https://doi.org/10.58086/yv38-0307

ARTIFICIAL INTELLIGENCE–BASED SUPER NODES FOR REAL-TIME THREAT DETECTION IN DISTRIBUTED ENVIRONMENTS BIBLIOMETRIC ANALYSIS

José Lopes[1]✉, **Mário Dias Lousã**[1,2], **José Carlos Morais**[1,3]

[1] Instituto Superior Politécnico Gaya (ISPGAYA), Portugal.
[2] Insight - Piaget Research Center for Ecological Human Developement, Portugal.
[3] CEOS.PP, ISCAP, Polytechnic of Porto, Portugal.
✉ Corresponding authors**:** ispg2021101231@ispgaya.pt

Abstract

The widespread adoption of distributed systems, driven by the growth of the Internet of Things (IoT), edge computing, and cloud infrastructure, has substantially expanded the attack surface of modern digital ecosystems. These environments, characterized by high heterogeneity, large data volumes, and stringent latency requirements, make real-time threat detection a complex task. Traditional, predominantly centralized security mechanisms reveal clear limitations in scalability and response time in the face of increasingly dynamic attack patterns. In this context, Artificial Intelligence (AI) and Machine Learning have emerged as essential enablers for more effective intrusion detection. At the same time, the concept of "super nodes" is gaining prominence: strategically positioned network elements with enhanced computational capabilities that act as intelligent intermediaries between edge devices and the central cloud. This study presents a bibliometric analysis of the use of AI-based super nodes for real-time threat detection. The analysis focuses on a sample of 300 publications indexed in the Lens.org database (2015–2025), selected according to the PRISMA 2020 guidelines. Through descriptive indicators and network analysis (such as keyword co-occurrence), research trends, thematic structures, and emerging directions in this field are identified.

Keywords: Artificial Intelligence; Machine Learning; IoT Security; Edge Intelligence; Bibliometric Analysis.

1. Introduction

Digital systems have undergone a profound transformation toward distributed architectures, supported by the proliferation of IoT devices, edge computing, and cloud. While these

technologies enable critical services, from healthcare to smart cities, they also introduce a substantial risk: a vastly expanded attack surface (Li et al., 2018; Conti et al., 2018).

The challenge lies in the nature of IoT and edge environments. They are heterogeneous, resource-constrained, and demand ultra-low latency. Traditional, centralized security models struggle here; they simply cannot scale or adapt fast enough to stop real-time threats (Roman et al., 2018; Zhou et al., 2019). The consequences are evident in the rise of devastating DDoS botnets and ransomware attacks targeting industrial systems (Shafiq et al., 2024; Singh et al., 2021; Vakulov, 2025).

In response, Artificial Intelligence (AI) and Machine Learning have emerged as essential tools. Moving beyond static rules, these models offer the ability to detect anomalies dynamically in large data streams (Al-Garadi et al., 2020; Goranin et al., 2024). However, software needs the right architecture. This is where 'super nodes' become relevant. Acting as powerful, intelligent intermediaries between the edge and the cloud, super nodes run AI algorithms locally to detect threats with minimal latency (Nguyen & Reddi, 2021; Zhou et al., 2019).

Yet, despite the clear synergy between AI and super nodes, scientific literature on the subject is scattered across engineering, security, and data science fields. This lack of cohesion makes it difficult to compare solutions or grasp the full scope of innovation (Goranin et al., 2024). To bridge this gap, this article provides the first comprehensive bibliometric analysis of AI-based super nodes for real-time threat detection, structuring existing knowledge and charting the course for future research.

2. Methodology

This research adopts a bibliometric approach with the aim of systematically and structurally analyzing the scientific production related to the use of AI-based super nodes for real-time threat detection in distributed environments. The choice of bibliometric methodology is justified by the need to synthesize a broad, interdisciplinary, and rapidly growing body of literature. This allows the identification of research trends, scientific collaboration structures, dominant thematic areas, and conceptual gaps that would be difficult to observe through traditional narrative reviews (Van Eck & Waltman, 2010).

The methodological design followed the PRISMA 2020 guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), widely recognized as a benchmark of

good practices for systematic and bibliometric reviews (Page et al., 2021; PeerJ Computer Science, 2024). The application of these guidelines allowed for the transparent and reproducible structuring of all phases of the study selection process, from the initial identification of records to the definition of the final sample. The complete process of identifying, screening, eligibility, and inclusion of studies is summarized in the flowchart presented in Figure 1.

