



Research article

Combining spatial and aspatial methods to identify European industrial leagues: Development perspectives[☆]

Z.T. Kosztyán^{a,b,1,*}, Z. Banász^c, M.T. Kurbucz^d, A.I. Katona^a,
A. Pribojszki-Németh^a, Á. Jakobi^e

^a Department of Quantitative Methods, University of Pannonia, Egyetem Street 10, H-8200 Veszprém, Hungary

^b Institute of Advanced Studies, Kőszeg (iASK), Charnel Street 14, H-9730 Kőszeg, Hungary

^c Department of Economics, University of Pannonia, Egyetem Street 10, H-8200 Veszprém, Hungary

^d Institute for Global Prosperity, University College London, 9-11 Endsleigh Gardens, London, WC1H 0EH, United Kingdom

^e Department of Regional Science, Eötvös Loránd University, Pázmány Péter sétány 1/C, Budapest, H-1117, Hungary

HIGHLIGHTS

- Biclustering identifies European industrial leagues from regional profiles.
- Spatial analysis distinguishes clustered from geographically diffuse patterns.
- The method identifies regionally specific and shared development priorities.
- Regional conditions and sectoral specialization shape Industry 4.0 readiness.
- NUTS 2 evidence supports targeted regional and sector-specific policies.

ARTICLE INFO

Keywords:

Industrial clusters
Biclustering analysis
Spatial autocorrelation
Regional development
Industry 4.0
European regional analysis
Technological specialization
NUTS 2 regions

ABSTRACT

This study addresses a critical need in regional development planning by combining aspatial bi-clustering with spatial autocorrelation analysis to examine European industrial competitiveness. Focusing on NUTS 2 regions, we move beyond broad categorizations by identifying data-driven “industrial leagues”—clusters of regions with similar strengths and weaknesses across 71 retained Industry 4.0 (I4.0)+ readiness indicators covering 320 regions. The proposed approach groups regions with similar economic and technological characteristics and then evaluates the spatial dependence of the resulting patterns. The analysis identifies seven regional leagues, each characterized by specific strengths and development challenges. For example, regions in Western Europe, such as Inner London–West and Brussels, exhibit high employment-based I4.0+ indicator values, whereas regions including Bucharest and Bratislava exhibit weaknesses in Research

[☆] Project no. K 143482 has been implemented with support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the K 22 “OTKA” funding scheme. M.T.K. received support from the ÚNKP-22-4-II-CORVINUS-55 New National Excellence Program of the Ministry for Culture and Innovation from the National Research, Development and Innovation Fund.

* Corresponding author at: Department of Quantitative Methods, University of Pannonia, Egyetem Street 10, H-8200 Veszprém, Hungary.

Email addresses: kosztyan.zsolt@gtk.uni-pannon.hu (Z.T. Kosztyán), banasz.zsuzsanna@gtk.uni-pannon.hu (Z. Banász), m.kurbucz@ucl.ac.uk (M.T. Kurbucz), katona.attila@gtk.uni-pannon.hu (A.I. Katona), pribojszki-nemeth.aniko@gtk.uni-pannon.hu (A. Pribojszki-Németh), jakobi.akos@ttk.elte.hu (Á. Jakobi).

URLs: <https://www.gtk.uni-pannon.hu/zsoltkosztyan/> (Z.T. Kosztyán), <https://www.gtk.uni-pannon.hu/zsuzsannabanasz/> (Z. Banász), <https://profiles.ucl.ac.uk/102765-marcell-t-kurbucz> (M.T. Kurbucz), <https://www.gtk.uni-pannon.hu/dr-attila-katona/> (A.I. Katona), <https://www.gtk.uni-pannon.hu/nemethaniko/> (A. Pribojszki-Németh), <https://ttk.elte.hu/en/staff/dr-akos-jakobi> (Á. Jakobi).

¹ First author

and Development (R&D) and workforce skills. Regional adoption conditions, including infrastructure quality and sectoral specialization, are associated with I4.0 readiness. The methodology also identifies sectoral concentrations within leagues, such as automotive manufacturing in Germany and ICT in Ireland, with implications for targeted development policy. Some regional groups combine strong employment indicators with weaker science-and-technology and R&D indicators, indicating a need for targeted investment in research infrastructure. Other groups share weak media-presence indicators and may require broader intervention. By identifying the indicators most closely associated with each league, this study provides policy-makers with evidence for tailored strategies that combine location-based and sector-specific measures.

Nomenclature		iBBiG	Iterative Binary Biclustering of Gene sets
<i>List of symbols and abbreviations</i>		KIBS	Knowledge-Intensive Business Services
BicARE	Biclustering Analysis and Results Exploration	LISA	Local Indicator of Spatial Association
C	contribution	O	research objective
EEA	European Economic Area	R&D	Research and Development
EFTA	European Free Trade Association	RQ	research question
EU	European Union	RRF	Recovery and Resilience Facility
I4.0	Industry 4.0		

1. Introduction

Well-designed regional development strategies are crucial for equitable and durable industrial growth [84]. Their implementation requires economic policies tailored to the needs and capabilities of individual regions, thereby strengthening local industrial competitiveness and supporting economic expansion [34]. Because they reduce regional inequalities, such strategies can foster social cohesion and distribute the benefits of industrialization more equitably. Regional policies can also support employment, skills development, and adaptation to broader changes such as digitalization and climate change, thereby contributing to a more inclusive and resilient economy [84]. Nevertheless, persistent regional disparities in Europe threaten economic progress, social cohesion, and political stability, indicating that current policy mechanisms remain insufficient [33,49].

Spatial and aspatial development strategies play complementary roles in strengthening I4.0 readiness [22]. Spatial strategies address the geographical organization of development, for example, by establishing innovation hubs and technology parks that foster collaboration [59]. Such concentrations can attract high-technology industries and provide infrastructure for advanced manufacturing and digital technologies [32,55]. Aspatial strategies address factors that are not inherently location bound, including policy frameworks, digital infrastructure, and workforce development [19]. Combining these approaches can address both the place-specific and systemic requirements of I4.0 adoption.

Recent research highlights the value of regional development strategies that combine spatial and aspatial evidence [88]. The incorporation of both location-based and nonspatial components can provide a more complete account of regional economic dynamics and support better targeted interventions [59]. However, methods for integrating these perspectives remain limited. Therefore, policy-makers and researchers need a coherent framework that distinguishes geographically concentrated challenges from those shared across distant regions.

Existing regional development frameworks capture only part of the complexity of I4.0 transformation. Simplified linear models may overlook interdependencies among technology, geography, and socioeconomic conditions [43]. Other approaches treat spatial and aspatial factors separately, even though technological capability, regional infrastructure, workforce skills, and innovation ecosystems are interdependent [24]. Predominantly national or broadly regional assessments also lack the NUTS 2-level resolution needed to represent heterogeneous readiness across Europe [25]. Therefore, an appropriate policy framework must identify groups with similar development profiles and then determine whether those profiles are geographically concentrated. This distinction indicates whether an intervention should be place-based or sector-wide.

This study addresses that need by combining aspatial biclustering [89] with Moran’s spatial autocorrelation analysis [67]. Biclustering identifies industrial leagues containing regions and indicators with similar profiles; spatial testing then evaluates the strength and significance of geographical dependence under a fully specified neighborhood definition. The sequence distinguishes geographically clustered challenges from league-membership patterns with weak or undetectable spatial concentration and thereby informs the balance between place-based and sector-wide interventions.

An *industrial league* simultaneously identifies a set of regions and the indicators on which those regions exhibit a similar profile. Therefore, the method supports development strategies tailored to comparable groups of NUTS 2 regions. These standardized territorial units provide a meaningful scale for examining regional disparities and the effects of Industry 4.0 policy within the European Union (EU). Their granularity facilitates a nuanced analysis of spatial patterns and targeted interventions. Combining aspatial league identification with a spatial analysis reveals regional strengths and weaknesses and helps distinguish place-based development opportunities from broader sector-specific priorities.

Table 1
The NUTS 2 levels examined.

