sciendo

Analyzing Regional Trends in Erasmus+ Project Funding and Thematic Clusters through Heatmaps Visualizations

Ioana-Diana PETRE*
Bucharest University of Economic Studies, Bucharest, Romania
**Corresponding author, diana.petre@csie.ase.ro*

Mihai GHEORGHE
Bucharest University of Economic Studies, Bucharest, Romania
mihai.gheorghe@csie.ase.ro

Mihail BUȘU
Bucharest University of Economic Studies, Bucharest, Romania
mihail.busu@fabiz.ase.ro

Abstract. *In the past decades, the Erasmus+ program has been a foundation of the European educational exchanges, promoting academic and professional mobility across the EU. This paper analyzes the regional trends in Erasmus+ project funding across Europe, using heatmap visualizations to represent project density and grant distribution. The study highlights key regions with high project participation and significant grant awards, suggesting a strong concentration in Central and Western Europe, particularly in France, Germany, Benelux, UK, Spain. The integration of thematic clusters provides insights into the types of projects funded and their regional prevalence and suggests that Erasmus-funded projects are more frequent in wealthier, central European countries where there is a stronger institutional network for research and education. These findings emphasize the need for more inclusive funding strategies to support project development in underrepresented regions.*

Keywords: Erasmus+, Mobility, Regional Trends, Thematic Clusters, Heatmap Visualization.

Introduction

The Erasmus+ program, launched in 1987 (Herz et al., 2018), has facilitated the mobility of millions of students, educators, and researchers across Europe. With substantial funding from the EU, the program promotes educational cooperation, intercultural understanding, and academic excellence.

This study builds on the Innovation Knowledge Graph project, analyzing Erasmus+ project data to identify regional trends, funding patterns, and thematic clusters, aiming to provide insights into the program's spatial and thematic dynamics (Ryazanova & McNamara, 2019). Previous studies have highlighted the importance of mobility programs in fostering educational cooperation and cultural exchange.

One key challenge in evaluating the overall impact of Erasmus and Horizon projects is the clear difference between Eastern/Southern countries and their Western/Northern counterparts. Participation rates for Erasmus tend to be higher in wealthier regions. Meanwhile, Horizon 2020 and Horizon Europe funding often concentrates in countries with established research infrastructures—such as Germany, France, the Netherlands, and Scandinavia (Nascimbeni et al., 2021)—leaving others striving to catch up. Financial constraints, linguistic barriers, and institutional capacity can all inhibit greater participation in these funding streams. Post-socialist states often face more severe challenges related to administrative resources or difficulty meeting co-financing requirements, thus

DOI: 10.2478/picbe-2025-0019
© 2025 I.-D. Petre; M. Gheorghe; M. Buşu, published by Sciendo.
This work is licensed under the Creative Commons Attribution 4.0 License.

limiting the number of research proposals they can successfully submit. Nevertheless, targeted measures within Horizon Europe, such as Twinning, Teaming, and ERA Chairs, aim to reduce these gaps by channeling resources toward less-performing research regions.

Through clustering algorithms and natural language processing (NLP), this paper identifies and categorizes Erasmus+ projects based on their thematic focus (Concilio et al., 2022). By linking thematic clusters with funding levels and participation rates, the analysis offers insights into the systemic inequalities that persist within European educational and research networks. The results also highlight how EU initiatives can better support less-developed regions by targeting funding and increasing institutional capacity-building efforts.

The paper continues with the Methodology section, in which the authors detail the data collection and processing steps, including data cleaning, normalization, and thematic clustering using NLP techniques and K-Means clustering (Greco, 2023). The Results section presents key findings on Erasmus+ project participation and funding trends, with heatmap visualizations highlighting disparities across European regions and thematic clusters. The Discussion section interprets these findings in the context of EU policies and educational equity goals, examining the underlying factors contributing to regional disparities and suggesting potential measures to improve participation (Dagen & Fink-Hafner, n.d.). Finally, the Conclusions summarize the study's main insights, acknowledge the study's limitations, and propose recommendations for future research to enhance regional equity and participation in Erasmus+ programs.