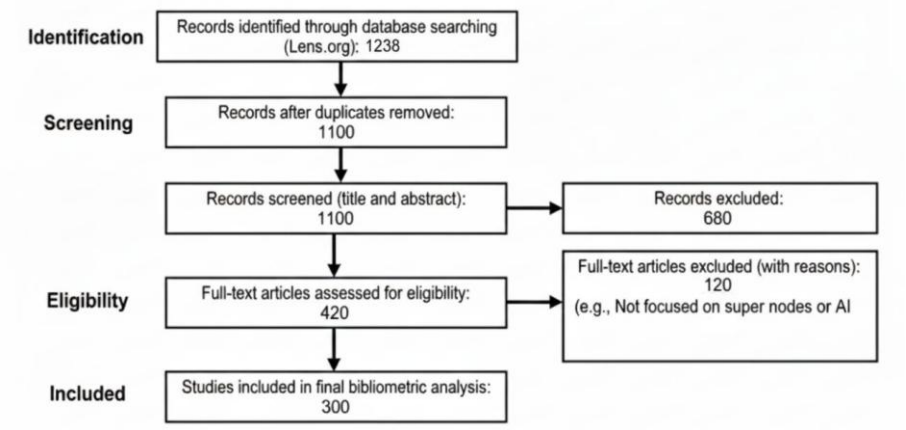

Fig 1. PRISMA 2020 flow diagram
Source: Authors' elaboration.

2.1. Data Source

The Lens.org database was selected as the primary source for collecting scientific literature. This choice is based on its broad interdisciplinary coverage, integrating relevant publications in the areas of cybersecurity, computer science, engineering, telecommunications, and distributed systems. Furthermore, Lens provides rich and structured bibliographic metadata, allowing the direct export of information compatible with bibliometric analysis tools, facilitating the replication of the study. When compared to commercial databases such as Scopus or Web of Science, the Lens platform presents specific advantages for the context of this work, namely open access and strong representation of technical literature associated with the Internet of Things and edge computing. However, it is recognized that the exclusive use of a single database may introduce limitations in terms of coverage of certain publications, an aspect that is critically considered in section four (Discussion).

2.2. Search Strategy

The search strategy was designed to capture studies that combined three central dimensions of the domain under analysis: (i) the application of Artificial Intelligence or Machine Learning, (ii) the focus on threat or intrusion detection, and (iii) the use in distributed environments, including IoT, edge computing, and hierarchical network architectures.

For this purpose, a Boolean search structure was applied to titles, abstracts, and keywords, according to the following general scheme:

("artificial intelligence" OR "machine learning") AND

("threat detection" OR "intrusion detection") AND

("IoT" OR "edge computing" OR "distributed systems")

The period considered encompassed publications between 2015 and 2025, reflecting the significant growth of research in IoT and edge intelligence from the second half of the 2010s, as well as the maturation of deep learning techniques applied to cybersecurity. No geographical or publication type restrictions were imposed, provided the studies were relevant to the topic under analysis.

2.3. Inclusion and Exclusion Criteria

The inclusion and exclusion criteria were defined beforehand, with the aim of ensuring thematic coherence and methodological quality of the final sample. Studies applying Artificial Intelligence or machine learning techniques to the detection of threats, intrusions, or anomalies in distributed environments were included, as well as works describing architectures involving nodes with enhanced analytical capabilities, explicitly or implicitly associated with the concept of super nodes.

On the other hand, studies unrelated to cybersecurity, those exclusively addressing centralized architectures without a distributed component, or those presenting insufficient metadata for a consistent bibliometric analysis were excluded. The clear definition of these criteria contributed to reducing biases in the selection process and ensuring the scientific relevance of the sample analyzed.

2.4. Study Selection Process

The selection process followed the PRISMA workflow. The initial search resulted in the identification of 1,238 records. After removing duplicates, 1,100 publications remained for title and abstract analysis. At this stage, 680 studies were excluded due to lack of thematic relevance.

The remaining 420 articles were analyzed in full text, leading to the additional exclusion of 120 publications that did not fully meet the defined inclusion criteria, namely because they did not address AI techniques or distributed architectures compatible with the concept of super nodes. The final sample comprised 300 studies, forming the basis of the bibliometric analysis presented in this work, as illustrated in Figure 1.

2.5. Bibliometric Analysis Techniques and Tools

The bibliometric analysis was conducted in two complementary phases. In the first phase, descriptive indicators were extracted using the analytical tools provided by Lens.org, including the temporal evolution of publications, distribution by scientific sources, disciplinary areas, and countries of affiliation of the authors.

In the second phase, the VOSviewer software was used to construct bibliometric networks, namely keyword co-occurrence networks and co-authorship networks. For the generation of these networks, a minimum threshold of five occurrences per term or author and a minimum link strength of two were defined, ensuring the analytical relevance and readability of the visualizations (Van Eck & Waltman, 2010). The interpretation of the identified clusters was carried out in conjunction with the concept of super nodes, allowing the observed bibliometric trends to be related to the conceptual and technological evolution of the domain under study.