	Country		NUTS 2 regions	Capital name	NUTS 2		Country		NUTS 2 regions	Capital name	NUTS 2
	code	name					code	name			
1.	AT	Austria	9	Vienna	AT13	19.	LU	Luxembourg	1	Luxembourg	LU00
2.	BE	Belgium	11	Brussels	BE10	20.	MT	Malta	1	Valletta	MT00
3.	BG	Bulgaria	6	Sofia	BG41	21.	ME	Montenegro	1	Podgorica	ME00
4.	HR	Croatia	2	Zagreb	HR04	22.	NL	Netherlands	12	Amsterdam/The Hague	NL32 / NL33
5.	CY	Cyprus	1	Nicosia	CY00	23.	MK	North Macedonia	1	Skopje	MK00
6.	CZ	Czechia	8	Prague	CZ01	24.	NO	Norway	7	Oslo	NO01
7.	DK	Denmark	5	Copenhagen	DK01	25.	PL	Poland	17	Warsaw	PL91
8.	EE	Estonia	1	Tallinn	EE00	26.	PT	Portugal	7	Lisbon	PT17
9.	FI	Finland	4	Helsinki	FI1B	27.	RO	Romania	8	Bucharest	RO32
10.	FR	France	23	Paris	FR10	28.	RS	Serbia	4	Belgrade	RS11
11.	DE	Germany	37	Berlin	DE30	29.	SK	Slovakia	4	Bratislava	SK01
12.	EL	Greece	13	Athens	EL30	30.	SI	Slovenia	2	Ljubljana	SI04
13.	HU	Hungary	8	Budapest	HU11	31.	ES	Spain	17	Madrid	ES30
14.	IS	Iceland	1	Reykjavík	IS00	32.	SE	Sweden	8	Stockholm	SE11
15.	IE	Ireland	3	Dublin	IE06	33.	CH	Switzerland	7	Bern	CH02
16.	IT	Italy	21	Rome	IT14	34.	TR	Turkey	26	Ankara	TR51
17.	LV	Latvia	1	Riga	LV00	35.	UK	United Kingdom	41	London	UK13-UK17
18.	LT	Lithuania	2	Vilnius	LT01	Total			320		

Given these research gaps, we formulate the following research objectives (Os).

- O₁ Identify European industrial leagues that jointly specify NUTS 2 regions and I4.0+ indicators with common strengths and weaknesses.
- O₂ Identify the indicators that require improvement for a NUTS 2 region to advance to a higher-performing league.
- O₃ Support decision-makers in distinguishing regionally specific development needs from challenges shared across regions.

The geographical scope comprises the EU-27 Member States, the European Economic Area (EEA) members Iceland and Norway, and countries that maintain NUTS-compatible statistical regions: Switzerland, Turkey, the United Kingdom, Montenegro, North Macedonia, and Serbia (Table 1). This diverse group contains economies at different levels of industrial development and provides a useful setting in which to examine how regional policies can promote balanced and sustainable growth. European regional policies support innovation, competitiveness, and the reduction of economic disparities through targeted investment and coordinated programs [48].

I4.0 represents an intelligent network of machines and processes for industry that emphasizes mass customization, responsiveness, and production tailored to individual consumer needs [6,66]. This paradigm focuses on the design, manufacturing, and provision of customized products and services according to specific consumer requirements, enabling unprecedented levels of flexibility and efficiency in production processes [45]. The integration of these new value-adding technologies could increase innovation and competitiveness significantly across regions [see, e.g., 7,29,40,53]. To exploit this advantage, the European Commission must face the challenge of the uneven geographical distribution of I4.0 research initiatives across European regions [see, e.g., 26,29,37,65]; nevertheless, countries of the EU emphasize different areas in their national policies related to I4.0 [81].

Industry readiness indices evaluate how well organizations can address the challenges and opportunities associated with I4.0 [78]. They commonly cover technological capability, digital infrastructure, workforce competence, organizational culture, and strategic alignment with I4.0 principles.

Readiness indices have been developed for diverse sectors and geographical settings. For example, Maisiri et al. [63] assessed South African industries and emphasized the need for evaluations adapted to regional conditions. Ghafar et al. [42] examined small and medium-sized aviation enterprises in Malaysia and identified sector-specific challenges and opportunities.

The development of I4.0 readiness indices draws on computer science [86], management, and business research [73]. This interdisciplinary basis supports comprehensive evaluation models and helps organizations identify gaps, respond to technological change, and strengthen long-term competitiveness [19].

The I4.0+ indicator system evolved from earlier I4.0 readiness assessments and extends beyond technological metrics to a 101-indicator framework for multidimensional digital-transformation readiness across European regions [46]. Developed from multiple European statistical sources from 2008 to 2019, it covers population education, youth employment, international student mobility (Erasmus), graduate specialization, scientific and technological capacity, and media presence [31]. Throughout this paper, “I4.0” denotes the general industrial paradigm, whereas “I4.0+” denotes this specific indicator system and its associated dataset.

We use the I4.0+ indicator system [31,46,47] to identify homogeneous leagues that jointly contain NUTS 2 regions and indicators (O₁). League composition and overlap reveal potential development priorities (O₂), whereas spatial autocorrelation indicates whether those priorities support place-based or geographically broader policy responses (O₃).

Although the I4.0+ data describe conditions in 2018, structural regional gaps in digital infrastructure, R&D capacity, and workforce skills tend to persist over multiyear periods [49]. Therefore, the identified leagues provide a historical baseline rather than a

current ranking. The principal contribution is methodological: the combination of aspatial biclustering and spatial autocorrelation can be reapplied when newer I4.0+ data become available, as discussed in [Section 7](#).

2. Literature review

Building upon the identified need for integrated spatial and aspatial approaches to regional development, the literature offers valuable insights into both types of strategies, although often in isolation. Regionally specific development strategies frequently emphasize leveraging unique regional strengths through approaches such as smart specialization and cluster development as well as the prioritization of spatial factors, such as geographic characteristics and the proximity of economic actors [71]. Conversely, sector-specific development strategies [82] focus on improvements in workforce skills, innovation, and sustainable practices [3], highlighting the importance of knowledge, technology, and environmental considerations, irrespective of spatial factors. The following section explores these approaches in detail, revealing the limitations of each when considered in isolation and underscoring the potential benefits of their integration.

2.1. Theoretical framework

Regional development strategies are theoretically anchored in two seminal conceptual frameworks that transcend mere descriptive categorization. Rooted in evolutionary economic geography and endogenous growth theory, smart specialization theory posits that regions achieve sustainable competitive advantages by concentrating resources on distinctive capabilities that leverage existing knowledge bases and institutional assets [10,16,17]. This approach fundamentally challenges the traditional convergence paradigm by suggesting that regional development should build upon heterogeneous territorial assets rather than attempting to replicate successful models from other contexts. Complementarily, grounded in Porter [75]’s seminal work on competitive advantage and later refined in the new economic geography literature by Krugman [52] and Martin and Sunley [64], cluster development theory demonstrates that geographic proximity facilitates knowledge spillovers, reduces transaction costs, and creates dynamic externalities that enhance collective innovation capacity. These theoretical frameworks are not merely policy tools but represent fundamental shifts in understanding how spatial agglomeration effects and knowledge-based specialization interact to create sustainable regional competitive advantages. The theoretical synthesis of these approaches suggests that effective regional development requires attention to both the specialization process (which regions should focus on) and the spatial clustering mechanisms (how proximity enhances performance), providing the conceptual foundation for integrating spatial and aspatial analytical approaches in contemporary I4.0 development strategies.

The urgency of developing robust regional development strategies for I4.0 readiness has been amplified by several high-level European policy initiatives that shape the contemporary policy landscape directly. The EU’s strategic autonomy agenda, particularly in terms of critical technologies and digital infrastructure, requires a nuanced understanding of regional capabilities and dependencies to reduce external vulnerabilities while building internal resilience [27]. The postpandemic Recovery and Resilience Facility (RRF), which allocates 723.8 billion EUR across member states with explicit mandates for digital and green transitions, presents both unprecedented opportunities and complex coordination challenges that demand precise regional targeting to maximize impact [27]. This massive funding mechanism creates a critical need for analytical frameworks that can identify which regions are best positioned to absorb and effectively utilize digital transformation investments while addressing the risk of exacerbating regional disparities.

Furthermore, the tension between national industrial policies, which are often driven by domestic political priorities and industrial legacies, and EU-wide cohesion goals creates a complex, multilevel governance challenge. Regional development strategies must balance local competitive advantages with broader European integration objectives [49]. This challenge is particularly acute in the context of I4.0, in which technological spillovers and supply chain interdependencies mean that a regional development decision in one area can have cascading effects across the European economic space. This policy complexity underscores the need for analytical approaches that can support national strategic autonomy while advancing EU-wide cohesion. Therefore, the identification of contexts in which spatially targeted or geographically broad development strategies are appropriate is both an academic question and a pressing policy concern.

Place-based strategies build on a region’s distinctive capabilities through approaches such as smart specialization and cluster development [59]. They account for physical characteristics and the proximity of economic actors [57], and they direct resources toward activities in which a region has a comparative advantage. This emphasis differs from traditional policies that seek to equalize outcomes through transfers and other geographically broad instruments [68]. Cluster development can connect firms and institutions within an industry, facilitating collaboration, innovation, and growth [76]. Although these approaches can reduce regional disparities and foster local innovation [38], an assessment of competitiveness must also consider factors that are not inherently spatial [4].

Sector-specific strategies emphasize workforce skills, innovation, and sustainable practices [20]. Knowledge, technology, and environmental conditions can influence regional performance even when they are not concentrated geographically [76]. Workforce programs can close sectoral skill gaps [61], whereas investment in R&D supports technological progress and global competitiveness [72]. The research on industrial specialization also shows that regions and cities can develop concentrated capabilities in particular industries [36,78,79]. Nevertheless, implementing these strategies requires coordination across sectors and levels of government to address inequality, infrastructure gaps, sociocultural barriers, environmental constraints, and governance challenges [39].

