Analysis of Thematic Critical Mass in the UNIMORE Knowledge Graph

Sofia Locicero, 282178@studenti.unimore.it

***Abstract*—This paper presents an analysis of the thematic critical mass within the publication network of the University of Modena and Reggio Emilia (UNIMORE). The study aims to evaluate the university's research activity across different areas by measuring the number of publications, involved researchers, interdisciplinarity, collaboration patterns. Using the UNIMORE Knowledge Graph, publications were associated with topics, keywords, SSDs, categories and macro areas (Area CUN). The findings provide insights into the university's strong research domains and suggest potential directions for internal and external collaborations. This analysis lays the groundwork for future research on interdisciplinarity, collaboration networks, and scientific impact, offering deeper insight into UNIMORE's thematic strengths.**

I. INTRODUCTION

The analysis of research output and thematic clusters is essential to understand the strengths and strategic positioning of a university. Bibliometric studies provide valuable insights into active research areas, interdisciplinarity, and collaboration patterns at both national and international levels. The University of Modena and Reggio Emilia (UNIMORE) maintain a rich repository of publications, authors, and research topics, structured in the UNIMORE Knowledge Graph, which enables a comprehensive analysis of the university's scientific activity.

The main objective of this study is to assess the thematic critical mass of UNIMORE by analyzing publication volume, researcher involvement, interdisciplinarity, collaboration networks, and preliminary measures of scientific impact. Understanding these aspects is crucial to identify the university's strong research domains and to support future collaborations and strategic planning.

The remainder of the paper is organized as follows: Section II describes the methodology, Section III presents the results, Section IV discusses the findings, and Section V concludes with potential extensions and future work.

II. METHODOLOGY

The dataset used in this study is derived from the UNIMORE Knowledge Graph, which contains all publications of the University of Modena and Reggio Emilia (UNIMORE). The graph provides a structured representation of the publication network, including information about documents, authors, research topics, keywords, SSDs (Settori Scientifico-Disciplinari), and macro areas (Area CUN).

Graph Nodes and Relationships

The UNIMORE Knowledge Graph contains a rich set of entities (nodes) and relationships that represent the publication network. The main nodes include:

- Document: publications
- Author: researchers contributing to publications
- Topic: topics associated with documents
- Source: journals or conferences where documents are published
- KeyWord: keyword extracted from publications
- SSD: scientific sector
- Area: macro areas (Area CUN) of SSDs
- Category
- LocalCollaboration
- InternationalCollaboration

The main relationships among these nodes are:

- (:Author)-[:WRITE]->(:Document)
- (:Document)-[:HAS]->(:Topic)
- (:Document)-[:HAS]->(:KeyWord)
- (:Author)-[:ASSOCIATED_WITH]->(:SSD)-[:PART_OF]->(:Area)
- (:Document)-[:PUBLISHED_IN]->(:Source)-[:BELONGS_TO_CATEGORY]->(:Category)
- (:Document)-[:IS]->(:Type)
- (:Author)-[:WORKS_IN]->(:Department)
- (:Document)-[:HAS_COLLABORATION]->(:LocalCollaboration)
- (:Document)-[:HAS_INTERNATIONAL_COLLABORATION]->(:InternationalCollaboration)
- (:Document)-[:PUBLISHED_BY]->(:Publisher)

Metrics Computed

The study focuses on evaluating the thematic critical mass through different quantitative measures. The number of publications was computed per topic, keyword, category, and Area. The number of authors involved was calculated for each thematic category. Interdisciplinarity was measured by the presence of co-authors from different SSDs or Areas. Internal and external collaborations were identified by analyzing the collaboration relationship linked to each document and then aggregated by category. Scientific impact was evaluated based on citations and authors' H-index derived from their publications.

Tools and Procedures

Neo4j was used to query the Knowledge Graph and extract the necessary metrics. Cypher queries were developed to count publications, authors and collaborations, as well as to compute each authors' H-index and the evolution of research activity. The extracted data were then aggregated at multiple levels to evaluate thematic strength and publication activity.