3. Bibliometric Results

This section presents the main results of the bibliometric analysis performed on the final sample of 300 publications, selected according to the process described in the Methodology section. The presentation of the results follows the same logic and numbering of the figures

in the original article, ensuring structural coherence and facilitating the reading and validation of the data. The results are organized into different analytical dimensions, including thematic analysis of the literature, patterns of scientific collaboration, publication sources, geographical distribution of research, and temporal evolution associated with scientific impact.

3.1. Keyword Co-occurrence and Thematic Clusters

Keyword co-occurrence analysis allowed us to identify the main research themes and understand the conceptual structure of the domain under study. The keyword network, shown in Figure 2, highlights the existence of several thematic clusters, reflecting different lines of research related to threat detection based on Artificial Intelligence in distributed environments.

The dominant cluster is centered on terms such as “machine learning”, “deep learning”, “intrusion detection”, and “anomaly detection”, representing the technical core of the domain. These concepts reflect the strong reliance on machine learning techniques for detecting malicious behavior. A second cluster aggregates terms related to IoT, edge computing, distributed systems, and fog computing, highlighting the growing concern about security in distributed environments with limited computing resources. The interconnection between these clusters suggests the emergence of hybrid architectures, in which super nodes assume an intermediary role of aggregation and intelligent analysis between edge devices and cloud infrastructures.

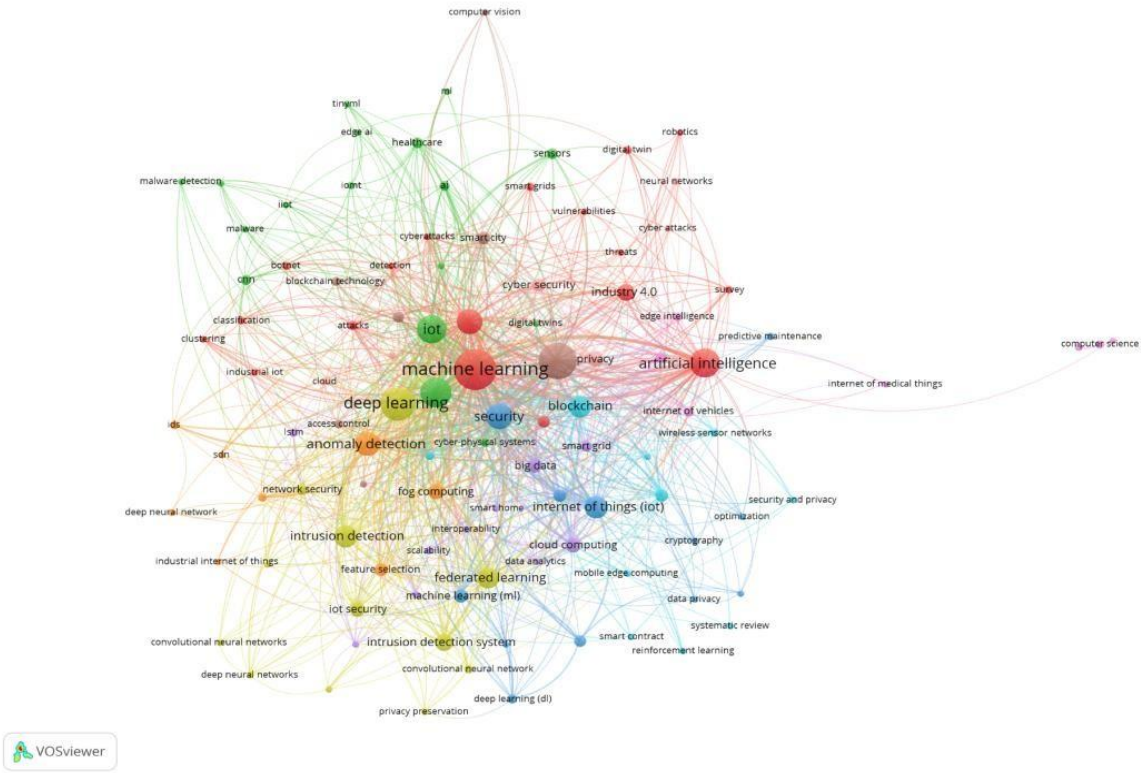

Fig 2. Keyword co-occurrence network generated, highlighting the main thematic clusters in the analyzed literature
Source: VOSviewer (2025).