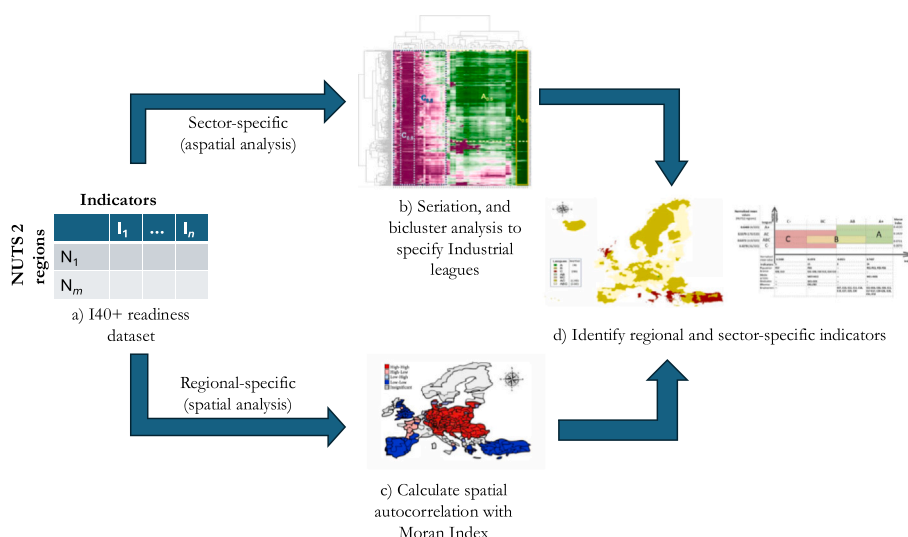


Fig. 1. Conceptual framework.

2.2. Conceptual framework

The conceptual framework of the study is shown in Fig. 1.

The I4.0+ readiness database contains development indicators for NUTS 2 regions (Fig. 1a). NUTS is the geographical classification used to subdivide the economic territory of EU Member States. NUTS 2 units are the basic regions for regional policy and generally contain 800,000 to 3 million inhabitants.² The classification supports the collection and harmonization of European regional statistics and socioeconomic analyses of regions.³ Because NUTS 2 regions are central to EU cohesion policy, they provide an appropriate scale for studying regional industrial performance and informing investment decisions [1,54].

The columns of the data matrix are the I4.0+ readiness indicators [31], described in Section 3.1. The analytical sequence has two stages. First, seriation reorders regions and indicators to reveal structure, and biclustering identifies submatrices containing regions and indicators with similar profiles [13,44,56] (Fig. 1b). Second, Moran statistics are used to assess the spatial dependence of the resulting memberships and indicators [67] (Fig. 1c). Integrating these stages indicates where place-based policies are warranted and where a broader sector-specific response may be more appropriate (Fig. 1d).

Regional Knowledge-Intensive Business Services (KIBS) and R&D initiatives can strengthen the capacity of NUTS 2 regions to specialize in emerging industries, promote innovation, and support economic progress [12,28]. Analyses at this scale can also reveal interregional supply chain connections and recurring transport links that are relevant to European economic integration [1,14]. Therefore, we use NUTS 2 regions as the unit of analysis (Fig. 1a).

Regions leading in earlier industrial transformations also tend to rank highly in I4.0 [26,65]. Regions with stronger capabilities in I4.0-related technologies may also be better positioned to diversify into new technological activities [8,43]. Nevertheless, regional potential varies substantially across Europe, and many regions remain weakly positioned [60]. Technological capabilities and spatial proximity strongly affect the diffusion of robotics and 3D printing, whereas big data and the Internet of Things are more geographically dispersed [29]. These differences motivate the research question (RQ)₁: *Which common strengths and weaknesses characterize European NUTS 2 regions, and which regions face similar development challenges?* Industrial biclusters, or leagues, answer this question by identifying groups of regions with common indicator profiles (Fig. 1b).

Research by Capello and Lenzi [21] operationalized Industry 4.0 technological transformation across European territories and reported that penetration into local economies depends on regional adoption conditions that reinforce sectoral specialization. Czvetkó et al. [31] developed the I4.0+ indicator system [see also 46,47] on the basis of measures of higher education and lifelong learning, labor markets, innovation, investment, and technology and applied it to EU NUTS 2 regions. They identified Helsinki–Uusimaa as the most advanced region of the EU, followed by Prague and Berlin, and reported that I4.0+ readiness is correlated with regional economic, innovation, and competitiveness indices. These findings motivate RQ₂: *How can we identify NUTS 2 regions and the I4.0+ indicators that should be developed simultaneously to improve regional performance?* This question is important because spatial development is sometimes treated merely as a local combination of parallel sectoral policies, leaving place-based considerations secondary to sectoral priorities [11]. Therefore, spatial autocorrelation analyses are needed to distinguish geographically concentrated strengths and weaknesses from those distributed across distant regions (Fig. 1c).

Development policy is often focused on either indicators [see 70] or regions [see 27]; the integration of vertical (sectoral) and horizontal (spatial) dimensions remains difficult [80]. Using Eurostat data, Castelo-Branco et al. [25], like Czvetkó et al. [31], identified

² https://www.destatis.de/Europa/EN/Methods/Classifications/OverviewClassification_NUTS.html.

³ <https://www.europarl.europa.eu/factsheets/en/sheet/99/common-classification-of-territorial-units-for-statistics-nuts>.

homogeneous groups and substantial disparities in national I4.0 readiness. Therefore, further questions remain (RQ₃): *Which development needs are regionally specific, which are shared across regions, and how can that distinction guide place-based and sector-specific action plans?* Answering these questions requires linking the indicators identified by biclustering to evidence about their spatial concentration (Fig. 1d).

The research objectives and questions address these gaps. Although regional policy increasingly requires combined spatial and aspatial evidence, the methodological support for integrating the two remains limited. Biclustering has been applied in marketing [85] and economics [77,87], but it remains rarely used in regional development. Kosztyán, Banász, Csányi and Telcs [50] demonstrated how biclustering can form leagues of countries; subsequent applications have considered universities [9] and individuals [51]. The present study extends this concept to European NUTS 2 regions.

The study aims are addressed by answering the following research questions.

- RQ₁ With respect to I4.0, what are the common strengths and weaknesses of European NUTS 2 regions, and which NUTS 2 regions face the same development challenges?
- RQ₂ Which NUTS 2 regions and I4.0+ indicators should be addressed jointly to improve regional performance?
- RQ₃ Which development needs are regionally specific, which are shared across regions, and how can that distinction guide place-based and sector-specific action plans?

The biclustering method identifies both European regions (RQ₁) and the indicators associated with their development needs (RQ₂). Moran's spatial autocorrelation analysis then helps distinguish needs that support place-based action from those that support broader sector-specific action (RQ₃).

Addressing the research questions yields the following contributions (Cs) to the current literature.

- C₁ The proposed method identifies industrial leagues that jointly group NUTS 2 regions and I4.0+ indicators. Regions within a league have comparable profiles across a homogeneous indicator subset, providing a basis for coordinated action plans.
- C₂ The method identifies the region–indicator combinations that permit policy-makers to focus on regional development priorities.
- C₃ The proposed combination of aspatial biclustering and spatial-dependence testing can indicate which I4.0+ development needs have a strong regional concentration and which are distributed more diffusely.

3. Data and methodology

We use a data-driven approach to identify European industrial leagues (O₁)—biclusters that jointly specify similar NUTS 2 regions and I4.0+ indicators—and to assess their development implications (O₂–O₃). The analysis is exploratory rather than based on a prespecified causal model, but it is guided by the stated research objectives and questions.

3.1. I4.0+ readiness dataset

Although an increasing number of studies have examined I4.0+ indicators and regional development across Europe, policy-oriented empirical applications remain limited. We identify strengths and weaknesses in regional I4.0+ profiles through biclustering and highlight potential directions for improvement.

Our analysis builds on an open-access secondary dataset⁴ compiled by Honti et al. [47]. This dataset covers 101 I4.0-related indicators aggregated at the NUTS 2 level over a 12-year period (2008–2019) and was used to construct a composite index of I4.0 development across European regions [46].

The biclustering method applied in this study requires a cross-sectional dataset, which involves selecting a single year and addressing incomplete data. We chose the year 2018 for three main reasons: (i) it had the *lowest rate of missing values* among the five most recent years; (ii) it offered the broadest indicator coverage across NUTS 2 regions; and (iii) it coincided with a key phase in I4.0 adoption marked by rising industrial automation, workforce transformation, and global economic uncertainty due to trade tensions [5,15,24]. Companies that adopted I4.0 technologies in 2018 positioned themselves for long-term competitiveness, and sustainability strategies gained momentum during this period.

The initial 2018 table contained 340 regions and 101 indicators, of which 10,977 of 34,340 cells (31.97%) were unavailable. We applied a temporally ordered substitution rule that retained the 2018 value when observed and otherwise used the first available value in the following fixed order: 2019, 2017, 2016, 2015, 2014, and 2013. In a reconstruction from the released 2008–2019 panel, this rule filled 6026 initially missing cells: 0 from 2019, 289 from 2017, 3520 from 2016, 1265 from 2015, 255 from 2014, and 697 from 2013. The procedure preserves observed regional values and avoids model-based interpolation, but it assumes that an adjacent-year observation is an acceptable proxy for the unobserved 2018 value. This assumption is more credible for slowly changing structural indicators than for volatile measures.