Literature review

The Erasmus+ programme plays a great role in developing international youth mobility across Europe, (Herz et al., 2018) highlight how historical and socio-economic factors continue to shape mobility patterns, particularly in post-socialist countries like Romania and Hungary, where student participation remains lower compared to Western European nations. The Erasmus+ programme helps balance opportunities by offering structured mobility, though funding and student experiences vary. Western European students participate more, while Eastern European students often use mobility to boost their careers and financial stability. These differences highlight the need to analyze funding distribution and project themes, which can be clearly shown using heatmaps. (Bryła & Domański, 2014) explores the professional and personal trajectories of former Erasmus students, emphasizing how international mobility serves as a tool for social mobility, particularly for middle-class youth from Southern Europe. The study highlights how institutionalized student mobility, such as the Erasmus+ programme, aligns with broader socio-economic trends.

(Dabasi-Halász et al., 2018) examine how international youth mobility within Europe intersects with key life transitions, such as moving from education to employment, from unemployment to work, and from youth to adulthood. They categorize mobile youth into three main groups: students pursuing higher education abroad, graduates seeking employment opportunities (higher-skilled workers), and non-graduates engaging in lower-skilled jobs. The study highlights that youth mobility is often driven by economic disparities, employment opportunities, and the need for skills recognition, with mobility serving as a strategy for career development. However, they also note that mobility patterns are shaped by structural inequalities, with students and workers from peripheral regions often using migration as a means of socio-economic advancement.

These findings underscore the need to analyze Erasmus+ funding distribution and thematic clustering, which can be effectively visualized through heatmap-based methodologies.

Methodology

The methodology for extracting, clustering, and analyzing the Erasmus+ project topics consists in a combination of data cleaning, text normalization, natural language processing (NLP) embedding techniques, and clustering algorithms designed to capture semantic similarities in the textual descriptions of Erasmus+ projects (Bryła & Domański, 2014). The goal is to identify and visualize patterns or trends in how projects align with predefined thematic clusters and understand how different institutions, regions, and partners engage with specific topics.

We have collected data from official Erasmus+ repositories, focusing on a set of core variables such as "project identifier," "project title," "topics," "project summary," and "EU grant award in euros," along with details regarding coordinating and partner organizations. Once these raw data were gathered, preprocessing was done to ensure consistency, validity, and ease of analysis. This involved cleaning duplicate records, fixing inconsistent naming conventions for institutions, and harmonizing country or region codes. Given that textual data can be inconsistent -especially when drawn from multiple Excel files, external data sources, and user inputs-an essential part of preprocessing involves normalizing text. Tokens within the project summaries or topic labels were stripped of punctuation, converted to lowercase, and grouped or united to maintain consistency. For instance, variations like "universities," "University," or "university-based" can be reduced to a base form of "university," this step is called lemmatization (Probierz et al., 2022) and represents the process of grouping together different inflected forms of the same word.

After the data was cleaned, the topic extraction step started. The goal is to construct a list of distinct topics that appear across all project summaries and formal "topics" fields. We scanned every project record and captured all unique terms and thematic designations that had been assigned by project coordinators or generated in the text descriptions. These extracted topics were stored alongside their frequencies in a specialized dictionary. Frequency analysis shows how commonly certain themes appear within the Erasmus+ portfolio, making it easy to understand focus areas (e.g., "digital education," "social inclusion," "language learning," etc.) across the data set.

With the topics identified and pre-processed, we then transformed these terms and short phrases into continuous vector representations using the modern NLP (Taskiran & Kaya, 2022) model called all-MiniLM-L6-v2 together with Sentence Transformers. This model creates embeddings that represent each topic. The benefit of using embeddings is that topics with similar meanings, such as "distance learning" and "online education" will be placed close together in the vector space, while unrelated topics like "waste management" and "arts education" will be far apart. After converting each topic to a vector, the clustering algorithm used was K-Means (Li & Wu, 2012), to group similar embeddings into cohesive clusters. K-Means aims to minimize the within-cluster variance, effectively ensuring that topics assigned to the same cluster are, on average, more like each other than to topics outside the cluster. For this specific analysis, we chose the following three predefined clusters:

1. **Cluster #1**: Sustainable Development, Education, and Innovation in Regional and International Contexts,
2. **Cluster #2**: Innovative and Digital Education Enhancing Key Competencies, Cultural Engagement, and European Citizenship,
3. **Cluster #3**: Promoting Social Inclusion, Equality, and Well-being through Civic Engagement and Support for Disadvantaged Groups.