3.2. Author Collaboration and Co-authorship Networks

The analysis of co-authorship networks allows us to understand the patterns of collaboration between researchers and institutions in the analyzed domain. Figure 3 presents the co-authorship network, highlighting the existence of several relatively isolated research groups with limited interconnection between them.

Despite the presence of some more central authors and institutions, the network is characterized by significant fragmentation, suggesting that the field is still in a phase of scientific consolidation. This fragmentation could be explained by the interdisciplinary nature of the topic, which crosses areas such as cybersecurity, networks, and Artificial Intelligence, as well as by the absence of a widely accepted formal conceptualization of the concept of super nodes.

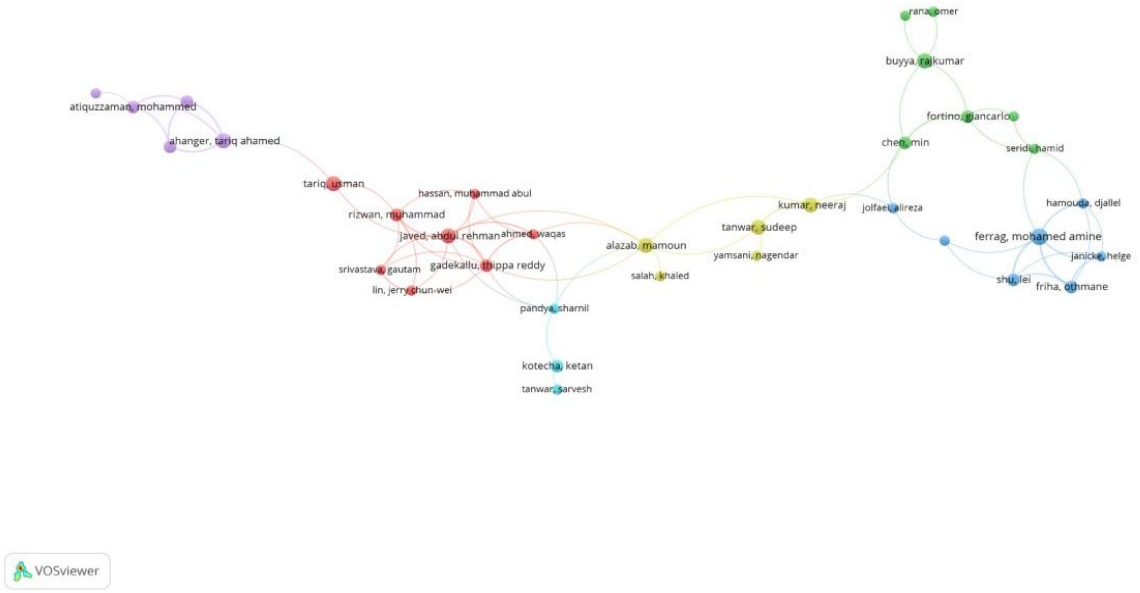

Fig 3. Co-authorship network illustrating collaboration clusters among authors publishing on AI-based threat detection in distributed environments
Source: VOSviewer (2025).

3.3. Publication Venues and Document Types

Analysis of publication sources reveals that scientific output is distributed across a diverse set of journals and conferences, primarily in the fields of computer science, engineering, and telecommunications. Figure 4 shows the distribution of publications among the main scientific journals, highlighting the predominance of specialized technical sources.

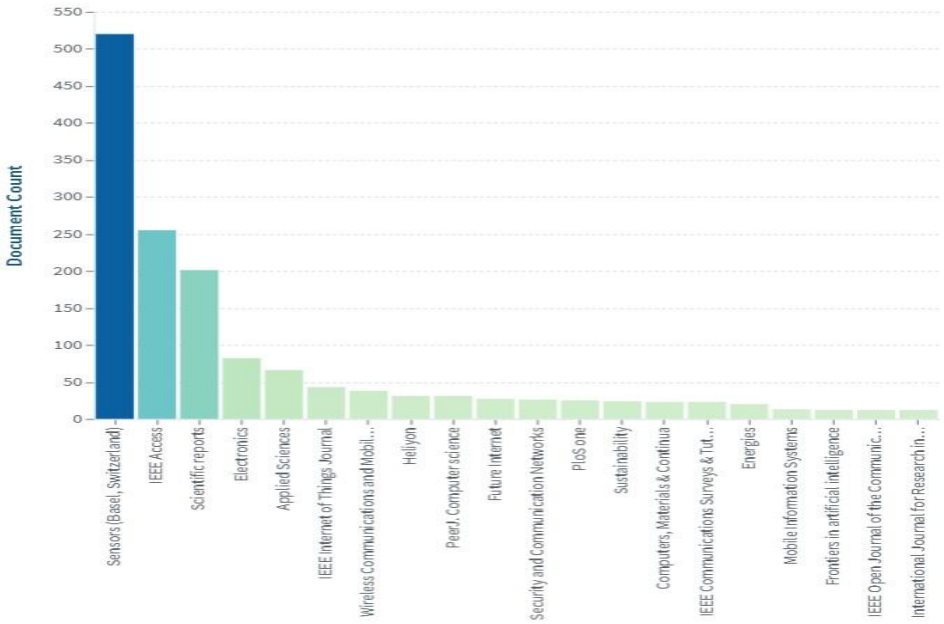

Fig 4. Distribution of publications by major journals
Source: The Lens (2025).