III. RESULTS

The UNIMORE Knowledge Graph was visualized using apoc.meta.graph to provide an overview of all nodes and relationships. As shown in Figure 1, the graph includes multiple types of nodes, such as documents, authors, topics, keywords, SSDs, areas, sources, categories, departments and collaboration entities. Relationships such as authorship, document–topic association, publication in sources, and internal or external collaborations are represented, offering a comprehensive view of the network. This overview highlights the structural complexity of UNIMORE's academic output and facilitates the analysis of thematic connections and collaboration patterns.

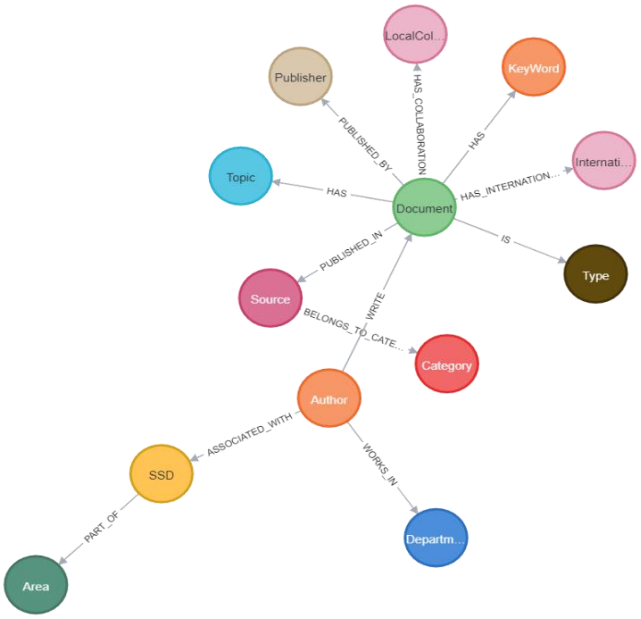

Figure 1: Overview of the UNIMORE Knowledge Graph.

Publications by Topic

Table I reports the most frequent topics in the dataset.

TABLE I.

Topic	Number of Publications
COVID-19	200
Deterioration	107
Animals	96
Atrial Fibrillation	89
Human Immunodeficiency Virus 1	76
Internet of Things	72

Top topics were identified by counting the number of publications associated with each topic. The dataset shows that medical topics dominate and topics such as COVID-19, Deterioration, and Atrial Fibrillation demonstrate significant research activity. Emerging technology topics like Internet of Things also show notable research presence.

Publications by Keyword

Table II presents the most frequent keywords associated with publications in the dataset.

TABLE II.

Keyword	Number of Publications
COVID-19	192
Atrial Fibrillation	94
HIV	89
Hadron-Hadron Scattering (experiments)	87
B physics	62

The keyword analysis confirms the medical domain's predominance, but also reveals activity in particle physics, highlighting the variety of disciplines represented in UNIMORE's research output. Overlap between topic and keyword frequency, such as COVID-19, further emphasizes key research priorities.

Publications by Research Area

Table III presents the number of publications for the main macro-areas (CUN) in the dataset.

TABLE III.

Macro areas (CUN)	Number of Publications
Medical Science	7716
Industrial and Information Engineering	4758
Biological Science	1729
Chemical Science	1708
Physical Science	919

Publications by research area (macro-areas CUN) highlight the largest research pillars, with Medical Sciences and Industrial and Information Engineering contributing the highest number of publications. Biological, Chemical, and Physical Sciences also contribute substantially, supporting an interdisciplinary profile.

Publications by Research Category

Table IV presents the number of publications in the top research categories.

TABLE IV.

Category	Number of Publications
Electrical and Electronic Engineering	735
Computer Science Applications	662
Mechanical Engineering	572
General Materials Science	554
General Medicine	493

Publications by research category highlight the most productive research areas within UNIMORE. Electrical and Electronic Engineering, Computer Science Applications, and Mechanical Engineering show the highest number of publications, followed by General Materials Science, General Medicine, Surgery, and Oncology.

Authors involved per Theme

The number of authors associated with each topic, keyword, and research area was computed. COVID-19 emerges as the most studied topic in terms of author involvement, with 5,211 distinct authors, followed by Atrial Fibrillation and Anticoagulant Agents. Among keywords, COVID-19, Mortality, and Atrial Fibrillation attract the highest number of authors. By research category, General Medicine, Oncology,

Multidisciplinary, Cardiology and Cardiovascular Medicine, and Infectious Diseases involve the largest author groups. Figure 2 shows that medical categories concentrate the highest number of authors, reflecting substantial human resource engagement in medical research.