Cluster #1 covers topics related to environmental sustainability, international partnerships, and educational methodologies that encourage innovation; cluster #2 emphasizes digital

transformation, creative competencies, and intercultural dialogue and cluster #3 covers equity, support for marginalized communities, and societal welfare initiatives.

For a short summary of the last steps, the data from the Erasmus+ project database were processed to extract key variables, including project title, topics, and grant amounts. Heatmaps (Netek et al., 2018) were then generated in ArcGIS Pro by spatially combining project density, average thematic similarity (between projects in each area), and total EU grant values, resulting in visualizations that highlight regions with concentrated, thematically coherent, and well-funded Erasmus+ activity.

For further analysis we have calculated the following metrics to evaluate the topic relationships within the clusters:

1. Centroid similarity,
2. Average similarly – this metric will be used for our heatmaps generation,
3. Maximum similarity.

ArcGIS Pro is used for spatial data processing and visualization (Pokojski et al., 2021), particularly for the generation of density-based heatmaps. The Erasmus+ project dataset was integrated into the ArcGIS Pro environment, where it was transformed into a shapefile format. The ArcGIS Pro Coordinate Identifier API (Van Duynhoven & Dragićević, 2024) was then used to geocode each participating institution, obtaining latitude and longitude coordinates. Separate heatmaps were then generated to represent different thematic clusters and corresponding EU grant allocations, enabling a comparative assessment of project density and financial distribution patterns across Europe. This way, we were able to achieve a holistic representation of Erasmus+ activities where regional patterns can be easily spotted.

Results and discussions

By analysing the generated heatmaps, we can observe an important difference between Western/Northern and Eastern/Southern European regions, highlighted also by (Bryła & Domański, 2014) and (King & Lulle, n.d.). Higher participation rates and larger grant allocations are concentrated in wealthier countries like France, Germany, and Spain, which host well-established research infrastructures while Eastern and Southern regions are not as popular due to the language barriers and limited institutional capacity.

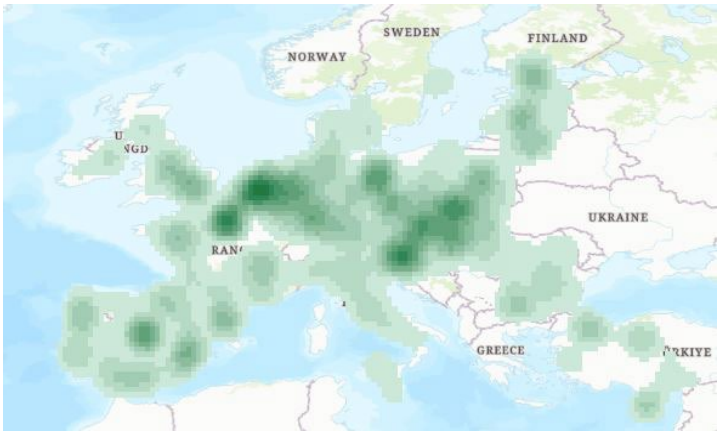

Figure 1. Generated Heatmap kernel density for Cluster 3

Source: Authors' own research.

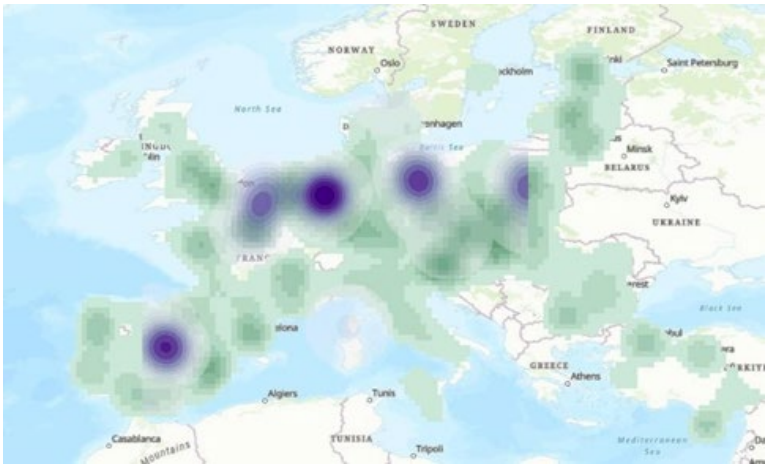

Figure 2. Generated Heatmap kernel density for Cluster 3 along with the normalised grant distribution layer

Source: Authors' own research.