Additionally, Figure 5 illustrates the distribution of document types included in the sample, highlighting the significant weight of conference papers and journal articles, reflecting the dynamic and emerging nature of the field under study.

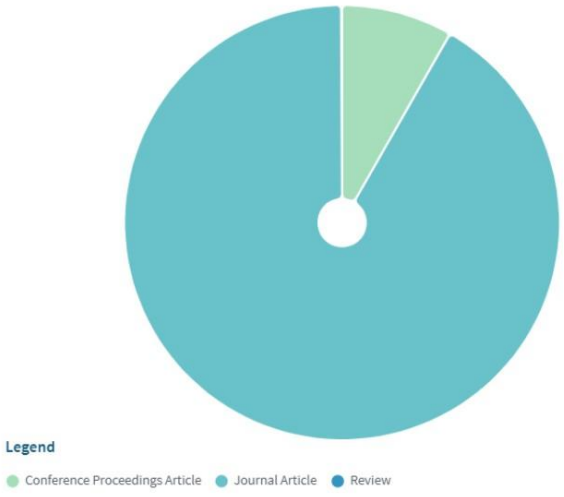

Fig 5. Distribution of document types (journal articles, conference papers, and reviews)
Source: The Lens (2025).

3.4. Geographical Distribution of Research

Analysis of the geographical distribution of publications indicates that research in this field is led by a relatively small number of countries. As illustrated in Figure 6, scientific output is concentrated mainly in China, the United States, India, and several European countries.

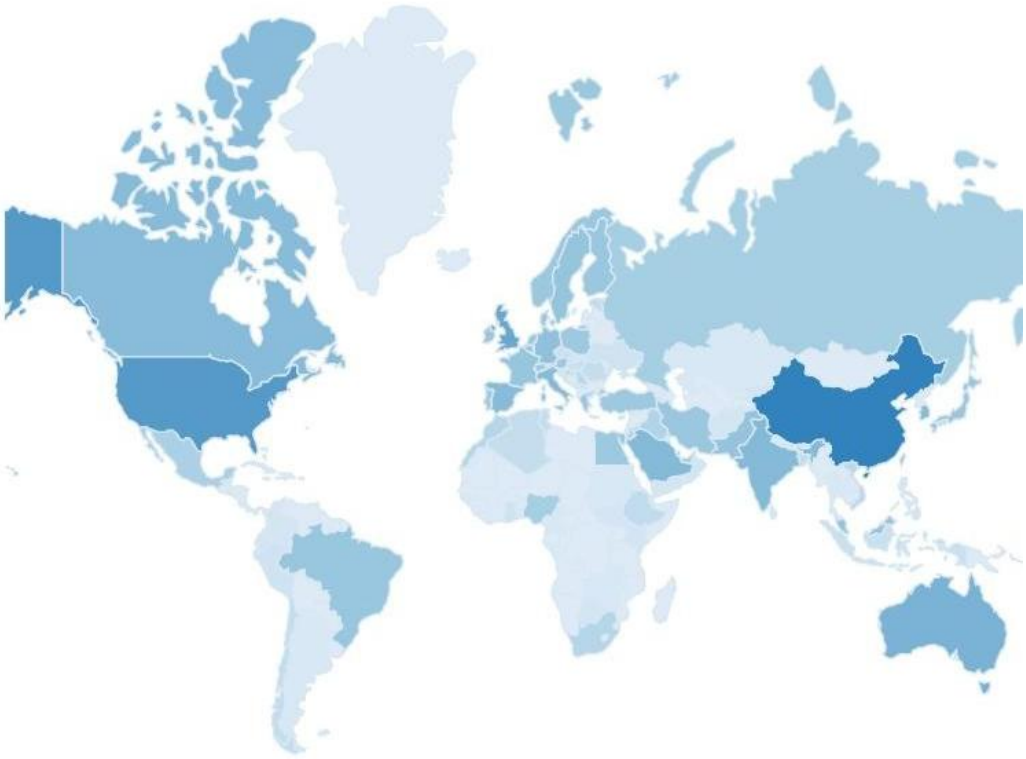

Fig 6. Global geographical distribution of publications by country
Source: The Lens (2025).

