary process of emerging technologies: A chronological network analysis of world wide web conference sessions," *Technological Forecasting and Social Change*, vol. 91, pp. 280–294, 2015.

[24] Y. Kajikawa and Y. Takeda, "Structure of research on biomass and biofuels: A citation-based approach," *Technological Forecasting and Social Change*, vol. 75, no. 9, pp. 1349–1359, 2008.

[25] Y. Kajikawa, J. Yoshikawa, Y. Takeda, and K. Matsushima, "Tracking emerging technologies in energy research: Toward a roadmap for sustainable energy," *Technological forecasting and social change*, vol. 75, no. 6, pp. 771–782, 2008.

[26] S. Iwami, J. Mori, I. Sakata, and Y. Kajikawa, "Detection method of emerging leading papers using time transition," *Scientometrics*, vol. 101, pp. 1515–1533, 2014.

[27] H. Small, "Tracking and predicting growth areas in science," *Scientometrics*, vol. 68, no. 3, pp. 595–610, 2006.

[28] H. Small, K. W. Boyack, and R. Klavans, "Identifying emerging topics in science and technology," *Research policy*, vol. 43, no. 8, pp. 1450–1467, 2014.

[29] C. Chen, "Citespace ii: Detecting and visualizing emerging trends and transient patterns in scientific literature," *Journal of the American Society for information Science and Technology*, vol. 57, no. 3, pp. 359–377, 2006.

[30] L. Huang, X. Chen, X. Ni, J. Liu, X. Cao, and C. Wang, "Tracking the dynamics of co-word networks for emerging topic identification," *Technological Forecasting and Social Change*, vol. 170, p. 120944, 2021.

[31] C. Mejia and Y. Kajikawa, "Emerging topics in energy storage based on a large-scale analysis of academic articles and patents," *Applied energy*, vol. 263, p. 114625, 2020.

[32] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[33] A. Charalampous, C. Djouvas, N. Tsapatsoulis, and E. Kouzaridi, "Emerging research topics identification using temporal graph neural networks," in *Artificial Intelligence Applications and Innovations*. Cham: Springer Nature Switzerland, 2024, pp. 192–205.

[34] Z. Wang, V. Baladandayuthapani, A. O. Kaseb, H. M. Amin, M. M. Hassan, W. Wang, and J. S. Morris, "Bayesian edge regression in undirected graphical models to characterize interpatient heterogeneity in cancer," *Journal of the American Statistical Association*, vol. 117, no. 538, pp. 533–546, 2022.
[35] M. E. Rose and J. R. Kitchin, "pybliometrics: Scriptable bibliometrics using a python interface to scopus," *SoftwareX*, vol. 10, p. 100263, 2019.
[36] A. Hagberg, P. Swart, and D. S Chult, "Exploring network structure, dynamics, and function using networkx," Los Alamos National Lab.(LANL), Los Alamos, NM (United States), Tech. Rep., 2008.
[37] M. Fey and J. E. Lenssen, "Fast graph representation learning with pytorch geometric," *arXiv preprint arXiv:1903.02428*, 2019.
[38] G. Rossetti, L. Milli, and R. Cazabet, "Cdlib: a python library to extract, compare and evaluate communities from complex networks," *Applied Network Science*, vol. 4, no. 1, pp. 1–26, 2019.