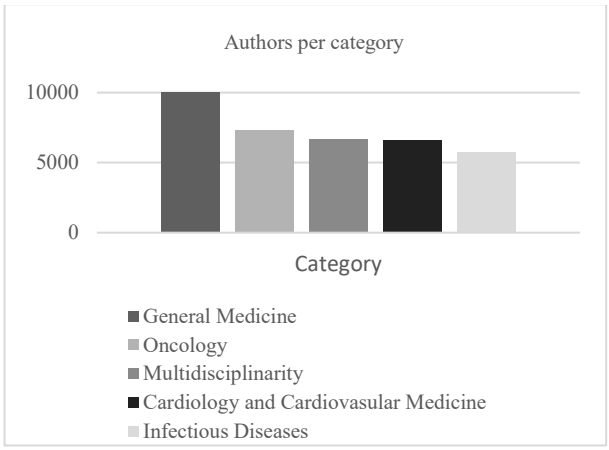

Figure 2: Number of authors involved in each research category.

Interdisciplinarity

Interdisciplinarity was assessed by measuring the diversity of SSDs and CUN areas among co-authors for each publication. For keywords with at least 10 publications, the average number of distinct SSDs and areas per document ranges from approximately 2.3 to 2.6, indicating a moderate degree of interdisciplinarity. Selected topics, such as Kaons, Quark Models, and Deterioration, show the highest average number of areas per document, up to 3.6, reflecting stronger cross-disciplinary collaboration. Figure 3 illustrates this trend for selected topics, while the overall analysis is primarily based on research categories. Some topics, particularly in physics, show stronger cross-disciplinary collaboration, suggesting that certain fields rely more heavily on interdisciplinary work. Moreover, it should be noted that several journals are indexed under multiple research categories. This may result in overlaps when counting publications per category, but at the same time it highlights the inherently interdisciplinary nature of many research areas.

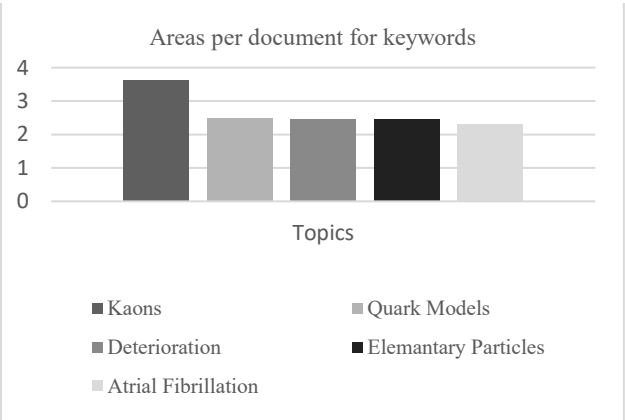

Figure 3: Average number of distinct CUN areas per document for selected topics, illustrating interdisciplinarity in research.

Collaborations

Local and international collaborations were evaluated based on publications involving multiple institutions. By keywords, COVID-19 exhibits the highest number of local collaborations (192 publications), while Hadron-Hadron Scattering ranks first for international collaborations (87 publications). By category, Electrical and Electronic Engineering, General Materials Science, and General Physics and Astronomy display the highest numbers of both local and international collaborations, suggesting that these areas are particularly active in collaborative research. Figure 4 compares local and international collaborations across research categories. Electrical and Electronic Engineering exhibits the highest number of both local (724) and international (305) collaborations, followed by Computer Science Applications and General Materials Science. Most categories show more local than international collaborations, with General Medicine presenting 484 local and 143 international collaborations. Technical and engineering fields generally demonstrate stronger collaborative activity than medical topics in terms of international reach.