This concentration reflects these countries' significant investment in emerging technologies, IoT infrastructure, and artificial intelligence research. However, the analysis also highlights limited international collaboration, reinforcing the perception of fragmentation in the field and pointing to future opportunities for transnational collaborative research.

3.5. Temporal Evolution and Citation Impact

The temporal evolution of scientific production, associated with the impact of publications, is represented in Figure 7. This figure combines the temporal dimension with impact metrics,

allowing us to observe not only the growth in the number of publications over time but also their scientific relevance measured through citations.

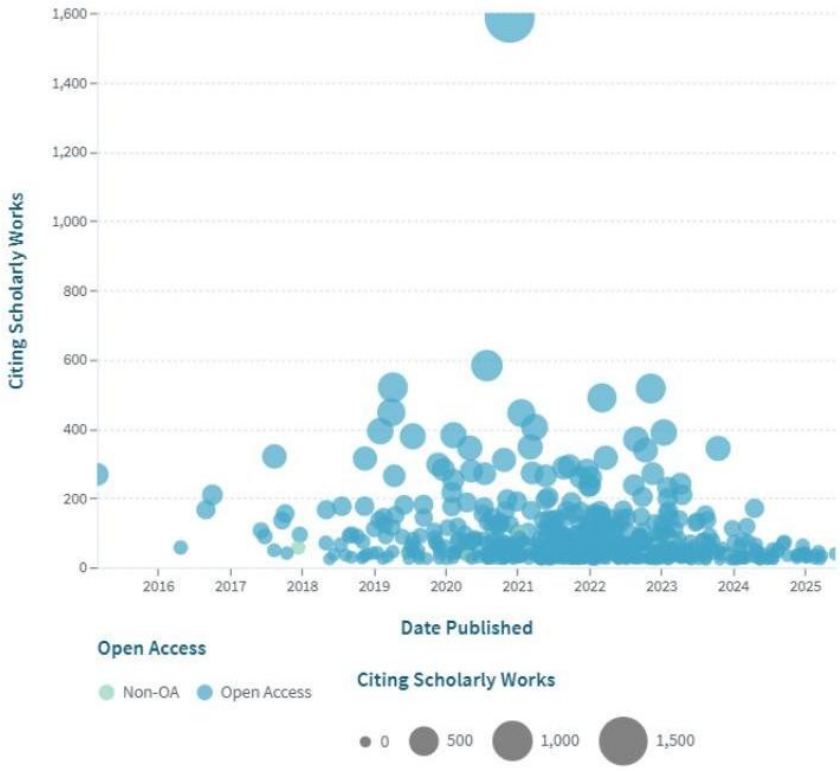

Fig 7. Temporal evolution of publications and citation impact over time
Source: The Lens (2025).

The results show a significant growth in research since 2018, coinciding with the maturation of IoT and edge computing technologies. The increased scientific impact associated with the most recent publications suggests a progressive consolidation of the domain and a growing interest in intelligent distributed architectures, in which super nodes play a central role.

3.6. Scientific Domains Involved

Finally, the analysis of the scientific domains of the included publications reveals a strong concentration in the areas of computer science, engineering, and information technologies. Figure 8 presents a visualization of the main scientific domains associated with the analyzed literature.