Region and indicator screening decreased the sample to 320 regions and 71 indicators. At that stage, 740 of 22,720 cells (3.26%) remained unavailable. Following the archived workflow, these cells were encoded as zeros before min–max normalization and biclustering. The missing data were concentrated: the most affected indicators were S13 (70 cells), ER2 (58), ER1 (56), E17 (46), E20

⁴ Available at <https://data.mendeley.com/datasets/23gwn43ygp/1> (retrieved: 22 September 2022).

Table 2
The indicator categories examined.

Indicator topics			Indicator unit		No. indicators
			absolute	relative	
1.	P	Population: percentage of the population aged 25–64 and 30–34 with different levels of education.		X	7
2.	E	Employment: employment rates of young people aged 15–34 years		X	32
3.	ER	Erasmus: number of students in the Erasmus program.	X		2
4.	G	Graduates: number of graduates in specific sciences.	X		3
5.	S	Science: rate of scientists and engineers, number of Industry 4.0 studies, number of research institutes.	X	X	15
6.	M	Media articles: number and average tone of the media articles, queried in the GDELT database matching certain expressions.	X	X	12
Total					71

(42), and E05 (40), whereas the most affected regions were FRY3 (21 cells); UKM6 and PT30 (20 each); and UKM5, FRF2, FRY1, and ITC2 (19 each). Complete indicator- and region-level counts are included in the reproduction archive.

Zero encoding has directionally dependent consequences. Of the 43 indicators with at least one residual missing value, zero lay below the observed range for 35, within the observed range for five media-tone indicators, and at the observed minimum for three graduate-count indicators. With respect to the first group, zero became an artificial minimum: missing cells were treated as maximally weak after normalization, and the normalized values of observed cells shifted upward. The maximum indicator-level shift among the observed normalized values was 0.516. In the reversed matrix used to identify League C, these same zero-coded cells became maximally high; therefore, they could reinforce low-performance membership. Across the 15 archived league-threshold blocks, zero-coded cells accounted for 0.99–5.75% of the included cells and were generally more prevalent in League C. Thus, the encoding does not dominate any league block, but it can affect marginal memberships and should be considered when low-performance leagues are interpreted. The released missing-data audit identifies every affected cell and reproduces these summaries. [Section 7](#) discusses the remaining inferential constraint.

Several NUTS 2 regions were excluded prior to the analysis because of data quality or definitional issues:

- Aggregated regions representing country-level or supranational averages (e.g., EU15, EU27, EU28, EA19, and the European Free Trade Association (EFTA)).
- Extraregional zones, such as maritime areas or airspace, identified by NUTS 2 codes ending in XX or ZZ.
- Regions with more than 30% missing data across indicators.

[Table 1](#) presents the included countries, the number of retained NUTS 2 regions per country, and the corresponding capital regions used to interpret regional I4.0+ profiles.

With respect to the columns, the original dataset contained 101 indicators. With respect to the remaining 320 NUTS 2 regions, we excluded the following indicators:

- Indicators with a missing data rate exceeding 30%.
- Absolute indicators for which a corresponding relative version was also available.

The final set of 71 indicators used in the analysis is categorized in [Table 2](#).

A description of each variable is provided in [Table A.1](#).

To improve comparability across NUTS 2 regions in different countries, absolute indicators were adjusted by population to produce relative values. In summary, biclustering was performed on the completed 2018 analysis matrix, which includes 71 indicators for 320 NUTS 2 regions. Supplementary Data S1 lists the complete NUTS 2 codes and names; Supplementary Data S2 provides the complete indicator dictionary; Supplementary Data S3 contains the 320-by-71 matrix used in the analysis; Supplementary Data S4 contains the archived region and indicator league memberships; and Supplementary Data S5 contains the spatial input, weights, diagnostics, permutation results, figures, code, and session information.

The statistical unit is the NUTS 2 region ($n = 320$), not the 22,720 region-indicator cells. Cells within a region and repeated values of a given indicator across regions should not be treated as mutually independent observations. The country composition is also unbalanced: retained countries contribute between 1 and 41 regions ([Table 1](#)), so the results represent the included European regions rather than a country-balanced sample. Moreover, spatial dependence reduces the amount of independent information, with the extent varying by indicator and league; consequently, no single effective sample size applies to all of the analyses. Therefore, the inferential results are interpreted as exploratory diagnostics for the observed regional sample rather than as estimates from 22,720 independent observations.

To prepare the data for seriation, all values were normalized using min–max normalization as follows:

$$\tilde{x}_{i,j} = \frac{x_{i,j} - \min_j x_{i,j}}{\max_j x_{i,j} - \min_j x_{i,j}}, \quad (1)$$

Table 3Means and standard deviations of normalized values within and outside of each league under different thresholds (tr).

Conditions	Cells	League A		League B		League C	
		Mean	SD	Mean	SD	Mean	SD
$tr = 0.5$ $N = 110, M = 34$	within	0.7636	0.2058	0.3341	0.2796	0.2081	0.2120
	remaining	0.2498	0.2446	0.5638	0.3389	0.7288	0.2364
$tr = 0.6$ $N = 100, M = 30$	within	0.7956	0.1856	0.2973	0.2708	0.1687	0.1905
	remaining	0.2979	0.2676	0.5608	0.3368	0.7059	0.2442
$tr = 0.7$ $N = 90, M = 26$	within	0.8384	0.1407	0.2396	0.2327	0.1027	0.1578
	remaining	0.3556	0.2939	0.5589	0.3357	0.6592	0.2658
$tr = 0.8$ $N = 80, M = 22$	within	0.8684	0.1276	0.1775	0.1760	0.0656	0.1185
	remaining	0.4128	0.3123	0.5553	0.3349	0.6373	0.2777
$tr = 0.9$ $N = 70, M = 18$	within	0.9788	0.0226	0.1319	0.1498	0.0409	0.0769
	remaining	0.4868	0.3268	0.5491	0.3346	0.6199	0.2882

Note: The overall mean is 0.5260, and the overall standard deviation is 0.3407.

where $x_{i,j}$ is the value of indicator i in NUTS 2 region j .⁵

In addition, one of the applied Iterative Binary Biclustering of Gene sets (iBBiG) methods binarizes the normalized data with respect to a given threshold ($tr \in [0, 1]$):

$$b_{i,j} = \begin{cases} 1 & \text{if } \tilde{x}_{i,j} > Q_{tr}(\tilde{X}) \\ 0 & \text{if } \tilde{x}_{i,j} \leq Q_{tr}(\tilde{X}), \end{cases} \quad (2)$$

where Q_{tr} is the quantile. In a special case, $Q_{0.5} = Me$, where Me is the median. $b_{i,j} \in B_{tr}$ is a binarized cell value from the binarized matrix B_{tr} with a given threshold tr . Binarization is also applied to the reverse normalized data, where the reversed value is $\tilde{x}'_{i,j} = 1 - \tilde{x}_{i,j}$.

Note that Q_{tr} is computed globally over the entire normalized matrix $\tilde{X}$ rather than separately for each indicator. Because every indicator is first rescaled to $[0, 1]$ by Eq. (1), the global quantile respects each indicator's observed range but does not equalize the distributional shape, including skewness. Applying a logarithmic or Box–Cox transformation [18] before normalization or computing Q_{tr} separately for each indicator could reduce the influence of distributional asymmetry on league membership. We did not apply these alternatives and we identify them as future robustness checks (Section 7).

Because $Q_{0.5}$ coincides with the median (as noted above), a threshold $< tr = 0.5$ would classify below-typical cells as 'high,' which is not meaningful for identifying above-average development; therefore, we restrict the sensitivity analysis to $tr \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$, which is already reported in Table 3 and Figs. 8 and A.1.

3.2. Spatial invariance methods—from seriation to biclustering

Seriation is an exploratory combinatorial data analysis technique used to reorder data matrices (columns = indicators and rows = NUTS 2 regions). Seriation facilitates pattern discovery at three information levels simultaneously: local fragments of relationships, sets of organized local fragments of relationships, and an overall structural pattern [56]. Through seriation, both rows and columns are reordered simultaneously to group similar cells as closely as possible. In this study, this method is used to develop the *pattern of industrial development in NUTS 2 regions*. Similar cells (specified as regions corresponding to an I4.0+ indicator) are grouped as closely as possible.

After seriation, biclustering jointly selects rows (NUTS 2 regions) and columns (indicators) to form a homogeneous submatrix, hereafter termed a league. League A identifies comparatively high-performing region–indicator combinations, whereas League C identifies comparatively low-performing combinations.