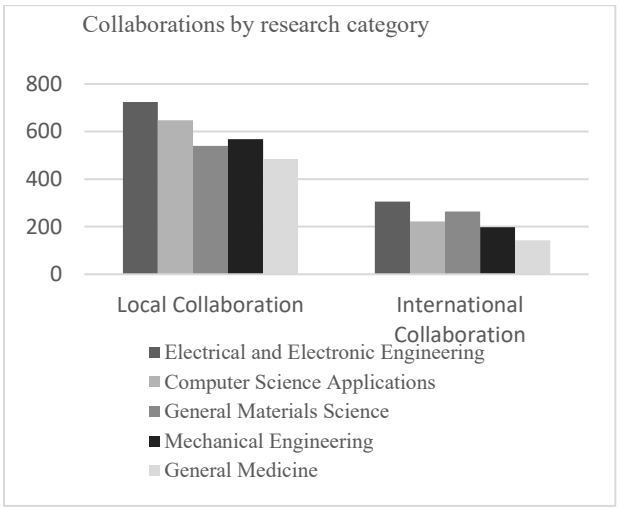

Figure 4: Number of local and international collaborations per subject category

Figure 5 shows a comparison of total publications, local and international collaborations, and author involvement per research category. General Medicine, while having the largest number of authors involved, shows a higher prevalence of local collaborations relative to international ones.

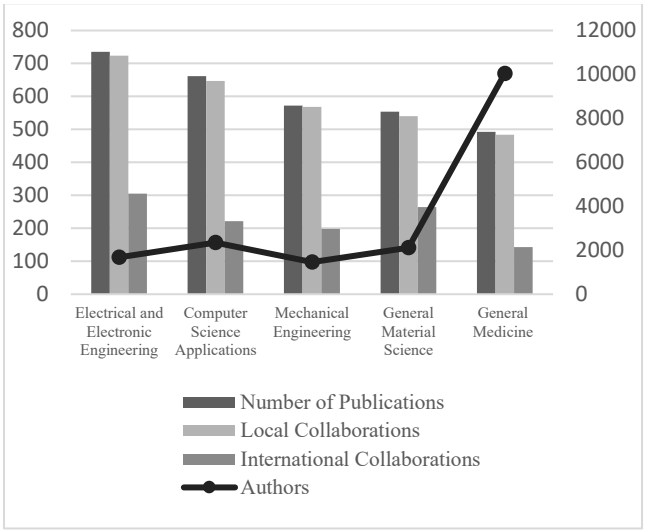

Figure 5: Comparison of total publications, local and international collaborations, and number of authors per research category

The percentage of publications with international collaborations was computed for each research category. Certain small categories, such as Complementary and Manual Therapy, Research and Theory, and Psychiatric Mental

Health, reach 100% international collaboration, although their publication volume is very low. Larger categories in the physical sciences, particularly Nuclear and High Energy Physics (212 out of 229 publications, 92.6%), maintain high absolute numbers of international collaborations. Categories such as Nature and Landscape Conservation (10 out of 12, 83.3%) and Transportation (23 out of 29, 79.3%) also show high international engagement despite modest publication counts.

Publications Impact

The scientific impact of publications was assessed using citation counts and authors' H-index. Citation analysis across keywords, topics, and categories shows that medical research remains highly cited, with "COVID-19," "Atrial Fibrillation," and related topics leading in total citations. At the category level, Electrical and Electronic Engineering, Computer Science Applications, General Materials Science, Mechanical Engineering, and General Medicine were selected for detailed comparisons, as they represent both core research areas and highly cited categories.

Figure 6 shows the distribution of citations per document for these main categories. Outliers are represented as individual points, corresponding to documents with exceptionally high citation counts. Other categories, such as Cell Biology and Cardiology and Cardiovascular Medicine, include documents with extremely high citations (up to 4,082 and 3,013, respectively) but are excluded from the figure for clarity. The category "General Biochemistry, Genetics and Molecular Biology" has a total of 8,314 citations across 254 documents, with an average of 32.7 citations per document.

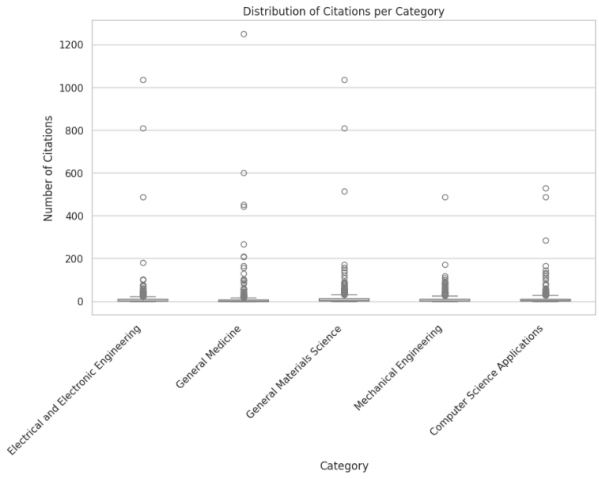

Figure 6: Distribution of citations per document for the main research categories considered