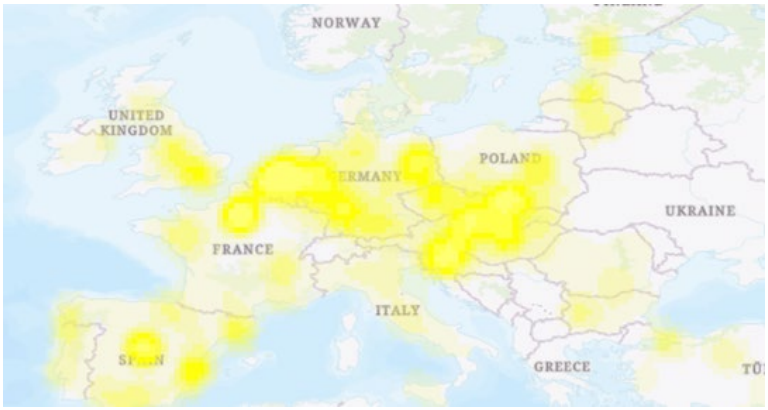

Figure 3. Generated Heatmap kernel density for Cluster 1

Source: Authors' own research.

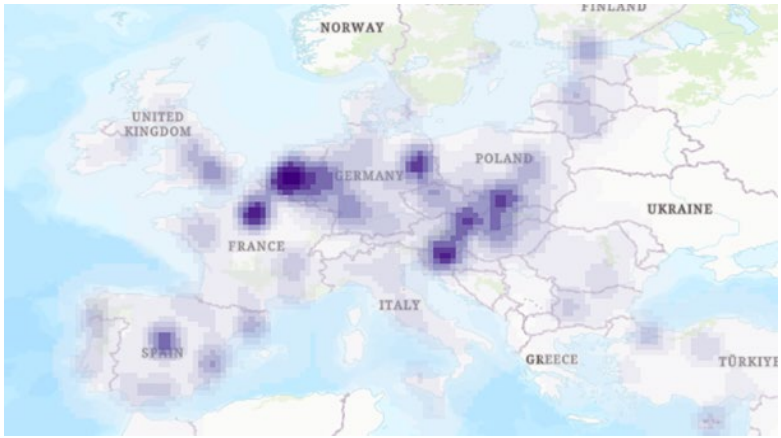

Figure 4. Generated Heatmap kernel density for Cluster 2

Source: Authors' own research.

Figure 5. Generated Layer for normalized grant allocation in Euros

Source: Authors' own research.

Our heatmaps (Figures 1–5) show that grant awards are heavily clustered in major cities within Western Europe, while Eastern regions display lower project density and funding. Post-socialist countries often struggle with administrative hurdles and co-financing requirements. While Erasmus+ provides some support for capacity-building, Horizon Europe initiatives such as Twinning, Teaming, and ERA Chairs specifically aim to address research disparities by directing resources toward underperforming regions. These findings align with existing research (Nascimbeni et al., 2021), reinforcing the need for more equitable funding distribution and enhanced support for less-developed areas.