We selected iBBiG [44] and Biclustering Analysis and Results Exploration (BicARE) [41] because their objectives complement the three-league framework of Kosztyán et al. (2019); Kosztyán et al. (2023) [50,51]. iBBiG identifies consistent binary blocks and is applied to the normalized matrix for high-performing League A and to the reverse-normalized matrix for low-performing League C. BicARE minimizes within-bicluster variance and identifies the intermediate League B. Together, the algorithms represent high values, intermediate homogeneous values, and low values while identifying the leagues without imposing a spatial constraint. Spatial dependence is tested only after the leagues have been selected.

The biclustering literature also includes the Cheng–Church mean-squared-residue method, plaid models, and spectral or information-theoretic methods such as ISA, FABIA, and SAMBA. These methods optimize different homogeneity criteria across row–column submatrices [62]. We did not use these methods because they do not jointly target the high/intermediate/low structure that

⁵ Since higher values indicate better development for all indicators except for indicator $x_1 = P01$ (percentage of the population with less than primary or lower secondary educational levels), only this indicator is reversed; furthermore, $\tilde{x}'_1 := 1 - \tilde{x}_1$.

the present league design needed. The entropy-based iBBiG objective in Eq. (3) and the within-bicluster variance objective of BicARE [41] directly correspond to the three block types used here.

Spatially constrained clustering methods—including SKATER, REDCAP, and max-p-regions—provide alternative approaches to regionalization [35]. They impose contiguity so that only adjacent NUTS 2 regions can be grouped, which is appropriate when the goal is to delineate contiguous planning areas. That constraint is unsuitable for the present design because it predetermines spatial clustering; therefore, it cannot distinguish a geographically concentrated similarity pattern from one dispersed across Europe—the distinction examined in RQ₃. Our two-stage design first identifies homogeneous leagues without spatial constraints and then tests their spatial concentration with Moran's I .

First, Kosztyán et al. [50] proposed three types of leagues. Type-A leagues contain indicators and, in the present application, NUTS 2 regions with comparatively high values. Therefore, League A represents the relatively developed group. League C is obtained by applying the biclustering procedure to the reversed binary matrix; it contains indicators and regions with comparatively low values; therefore, it represents the relatively underdeveloped group.

Both type-A and type-C leagues are identified by the iBBiG algorithm [44]. This algorithm assumes that the utilized dataset is binary; if this assumption is not met, the first step is to binarize the dataset on the basis of a given threshold (τ) [see, e.g., 50].

This algorithm balances the homogeneity (in this case, entropy) of the selected submatrix with the size of the selected block. Formally, the iBBiG algorithm maximizes the following objective where the binarized dataset at the threshold τ is denoted by B_{τ} :

$$\max \leftarrow score := (1 - H_B)^\alpha \begin{cases} \sum_i \sum_j [B]_{i,j} & , \text{ if } Me(B) > \tau \\ 0 & , \text{ if } Me(B) \leq \tau, \end{cases} \quad (3)$$

where $score$ is the score value of the submatrix (bicluster, league) $B \in B_{\tau}$. H_B is the entropy of the submatrix B , with a given threshold τ ; $Me(B)$ is the median of the bicluster B ; $\alpha \in [0, 1]$ is the exponent; and τ is the cutoff threshold. If τ , τ , or α is increased, we obtain a smaller but more homogeneous submatrix. Following previous studies [see, e.g., 44, 50], the balance exponent was set to $\alpha = 0.3$, and the cutoff threshold was set to $\tau = 0.5$. The binarization threshold τ was varied over $\{0.5, 0.6, 0.7, 0.8, 0.9\}$.

For every value of τ , iBBiG was run with one requested module (`nModules=1`), a population size of 100, a stagnation limit of 50, a selection pressure of 1.2, a maximum selection pressure of 15, and a success ratio of 0.8. The mutation probability was 0.30 for League A and 0.08 for League C. League A was initialized from the binarized normalized matrix, whereas League C was initialized from the binarized reverse-normalized matrix. The algorithm generates its initial genetic algorithm population internally.

The historical notebooks did not record a random-number seed. Consequently, the original stochastic random-number stream and the exact sequence of fitted biclusters cannot be reproduced retrospectively. The archived membership files instead preserve the exact League-A, League-B, and League-C solutions reported in the manuscript at every threshold. These archived memberships can be verified deterministically against the reported sizes, overlaps, descriptive statistics, and relevance coefficients; that verification requires no random seed. Separately, the reproduction archive provides a new stochastic verification run based on Master Seed 20,260,814. Each algorithm–threshold–league fit and each bootstrap task receives a distinctly named seed recorded in `seed_manifest.csv`, so the result does not depend on the execution order. This new run tests whether the released pipeline is executable and repeatable; it is not presented as a reconstruction of the unknown historical run, and its memberships may differ from the archived solution.

To support the identification of patterns, another biclustering method is applied, BicARE, which minimizes the variance within each bicluster. Notably, in this case, there is no need to binarize or normalize the original data. Kosztyán et al. [50] define this type of bicluster as a type-B league:

$$\min \leftarrow score := \frac{VAR(\mathcal{X})}{N \cdot M}, N \geq N_{\min}, M \geq M_{\min}, \quad (4)$$

where $\mathcal{X} \subseteq X$ is a submatrix, i.e., a bicluster (league), of the original completed analysis matrix X . $N \cdot M$ is the size of the bicluster, where N is the number of rows (i.e., NUTS 2 regions), and M is the number of columns (i.e., indicators). The minimum numbers of rows ($N_{\min}$) and columns ($M_{\min}$) must be specified.

The BicARE FLOC implementation was run on the complete, non-normalized matrix with one requested bicluster (`k=1`), `pGene=1`, `pSample=1`, residue threshold $r = 10^{-16}$, 10,000 iterations, and no user-supplied initial row or column blocks (`blockGene=NULL`, `blockSample=NULL`). Thus, all rows and columns were eligible at initialization. The minimum dimensions ($N_{\min}, M_{\min}$) were (110, 34), (100, 30), (90, 26), (80, 22), and (70, 18) for $\tau = 0.5, 0.6, 0.7, 0.8$, and 0.9 , respectively.

The biclustering methods balance block size against within-block heterogeneity, thereby identifying common properties within each league.

Row and column effects were evaluated with the `computeObservedFstat` procedure used in the archived R notebooks. It compares the observed within-bicluster row and column effects through F statistics. A bicluster was classified as significant only when both its row and column effects were significant. These tests summarize the fitted biclusters; they do not render the 22,720 cells as independent observations, and their p values are interpreted as exploratory diagnostics.

Identifying leagues through biclustering is a heuristic process. Therefore, in addition to conducting significance tests, we evaluated bicluster stability separately for rows and columns with `diagnoseColRow` (`number=1`, `nResamplings=100`, `replace=TRUE`). Thus, each reported stability result is based on exactly 100 bootstrap resamples with replacement. A bicluster was classified as stable only when both the row and column bootstrap tests were significant. With 100 resamples, a displayed empirical value of zero means that no bootstrap replicate was at least as extreme; therefore, it is reported as $p < 0.01$, not as $p = 0$.

League relevance was measured by rank preservation, not by partial correlation; no control variables were partialled out. With respect to the region coefficient, s_j^r is the equally weighted mean of the normalized indicators selected for a league, and s_j^a is the

Table 4
Significance, stability, relevance, and spatial autocorrelation statistics of the leagues.

Conditions	Statistics	League A		League B		League C	
		NUTS 2	I4.0 + ind.	NUTS 2	I4.0 + ind.	NUTS 2	I4.0 + ind.
tr = 0.5 N = 110 M = 34	Size	284	43	110	34	316	28
	Significance (p)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
	Stability (p)	< 0.01	< 0.01	< 0.01	< 0.01	0.3366	< 0.01
	Relevance (Spearman's ρ)	0.9004	0.9968	0.8000	0.9936	0.7648	0.9995
	Relevance (p value)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
	Moran's I	0.4885	0.5512	0.1076	0.5775	−0.0007 [†]	0.4297
tr = 0.6 N = 100 M = 30	Size	254	41	100	30	317	24
	Significance (p)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
	Stability (p)	< 0.01	< 0.01	< 0.01	< 0.01	0.4059	< 0.01
	Relevance (Spearman's ρ)	0.8937	0.9960	0.7175	0.9924	0.7539	0.9991
	Relevance (p value)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
	Moran's I	0.5707	0.5303	0.0932	0.5447	−0.0018 [†]	0.3699
tr = 0.7 N = 90 M = 26	Size	211	38	90	26	320	17
	Significance (p)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
	Stability (p)	< 0.01	< 0.01	< 0.01	< 0.01	0.4257	< 0.01
	Relevance (Spearman's ρ)	0.8184	0.9718	0.3267	0.9952	0.7030	1.0000
	Relevance (p value)	< 0.001	< 0.001	0.0017	< 0.001	< 0.001	< 0.001
	Moran's I	0.4905	0.5693	0.1239	0.4753	NA	0.2246
tr = 0.8 N = 80 M = 22	Size	171	33	80	22	316	14
	Significance (p)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
	Stability (p)	< 0.01	< 0.01	< 0.01	< 0.01	0.4851	< 0.01
	Relevance (Spearman's ρ)	0.7284	0.9525	0.2918	0.9921	0.5623	1.0000
	Relevance (p value)	< 0.001	< 0.001	0.0086	< 0.001	< 0.001	< 0.001
	Moran's I	0.3995	0.5571	0.1243	0.3863	−0.0004 [†]	0.1492
tr = 0.9 N = 70 M = 18	Size	302	6	70	18	307	12
	Significance (p)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
	Stability (p)	< 0.01	< 0.01	< 0.01	< 0.01	0.4455	< 0.01
	Relevance (Spearman's ρ)	−0.0680	1.0000	0.1013	0.9938	0.5346	0.9930
	Relevance (p value)	0.2384	< 0.001	0.4039	< 0.001	< 0.001	< 0.001
	Moran's I	0.0194 [†]	0.0215	0.1172	0.2868	0.0153 [†]	0.1174