The H-index was computed for individual authors affiliated with UNIMORE departments and for co-authors with at least two shared publications. Top authors by H-index include Bizzeti A. (38), Cardini A. (37), and Vinceti M. (35), indicating a highly productive group contributing disproportionately to university's research visibility. The average H-index of strong co-authors ranges from approximately 15 to 20, indicating that high-impact authors tend to collaborate with other highly cited researchers. Additionally, authors with higher H-index also demonstrate many international collaborations. Average, maximum, and minimum H-index values were also evaluated at the Area CUN level, with Medical Sciences showing the highest average H-index, followed by Chemical Sciences.

These metrics provide a comprehensive view of the impact and visibility of research output across different fields, illustrating both domain-specific and interdisciplinary contributions.

Publications Trend

To further investigate the evolution of research activity, the number of publications per category was analyzed over time (2017–2021). General Medicine shows a substantial increase, from 51 publications in 2017 to 179 in 2021. Materials Science more than doubled, from 62 to 155 publications in the same period. Engineering-related fields, including Electrical and Electronic Engineering, Computer Science Applications, and Mechanical Engineering, maintain consistently high output over the years. Figure 7 illustrates these trends for selected research categories.

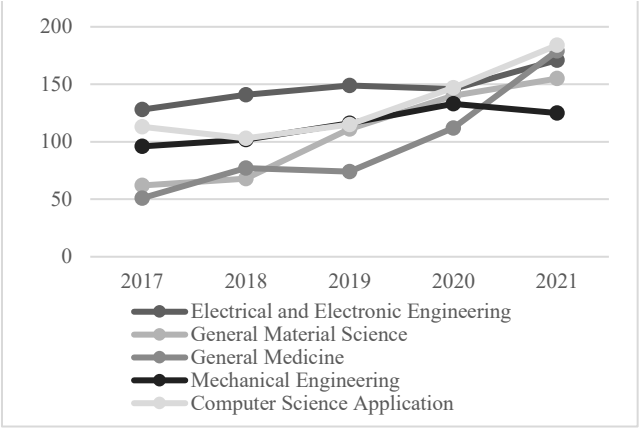

Figure 7: Evolution trends of publications for each research category

IV. DISCUSSION

Thematic Strengths and Author Engagement

The distribution of publications and authors highlights that Medical and Physical Sciences and Engineering-related fields represent the university's strongest domains, both in volume of research and in human resource involvement. High engagement in topics such as COVID-19 and Atrial Fibrillation confirms the university's focus on pressing and high-priority areas, guiding potential future collaboration.

Interdisciplinarity and Collaboration Patterns

Moderate interdisciplinarity was observed for most topics, with higher diversity in areas such as Kaons and Quark Models, pointing to potential opportunities for strengthening cross-disciplinary collaboration in other research areas. Local and international collaborations are concentrated in medical or technical fields (e.g., Electrical and Electronic Engineering, General Materials Science, and Computer Science Applications), reflecting both high research output and robust collaboration networks. These findings suggest that UNIMORE's research influence is driven not only by publication volume but also by effective collaboration, highlighting areas with strong collaborative potential.

The combined comparison of total publications, collaborations, and authors (Figure 5) highlights differences

in how research categories mobilize resources and engage in collaborative activity. Electrical and Electronic Engineering and Computer Science Applications combine high publication output with substantial number of both local and international collaborations. In contrast, General Medicine shows a particularly high number of contributing authors, mostly involved in local collaborations. General Materials Science and Mechanical Engineering display a balanced pattern of publications and collaborations. These patterns suggest that engineering and technical fields rely on international collaborative networks, whereas medical research engages a larger author base but with a stronger local focus. These insights can guide UNIMORE in prioritizing support, fostering international partnerships and enhancing collaboration strategies in areas with high potential for collaborative impact.