Table 1. Summary Statistics of Countries with Coordinated Projects, Total Grants, and Largest Awarded Grants

Coordinator's country	Number of Projects	Total Grant (Euros)	Largest Grant (Euros)
ES	1977	101,092,222	3,019,160
DE	1878	131,830,763	2,651,000
FR	1695	111,167,770	2,410,000
PL	1112	81,026,912	3,464,883
TR	881	66,646,758	2,356,552
IT	855	10,5904,288	3,600,025
UK	759	71,496,651	2,837,000
RO	483	4,1604,141	1,877,019
AT	473	25,081,213	2,122,800
CZ	443	26,943,197	1,973,445
BE	402	32,862,722	2,779,000
HU	392	23,577,652	1,354,890
NL	338	38,657,330	2,165,000
SK	336	17,595,746	1,438,795
FI	333	20,071,498	776,931
PT	316	28,481,585	1,352,990

Coordinator's country	Number of Projects	Total Grant (Euros)	Largest Grant (Euros)
BG	308	17,720,950	873,885
SE	291	19,711,547	1326,000
LV	282	11,552,022	630,625
LT	280	16,247,471	1,178,326
SI	255	10,645,649	3,427,758
HR	221	10,271,608	2,022,206
EE	201	8,714,876	921,215
DK	175	13,594,941	1,010,551
IE	159	10,661,245	1,121,625
CY	145	5546,697	98,2520
LU	65	2,691,478	860,631
MT	59	3,119,160	110,1002

Source: Authors' own research.

An expanded view of total grant sums, largest single grants, and the number of projects per country strengthens the overall picture of uneven funding allocations and participation across Europe. Germany ranks highest in total Erasmus+ funding with over 131 million euros spread across 1,878 projects, yet its single largest grant is approximately 2.65 million, smaller than Italy's top single award of 3.6 million. This suggests a balanced grant distribution in Germany, where many medium-sized projects are fund ed rather than a small number of high-value initiatives, which aligns with the findings studied by (Finn & Darmody, 2017). France and Spain each surpass 100 million euros in total grants, being actively involved in Erasmus+ programs. France's largest individual award reaches 2.41 million, whereas Spain's stands at approximately 3 million, indicating that both countries host large-scale and mid- level projects. Among Eastern and Central European states, Poland stands out with a considerable total grant amount of 81 million euros over 1,112 projects, accompanied by a large single grant of 3.46 million. This shows that some Eastern countries can win high-value projects.

The broader data show that many Eastern and Southern European countries (e.g., Romania, Bulgaria, or Croatia) trend lower in both total funding and largest single awards, as it is also suggested by (Birle et al., 2020). Romania, for instance, has a total of around 41.6 million euros for 483 projects, and Bulgaria stands at 17.7 million across 308 projects, with largest single grants below the million-euro mark. These differences confirm what the heatmaps showed: W estern and Northern countries have more projects and receive higher total funding (Concilio et al., 2022). In contrast, many Eastern and Southern countries have fewer projects and smaller overall budgets, even if they occasionally secure a large grant.

Conclusion

Taken together, the scholarly literature on Erasmus mobility programs and Horizon-funded research collaborations underscores the multidimensional impact of these EU initiatives on migration and mobility patterns. Short-term study and research placements have immediate benefits for personal development, language skill acquisition, and cross-cultural exchange. In the longer term, participants often enjoy heightened career prospects, with some choosing to settle

permanently abroad and others returning home with a renewed sense of transnational connection. As Europe continues to integrate, debates around balancing open mobility with regional equity persist. On one hand, mobile students and researchers expand their horizons, gain language skills, and encourage the cross-pollination of ideas. On the other hand, the possibility of "brain drain" or an unequal concentration of knowledge resources remains a pressing concern.

This study has certain limitations. First, the data analysis relied primarily on publicly available Erasmus+ project data, which may not capture all nuances of funding decisions or project outcomes. Additionally, the clustering algorithms (Probierz et al., 2022) used were based on textual descriptions of project themes, which can be subjective and influenced by inconsistencies in project documentation. Another limitation involves the static nature of the data. Since funding trends and institutional participation evolve over time, longitudinal studies that track changes in participation and funding allocation could offer deeper insights. Future research could explore more advanced NLP models or deep learning techniques to refine topic clustering and improve thematic accuracy. Future research should also consider conducting interviews with project coordinators and policymakers to understand the qualitative aspects of institutional capacity and funding barriers.

Lastly, comparative analyses between Erasmus+ and Horizon Europe programs could help identify best practices for increasing participation from underrepresented regions. By addressing these limitations and expanding the scope of analysis, future research can further inform EU policies aimed at promoting educational equity and fostering greater cross-border cooperation across Europe.