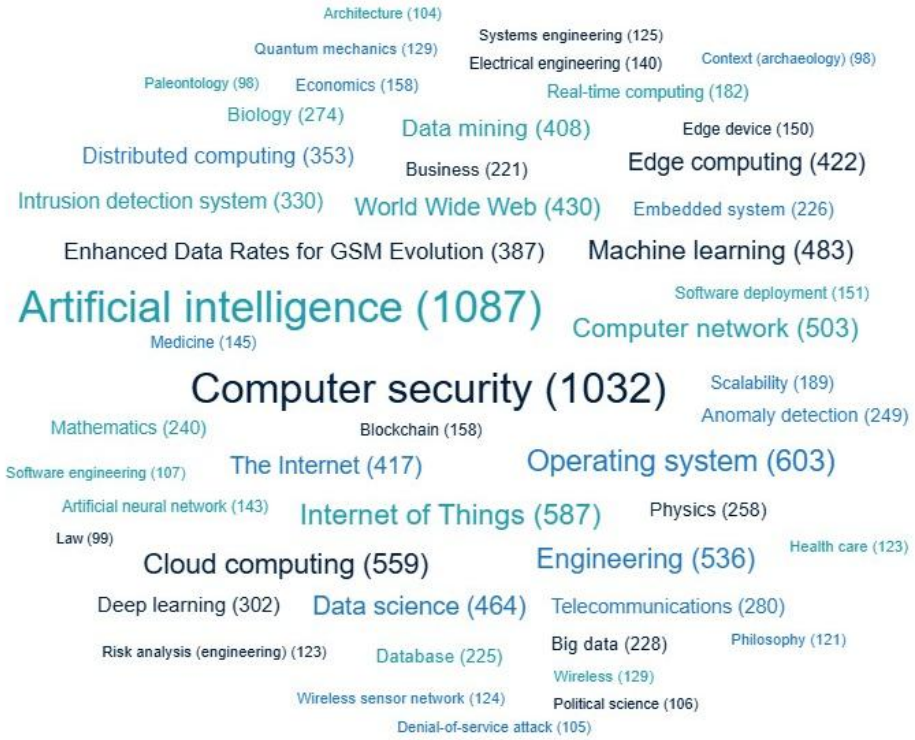

Fig 8. Scientific domains associated with the analyzed publications represented as a word cloud or domain-frequency visualization
Source: The Lens (2025).

This technical focus indicates that research on super nodes and threat detection is still in a phase of technological consolidation, and there is room for future, more integrated approaches that consider organizational, regulatory, and social dimensions.

4. Discussion

The bibliometric analysis carried out allows us to draw a set of relevant conclusions about the scientific evolution of the field of threat detection based on Artificial Intelligence in distributed environments, as well as about the emerging role of super nodes in this context. The results show a rapidly growing but still fragmented research field, both at the conceptual level and at the level of scientific collaboration, confirming several of the gaps identified in the literature.

First, the temporal evolution and impact of publications indicate that scientific interest in this field has intensified significantly since 2018. This growth coincides with the maturation of Internet of Things and edge computing technologies, as well as advances in deep learning techniques applied to cybersecurity. This technological convergence has created the

necessary conditions for the development of more distributed and intelligent architectures, capable of responding to the scalability and latency limitations of traditional centralized models. The increasing relevance of the most recent publications, reflected in their scientific impact, suggests that the topic is in a phase of consolidation and recognition within the academic community.

Keyword co-occurrence analysis revealed a relatively well-defined thematic structure, organized into clusters that reflect the main research lines of the domain. On the one hand, there is a technical core strongly focused on machine learning and deep learning techniques for intrusion and anomaly detection. On the other hand, a set of themes associated with distributed environments emerges, such as IoT, edge computing, and hybrid architectures. The articulation between these clusters suggests that research has been evolving towards the integration of advanced analytical capabilities and distributed architectures, creating a clear conceptual space for the emergence of super nodes as intermediate elements of intelligence.

However, despite this thematic convergence, the literature reveals an absence of a formal and widely accepted conceptualization of the concept of super nodes. Many studies describe architectures that incorporate nodes with enhanced computational capabilities and local aggregation and analysis functions, but they do so use diverse and sometimes inconsistent terminology. This conceptual fragmentation hinders systematic comparison between approaches, as well as the construction of reference models that can guide future investigations and practical implementations.

The identified patterns of scientific collaboration reinforce this perception of fragmentation. The analysis of co-authorship networks reveals the existence of multiple relatively isolated research groups, with reduced interconnection between them. This configuration can be partially explained by the interdisciplinary nature of the domain, which crosses areas such as cybersecurity, networks, artificial intelligence, and systems engineering. However, the absence of strongly interconnected scientific communities may also limit the consolidation of common approaches and the evolution of the field towards more mature and standardized solutions.

The geographical distribution of scientific production also reveals a significant concentration of research in countries with strong investment in emerging technologies, such as China, the United States, and several European countries. Although this concentration is expected, considering the resources and infrastructure available in these contexts, the analysis also shows limited international collaboration. This aspect represents both a current limitation

and a future opportunity, namely for the development of collaborative projects that allow the validation of super node-based architectures in different geographical and operational contexts.

Regarding the predominant scientific domains, the results indicate that research is strongly anchored in technical areas, such as computer science and engineering. The limited presence of approaches originating from areas such as management, public policy, or regulatory frameworks suggests that the domain is still in a phase of technological consolidation. As super node-based solutions evolve into real-world application contexts, it will be crucial to integrate these dimensions, especially in critical environments where privacy, accountability, and regulatory compliance issues play a central role.

In summary, the results of the bibliometric analysis confirm that super nodes constitute an emerging concept with high potential to address the challenges of threat detection in distributed environments. However, the consolidation of this paradigm requires further effort in conceptual systematization, empirical validation, and scientific collaboration. This research contributes to this effort by structurally mapping the state of the art, identifying relevant gaps, and providing a solid foundation for future research in this field.