Comparing both the volume and the percentage of international collaborations allows identification of categories that are both highly collaborative in absolute terms and those with particularly strong international propensity. Physical sciences, such as Nuclear and High Energy Physics and Physics and Astronomy, combine high publication output with substantial international collaboration, whereas smaller categories, while producing fewer publications, can still achieve relatively high international engagement. These insights provide a nuanced understanding of collaborative dynamics across research areas and may guide UNIMORE in prioritizing support for international partnerships.

Scientific Impact

The distribution of citations (Figure 6) confirms that the categories with the highest number of publications and authors are not necessarily those with the highest single-document citation counts. Electrical and Electronic Engineering, Computer Science Applications, General Materials Science, Mechanical Engineering, and General Medicine show the highest volume of research activity and engagement. However, categories such as Cell Biology and Cardiology and Cardiovascular Medicine contain individual documents with exceptionally high citations, which significantly affect total citation metrics. General Biochemistry, Genetics and Molecular Biology also shows high average citations per document, indicating that scientific impact is influenced both by volume and by standout publications. These results emphasize that while the five selected categories are strategic pillars in terms of collaboration and author involvement, high-impact contributions can also emerge from smaller or more specialized areas.

Authors with high H-index values tend to collaborate with similar influential co-authors, underscoring the role of strong research networks. Additionally, a positive correlation emerges between authors' H-index and their level of international collaboration, suggesting that global engagement contributes to research impact. The H-index also allows identifying key contributors who disproportionately influence UNIMORE's visibility, providing insights into mentorship, collaboration dynamics, and potential strategic partnerships.

Annual Trends

The analysis of publication trends over time highlights a clear growth in medical and materials science research at UNIMORE, particularly during 2020–2021. General Medicine publications increased from 51 in 2017 to 179 in 2021, while Materials Science more than doubled from 62 to 155. Engineering fields maintain consistently high output. The surge in medical publications is strongly influenced by COVID-19, which emerged as a dominant topic and keyword, reflecting the university's responsiveness to urgent research needs. These temporal patterns demonstrate that UNIMORE's thematic critical mass is not static but dynamically evolving in response to emerging scientific challenges.

Strategic Implications

These results help identify UNIMORE's strong thematic areas and fields with the greatest potential for future partnerships. By leveraging strengths in Medical, Physical and Engineering fields, the university can continue to participate in high-profile collaborations while promoting interdisciplinary initiatives in emerging topics, such as Internet of Things, Machine Learning and AI, which currently show moderate engagement. Furthermore, the observed link between high-impact researchers, collaboration intensity, and global engagement suggests that fostering international networks and supporting key contributors could significantly enhance UNIMORE's overall research visibility and impact.

V. CONCLUSION AND FUTURE WORK

This study analyzes the UNIMORE Knowledge Graph to provide a systematic assessment of research activity, author engagement, collaboration patterns, and scientific impact. The analysis shows how structured graph-based metrics can support strategic decision-making by identifying areas of strong research performance, high-impact authors, and effective collaboration networks, offering actionable insights for institutional planning, including resource allocation, fostering collaborations, and supporting interdisciplinary initiatives.

Unlike traditional databases, the Knowledge Graph enables a richer and more flexible representation of research activity. By explicitly modeling entities such as authors, topics, keywords, SSDs, areas, and collaborations, it allows multi-level analyses that capture structural relationships across disciplines and institutions. This graph-based approach facilitates the identification of thematic clusters, interdisciplinarity, and collaboration patterns that would be difficult to detect with conventional datasets, providing both quantitative and qualitative insights into UNIMORE's strategic strengths and emerging fields.

Future work will focus on continuously updating the Knowledge Graph to capture new publications and collaborations, enabling dynamic monitoring of research trends. Additional analyses may explore co-author network structures, inter-SSD interactions, and the temporal evolution of thematic areas. Predictive tools could also be developed to identify emerging research topics and guide strategic planning for research partnerships and investments.

REFERENCES

[1] Neo4j. "Neo4j Documentation". In: (2025).

[2] J. D. Hunter, "Matplotlib: A 2D Graphics Environment", Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, 2007.
[3] Waskom, M. L., (2021). seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021, https://doi.org/10.21105/joss.03021.
[4] McKinney, W., et al. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51–56).