Note: biclustering significance tests are nominal and unadjusted for multiplicity. Stability values are based on 100 bootstrap resamples; $p < 0.01$ denotes no exceedances among those resamples. Relevance is Spearman's rank correlation between the all-indicator ranking and the ranking restricted to the selected dimension; the corresponding two-sided, large-sample p values were computed separately without multiplicity adjustment. Moran's statistics were rerun with the fully specified row-standardized spatial weights. A dagger (†) denotes a statistic that was not significant at the 0.05 level after Benjamini–Hochberg correction within its 15-test spatial family. NA denotes the constant League C membership vector at tr = 0.7, for which Moran's I is undefined.

equally weighted mean of all 71 normalized indicators, both of which are evaluated for the regions selected for that league. With respect to the indicator coefficient, the analogous scores are equally weighted means across the selected regions and across all 320 regions, respectively, evaluated for the indicators selected for the league. Let $R(\cdot)$ denote ranks, with average ranks assigned to ties. In both cases, the relevance coefficient is the ordinary Spearman correlation

$$\rho = \text{cor}\left[R(s^I), R(s^{\text{all}})\right], \quad (5)$$

computed in R with `cor.test(..., method="spearman", exact=FALSE)`. Table 4 reports the Spearman coefficient ρ and its two-sided large-sample p value in separate rows. Values $|\rho| > 0.5$ were treated as high-rank preservation. A high region coefficient means that the selected indicators preserve the ordering of the selected regions; a high indicator coefficient means that the selected regions preserve the ordering of the selected indicators. The previous term “partial rank correlation” was misleading and has been removed.

The reported significance, stability, and relevance tests are nominal and were not adjusted for multiplicity. Because the biclusters and threshold comparisons are exploratory and statistically dependent, we emphasize effect sizes, stability across thresholds, and the substantive pattern rather than treating isolated nominal p values as confirmatory evidence.

The biclustering methods can identify multiple, potentially overlapping leagues and order them by their objective scores. The first bicluster in an ordered solution is termed the *dominant bicluster* because it contains the most cells. A league is treated as well supported when it is significant, stable, and relevant. The analysis reports only the dominant bicluster for each league type.

3.3. Spatial autocorrelation and Moran's I

Spatial autocorrelation measures whether values observed in nearby regions are more similar or dissimilar than expected under spatial randomness. Positive spatial autocorrelation occurs when neighboring regions tend to have similar values, whereas negative spatial autocorrelation occurs when high and low values tend to be adjacent. Values consistent with spatial randomness show no

detectable autocorrelation. This framework reflects Tobler's first law of geography: nearby phenomena are generally related more strongly than distant phenomena are [83].

Spatial-autocorrelation statistics differ primarily in how they define neighborhood relationships and summarize spatial comovement. We use the global Moran's I [67]:

$$I = \frac{N}{S_0} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} z_i z_j}{\sum_{i=1}^N z_i^2}, \quad z_i = x_i - \bar{x}, \quad S_0 = \sum_{i=1}^N \sum_{j=1}^N w_{ij}, \quad (6)$$

where w_{ij} is the element (i, j) of the spatially weighted matrix, and N is the number of regions. Under the randomization null, the expected global value is $E[I] = -1/(N-1)$, which is not exactly zero. Values above this expectation indicate positive spatial association and values below it indicate negative spatial association; whether the departure is statistically significant must be determined by the stated reference distribution rather than by the sign alone. Global Moran's I summarizes the complete regional pattern with one statistic.

Whereas global Moran's I summarizes the complete regional pattern, the local indicator of spatial association (Local Indicator of Spatial Association (LISA)) provides a statistic for each region and can identify local clusters and spatial outliers [2]:

$$I_i = \frac{z_i}{m_2} \sum_{j=1}^N w_{ij} z_j, \quad m_2 = \frac{1}{N} \sum_{i=1}^N z_i^2, \quad (7)$$

Here, I_i is the local statistic for region i , and m_2 is the second moment of the centered indicator. A positive local statistic is consistent with a high-high or low-low cluster, whereas a negative statistic is consistent with a high-low or low-high spatial outlier; cluster type is determined jointly from z_i and its spatial lag. Local significance must likewise be assessed against an explicit reference distribution. Thus, the global and local statistics answer different questions: the former evaluates overall spatial association, whereas the latter identifies significant clusters and outliers.

Spatial weights were reconstructed, and all spatial results were rerun from the released data. We used the Eurostat/GISCO NUTS 2016 1:60-million polygon file in EPSG:4326. For each retained NUTS 2 polygon, we calculated the area-weighted polygon centroid in the source coordinate system and then calculated the great-circle angular distances between the centroids. The initial symmetric binary matrix was defined as $d_{ij} = 1$ when $0 < D_{ij} \leq 5^\circ$ and $d_{ij} = 0$ otherwise, with $d_{ii} = 0$ set explicitly. Five regions had no neighbor inside this band: FRY1, FRY3, FRY4, IS00, and PT20. We handled these islands deterministically by adding a symmetric edge to the nearest retained centroid: FRY1–FRY3 (14.684°), FRY4–CY00 (60.047°), IS00–UKM6 (9.979°), and PT20–PT30 (11.726°). The corrected binary matrix was then row standardized, $w_{ij} = d_{ij} / \sum_j d_{ij}$. Consequently, every row sums to one, $S_0 = 320$, and the final minimum, median, and maximum neighbor counts are 1, 45, and 104, respectively. Unusually long fallback connections are reported explicitly rather than being concealed as ordinary 5-degree neighbors; sensitivity to alternative treatments of geographically remote territories remains a limitation.

Global significance was assessed with 999 unconditional random permutations of the regional values and a two-sided pseudo- p value centered on $E[I] = -1/(N-1)$; therefore, the smallest attainable value is 0.001. We used the fixed seed 20,260,814. Benjamini–Hochberg false-discovery-rate correction was applied separately to the 71 indicator tests, the 15 league-region membership tests, and the 15 league-indicator composite tests. With respect to each mapped indicator, local significance was assessed with 999 conditional permutations: z_i was held fixed, and k_i values were sampled without replacement from the other $N-1$ regions. The resulting 320 local p values were Benjamini–Hochberg adjusted within that indicator, and a region was assigned a LISA cluster/outlier category only when $q < 0.05$. The released spatial script, weight matrix, permutation results, diagnostics, and session information reproduce every value reported below.

4. Results

4.1. Set of leagues

Fig. 2 presents the seriated normalized matrix, with higher values in green and lower values in purple, together with selected League-A and League-C solutions. League A has moderate to high regional Moran statistics under the first four thresholds, whereas League-C membership is spatially diffuse under every nonconstant setting (Table 4). Thus, the identified strengths are more geographically concentrated than the low-performance league memberships, directly informing O_3 and RQ_1 .

The vertical pattern of the seriated normalized matrix indicates that the indicators with the highest (M01–M06) and lowest (M07–M12) normalized means are both related to media coverage and therefore reflect different dimensions of industrial content. Employment indicators E19, E23, E24, and E26 form the next group, with high means and high Moran statistics. In contrast, ER2 (incoming Erasmus students), S12 (I4.0-related studies), S13 (R&D personnel and researchers in the business-enterprise sector), and S14 (research institutes) form the next-lowest group and have moderate or low Moran statistics (Table A.2). These groups characterize Leagues A and C at the strictest threshold ($\tau = 0.9$). No indicator is assigned simultaneously to League A and League C at any tested threshold; this mutual exclusivity follows from the construction of League C from the reversed normalized matrix. It is distinct from the narrower notion of “common” challenges used in relation to RQ_3 , which refers to weaknesses that are not geographically concentrated. In contrast, at the NUTS 2-region level, Leagues A and C continue to overlap substantially (Section 4.2); several regions belong to both leagues, indicating strong performance on some indicators alongside remaining development needs on others (O_1).

Fig. 3 shows League B under different conditions.