5. Conclusion

This article aimed to systematically analyze the scientific production related to the use of AI-based super nodes for real-time threat detection in distributed environments. To this end, a bibliometric analysis of 300 scientific publications was conducted, selected according to the PRISMA 2020 guidelines and extracted from the Lens.org database. This allowed for mapping the evolution of the domain, identifying research trends, thematic structures, and relevant gaps in the literature.

The results show a significant growth in research in this field since 2018, reflecting the convergence between the maturation of Internet of Things technologies, edge computing, and advances in machine learning and deep learning techniques applied to cybersecurity. The thematic analysis revealed well-defined clusters, centered, on the one hand, on intrusion detection techniques based on machine learning and, on the other hand, on distributed and hybrid architectures. This thematic convergence reinforces the growing relevance of

architectures that integrate intelligent intermediate nodes, conceptualizable as super nodes, capable of mitigating limitations associated with centralized security models.

Despite this progressive growth and maturation, bibliometric analysis has also revealed a set of structural limitations in the field. In particular, the literature appears fragmented, both at the conceptual level and at the level of scientific collaboration, with an absence of a formal and widely accepted definition of the concept of super nodes. Many studies address functionally similar architectures but resort to diverse and sometimes inconsistent terminology, hindering the systematic comparison of approaches and the consolidation of scientific knowledge.

Additionally, the patterns of scientific collaboration and the geographical distribution of publications suggest that research is concentrated in certain countries and research groups, with relatively small levels of international collaboration. This fragmentation could limit the empirical validation of proposed solutions in diverse operational contexts, as well as delay the emergence of widely recognized reference models or best practices.

In this context, super nodes emerge as a promising paradigm for threat detection in distributed environments, offering an intermediate approach between resource-limited edge devices and centralized cloud platforms. Their ability to aggregate data, perform local AI-based inference, and make decisions in reduced time makes them particularly suitable for scenarios that demand rapid responses, scalability, and resilience.

Future research areas include the need to develop a formal and standardized conceptualization of the super node concept, as well as conducting empirical studies to evaluate their performance in real or simulated environments. Future investigations could also explore the integration of super nodes with distributed orchestration mechanisms, federated learning techniques, and adaptive security approaches. Finally, it will be relevant to incorporate organizational, regulatory, and ethical perspectives to ensure that the adoption of these architectures occurs in a secure, transparent manner, aligned with legal and social requirements.

In summary, this work contributes to the systematization of the state of the art on super nodes and threat detection based on Artificial Intelligence, providing a solid foundation for advancing research in this emerging and technologically relevant field.

References

Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., Ali, I., & Guizani, M. (2020). A survey of machine and deep learning methods for Internet of Things (IoT) security. *IEEE Communications Surveys & Tutorials, 22*(3), 1646–1685. https://doi.org/10.1109/COMST.2020.2988293

Conti, M., Dehghantanha, A., Franke, K., & Watson, S. (2018). Internet of Things security and forensics: Challenges and opportunities. *Future Generation Computer Systems*, *78*, 544-546.

Goranin, N., Hora, S. K., & Čenys, H. A. (2024). A bibliometric review of intrusion detection research in IoT: Trends, collaboration, and thematic evolution. *Electronics, 13*(16), 3210. https://doi.org/10.3390/electronics13163210

Li, C., Xue, Y., Wang, J., Zhang, W., & Li, T. (2018). Edge-oriented computing paradigms: A survey on architecture design and system management. *ACM Computing Surveys (CSUR)*, *51*(2), 1-34. https://doi.org/10.1145/3154815

Nguyen, T. T., & Reddi, V. J. (2023). Deep reinforcement learning for cyber security. *IEEE Transactions on Neural Networks and Learning Systems*, *34*(8), 3779–3795. https://doi.org/10.1109/TNNLS.2021.3121870

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Roman, R., Lopez, J., & Mambo, M. (2018). Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges. *Future Generation Computer Systems*, *78*, 680-698.

Shafiq, M., Tian, Z., Liu, Y., Aljuhani, A., & Li, Y. (2024). ESC&RAO: Enabling seamless connectivity resource allocation in tactile IoT for consumer electronics. *IEEE Transactions on Consumer Electronics*, *70*(3), 5506–5515. https://doi.org/10.1109/tce.2023.3327136

Singh, P., Kaur, A., & Singh, D. (2021). Edge-centric network intrusion detection using deep learning. *arXiv Preprint*, arXiv:2102.01873. https://arxiv.org/abs/2102.01873

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2019). Edge intelligence: Paving the last mile of artificial intelligence with edge computing. *Proceedings of the IEEE, 107*(8), 1738–1762. https://doi.org/10.1109/JPROC.2019.2918951