Acknowledgements

This work was funded by the EU's NextGenerationEU instrument through Romania's National Recovery and Resilience Plan - Pillar III-C9-I8, managed by the Ministry of Research, Innovation, and Digitalization, as part of the project titled "CauseFinder: Causality in the Era of Big Data and AI and its Applications in Innovation Management," contract no. 760049/23.05.2023, code CF 268/29.11.202.

References

Birle, P., Göbel, B., Boyer, M., & Krusche, J. (2020). Mobility, diversity, inequality, sustainability: cross-cutting issues of cultural, scientific and social relations between the European Union and Latin America/the Caribbean. Eu-LAC Foundation.

Bryła, P., & Domański, T. (Eds.). (2014). The impact of the Erasmus programme on mobility and employability. Wydawnictwo Uniwersytetu Łódzkiego.

Concilio, G., Costa, G., Karimi, M., Vitaller Del Olmo, M., & Kehagia, O. (2022). Co-Designing with Migrants' Easier Access to Public Services: A Technological Perspective. Social Sciences, 11(2), 54. https://doi.org/10.3390/socsci11020054

Dagen, T., & Fink-Hafner, D. (n.d.). Impact of Globalisation on Internationalisation of Universities.

Finn, M., & Darmody, M. (2017). Examining student immobility: A study of Irish undergraduate students. Journal of Higher Education Policy and Management, 39(4), 423–434. https://doi.org/10.1080/1360080X.2017.1335265

Greco, S. (2023). The Role of Higher Education in the EU Development Cooperation Policy Stefano Greco. In N. Inamdar & P. Kirloskar, Reimagining Border in Cross-border Education (1st ed., pp. 237–259). Routledge. https://doi.org/10.4324/9781003427827-12

Herz, A., Díaz-Chorne, L., Díaz-Catalán, C., Altissimo, A., & Carignani, S. S. (2018). Are you mobile, too? The role played by social networks in the intention to move abroad among youth in Europe. Migration Letters, 16(1), 93–104. https://doi.org/10.33182/ml.v16i1.622

King, R., & Lulle, A. (n.d.). International Youth Mobility and Life Transitions in Europe: Questions, Definitions, Typologies and Theoretical Approaches.

Li, Y., & Wu, H. (2012). A Clustering Method Based on K-Means Algorithm. Physics Procedia, 25, 1104–1109. https://doi.org/10.1016/j.phpro.2012.03.206

Nascimbeni, F., Burgos, D., Spina, E., & Simonette, M. J. (2021). Patterns for higher education international cooperation fostered by Open Educational Resources. Innovations in Education and Teaching International, 58(3), 361–371. https://doi.org/10.1080/14703297.2020.1733045

Netek, R., Pour, T., & Slezakova, R. (2018). Implementation of Heat Maps in Geographical Information System – Exploratory Study on Traffic Accident Data. Open Geosciences, 10(1), 367–384. https://doi.org/10.1515/geo-2018-0029

Pokojski, W., Panecki, T., & Słomska-Przech, K. (2021). Cartographic visualization of density: Exploring the opportunities and constraints of Heat Maps. Polish Cartographical Review, 53(1), 21–36. https://doi.org/10.2478/pcr-2021-0003

Probierz, B., Kozak, J., & Hrabia, A. (2022). Clustering of scientific articles using natural language processing. Procedia Computer Science, 207, 3449–3458. https://doi.org/10.1016/j.procs.2022.09.403

Ryazanova, O., & McNamara, P. (2019). Choices and Consequences: Impact of Mobility on Research-Career Capital and Promotion in Business Schools. Academy of Management Learning & Education, 18(2), 186–212. https://doi.org/10.5465/amle.2017.0389

Taşkiran, S. F., & Kaya, E. (2022). ACADEMIC TEXT CLUSTERING USING NATURAL LANGUAGE PROCESSING. Konya Journal of Engineering Sciences, 10, 41–51. https://doi.org/10.36306/konjes.1081213

Van Duynhoven, A., & Dragićević, S. (2024). The Geographic Automata Tool: A New General-Purpose Geosimulation Extension for ArcGIS Pro. Applied Sciences, 14(15), 6530. https://doi.org/10.3390/app14156530