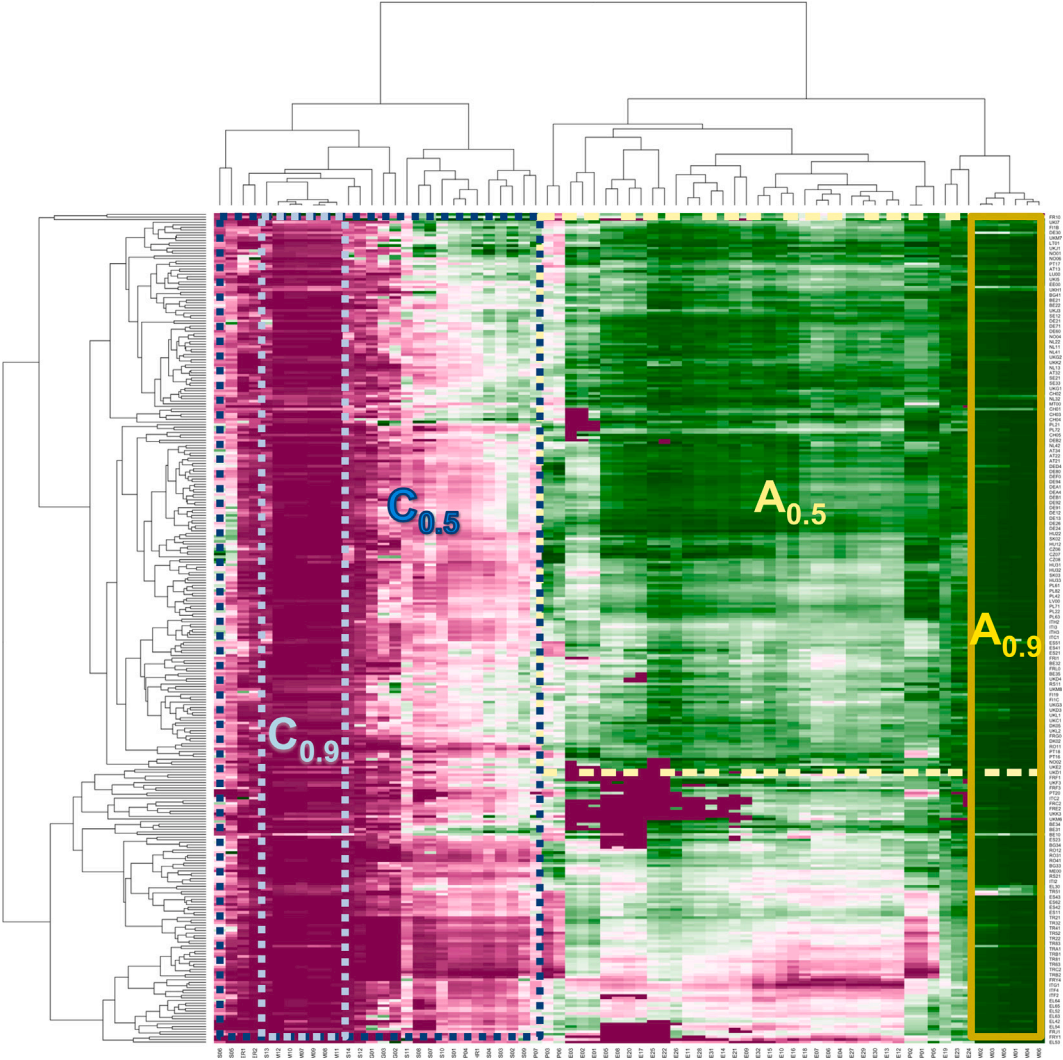


Fig. 2. Seriation and biclustering results under different thresholds. [Source: figure designed by the authors and generated in R.]
Notes: A_{50} , C_{50} , A_{90} , and C_{90} give the dominant leagues for thresholds $\tau_r = 0.5$ and 0.9 , respectively.

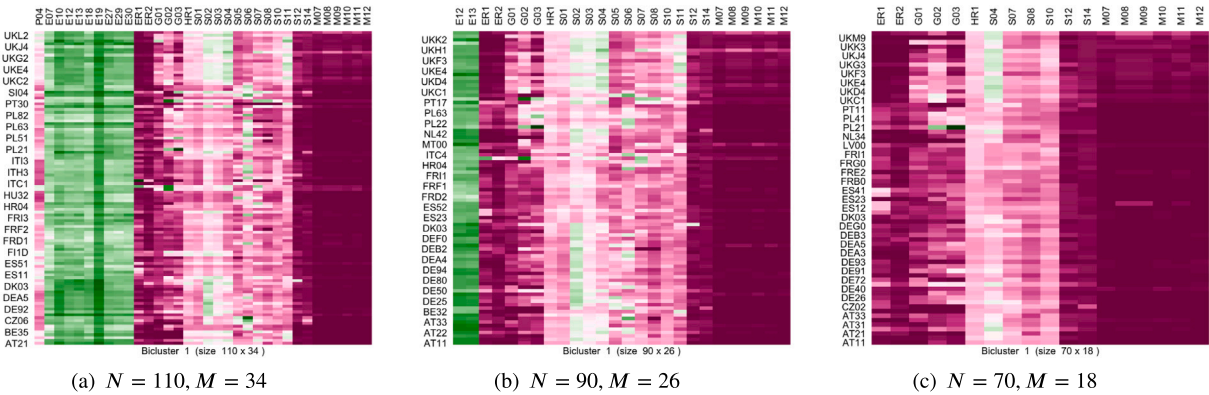


Fig. 3. League B under different threshold settings. [Source: figure designed by the authors and generated in R.]

The League B regional membership patterns have small but statistically detectable Moran statistics ($I = 0.093$ – 0.124 after correction; see Table 4). Thus, emerging regions are geographically clustered only weakly rather than being wholly independent of location (see O_3). Tightening the conditions also decreases the number of included indicators, which means that the identified weaknesses are more homogeneous than the strengths are (see RQ_1).

Table 3 shows the cell statistics for the included (within a bicluster) and excluded (i.e., remaining) cells. The overall mean is 0.5260, and the overall standard deviation is 0.3407.

As the threshold is tightened, the within-league mean increases for League A (0.7636 to 0.9788) and decreases for League C (0.2081 to 0.0409), so the separation increases between the high- and low-performing blocks. Their within-block standard deviations also decrease. With respect to League B, both the within-block mean and the standard deviation decrease as the minimum row and column constraints are tightened. These patterns confirm that stricter settings select smaller, more homogeneous blocks.

Table 4 reports the bicluster significance, stability, relevance, and spatial autocorrelation for every league type and threshold. The statistics are reported separately for rows (NUTS 2 regions) and columns (indicators).

League C reflects relatively weak indicators across the NUTS 2 regions. Tightening the threshold has almost no effect on the set of regions; it primarily changes the indicator set. Its binary regional membership pattern is not significant after false-discovery-rate correction at $tr = 0.5, 0.6, 0.8$, or 0.9 ; at $tr = 0.7$, all 320 regions are members, so the statistic is undefined. This diffuse membership pattern supports common interventions for shared weaknesses such as the number of industry-related studies (S12), total R&D researchers (S13), number of research institutes (S14), and their reflection in the media (M07–M12). The pattern does not imply that every underlying indicator is individually aspatial. The strong relevance coefficients indicate that rankings based only on the selected indicators are similar to rankings based on all 71 indicators (see O_1 and O_3).

In contrast, the developed league (League A) has strongly clustered region membership at $tr = 0.5$ – 0.8 ($I = 0.399$ – 0.571 , adjusted $q < 0.05$); the $tr = 0.9$ membership pattern is not significant. Except at $tr = 0.9$, League A includes more indicators but fewer NUTS 2 regions than League C does. The results indicate geographically concentrated strengths under the first four settings but also show that developed regions can still contain indicators that require improvement. The last setting should be interpreted cautiously because its regional relevance is negative and its spatial membership pattern is not significant.

Leagues of Type B include emerging NUTS 2 regions. All the biclusters are significant and stable, but only the first two meet the stated relevance criterion for both dimensions. In every setting, the selected numbers of rows and columns equal the minimum constraints, indicating that no larger equally homogeneous block was found. The small positive Moran statistics for League-B membership show weak spatial clustering: these patterns are not entirely aspatial, but spatial location explains much less of them than it does for League A under the first four thresholds.

4.2. Overlap analysis

We analyze the overlap among the archived biclusters for both NUTS 2 regions and indicators. Fig. 4 shows the named overlap categories at $tr = 0.5$. iBBiG identifies League A from high values in the normalized matrix and League C from high values in the reverse-normalized matrix, corresponding to low values in the original orientation. BicARE identifies the intermediate, low-variance League B. The associated significance results from the F -statistic diagnostics are reported in Table 4.

Regions may belong to more than one league because the same region can perform strongly on indicators selected for League A and poorly on indicators selected for League C. Indicators are never assigned jointly to Leagues A and C at any tested threshold; only A–B and B–C indicator overlaps occur (Table A.2). Therefore, the bicluster overlaps in Fig. 4 describe distinct regional and indicator relationships and clarify the similarities and differences addressed by O_1 .

League A +: The non-overlapping part of League A includes four capital regions and 34 indicators, most of which are employment indicators.

League B: The non-overlapping part of League B is empty.

League C: The non-overlapping part of League C includes 36 regions, primarily from North Macedonia, Turkey, and Greece, and three indicators, primarily from the science category.

League AB: The overlap between Leagues A and B contains the remaining employment indicators.

Leagues AC and ABC: The A–C and A–B–C regional overlaps contain 280 of the 320 regions. League AC contains no indicators and 170 regions, primarily from Germany, Austria, Norway, Portugal, and Bulgaria. League ABC contains 110 regions, primarily from the Netherlands, Romania, Sweden, Switzerland, Cyprus, Iceland, Luxembourg, Montenegro, Malta, Croatia, Slovenia, Estonia, and Latvia.

Across all five tested thresholds, no indicator is assigned simultaneously to both League A and League C (Table A.2). By contrast, the share of NUTS 2 regions belonging to both leagues (categories AC and ABC) is substantial and generally declines as the threshold is tightened: 87.5% ($tr = 0.5$) $\rightarrow$ 78.4% ($tr = 0.6$) $\rightarrow$ 65.9% ($tr = 0.7$) $\rightarrow$ 52.8% ($tr = 0.8$). At $tr = 0.9$, the share increases to 92.5%; however, League A's regional relevance then becomes negative (Table 4, $\rho = -0.068$; $p = 0.2384$), so this setting should be interpreted with caution rather than as a substantive reversal. The number of indicators shared between adjacent leagues decreases monotonically as the threshold is tightened: the AB counts are 9, 6, 2, 0, and 0, and the BC counts are 25, 22, 16, 13, and 11 for $tr = 0.5$ through $tr = 0.9$. This indicator–region distinction resolves the apparent tension: League-A and League-C indicator memberships are mutually exclusive by construction, whereas their region memberships are not because a region can perform well on some indicators and poorly on others.

League BC: The overlap between Leagues B and C contains the second-largest indicator set, primarily the science, graduate, and Erasmus indicators.

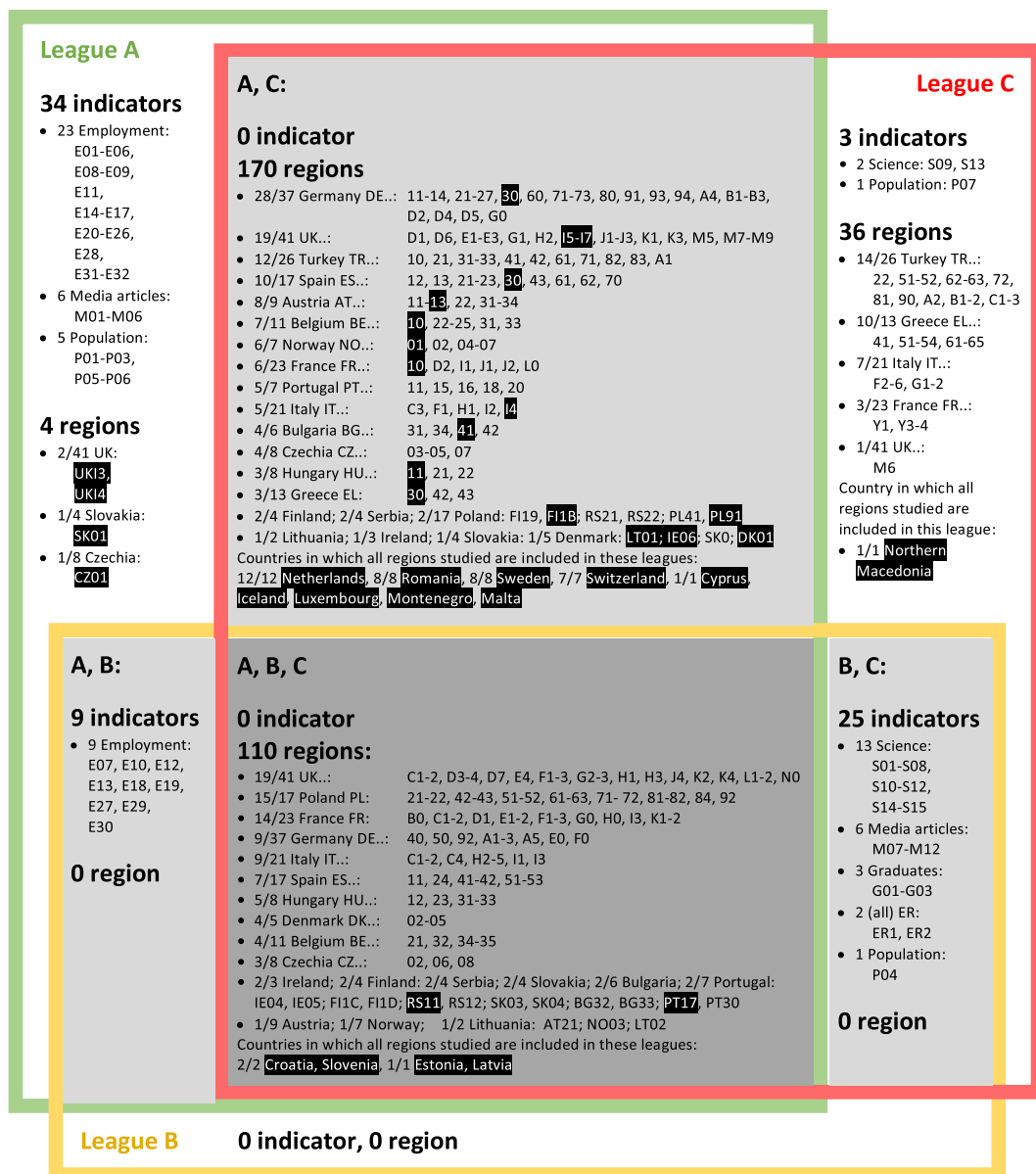


Fig. 4. Venn diagram of the leagues identified from 71 indicators and 320 NUTS 2 regions ($tr = 0.5$, $N = 110$, $M = 34$). Regions containing a capital are shown with a black background. [Source: Figure designed by the authors.] (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 5 summarizes the indicators that may require improvement for regions to move toward higher-performing leagues and identifies indicators shared across overlapping biclusters. The comparison highlights strengths and weaknesses among regions in League A+ and League C- and identifies potential intervention points relevant to O_2 .

4.3. Influence of thresholds

Figs. 6 and 3 show how indicator membership changes across thresholds. League A+ contains the largest indicator set at $tr = 0.5$, but more indicators become unmatched (code 000) as the threshold increases. League AB becomes empty, whereas League C- remains relatively stable at higher thresholds.

Fig. 7 shows the alluvial plot for the NUTS 2 regions.

Fig. 7 shows how the regions are distributed across the leagues at different thresholds. We find that most regions are in leagues AC and ABC, indicating mixed development levels. Only four capital regions are in League A, with greater development than their surrounding regions. League C has more regions at lower thresholds, and its membership increases as the threshold increases, except

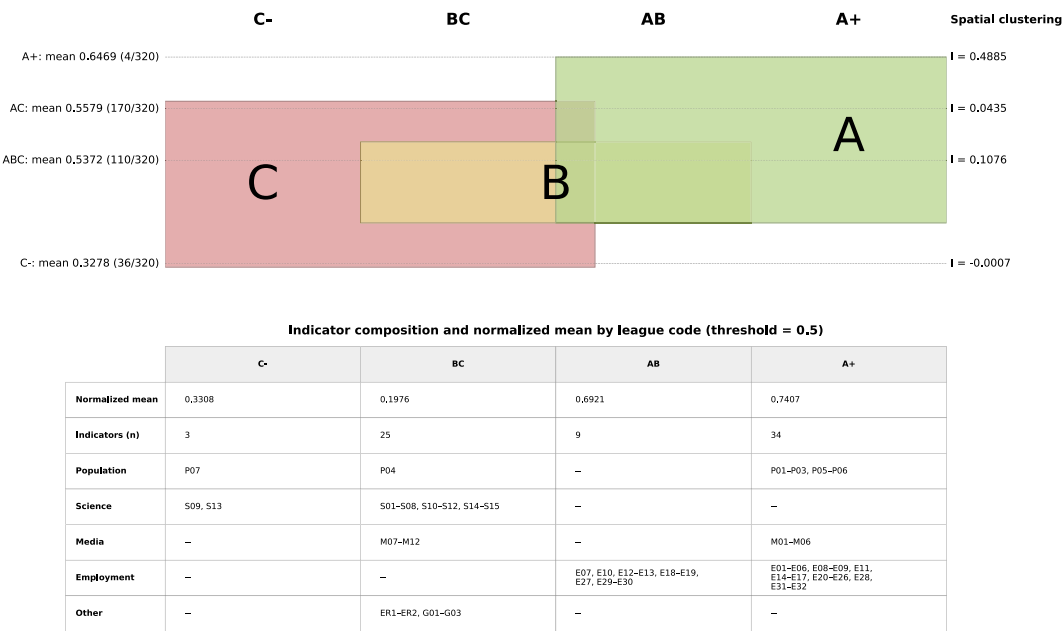


Fig. 5. Development opportunities among leagues ($\alpha = 0.5$, $N = 110$, and $M = 34$). The spatial clustering values were recomputed with the fully specified released weight matrix; the A–C value describes the exact AC overlap category. [Source: Figure designed by the authors and generated with the released reproduction code.]

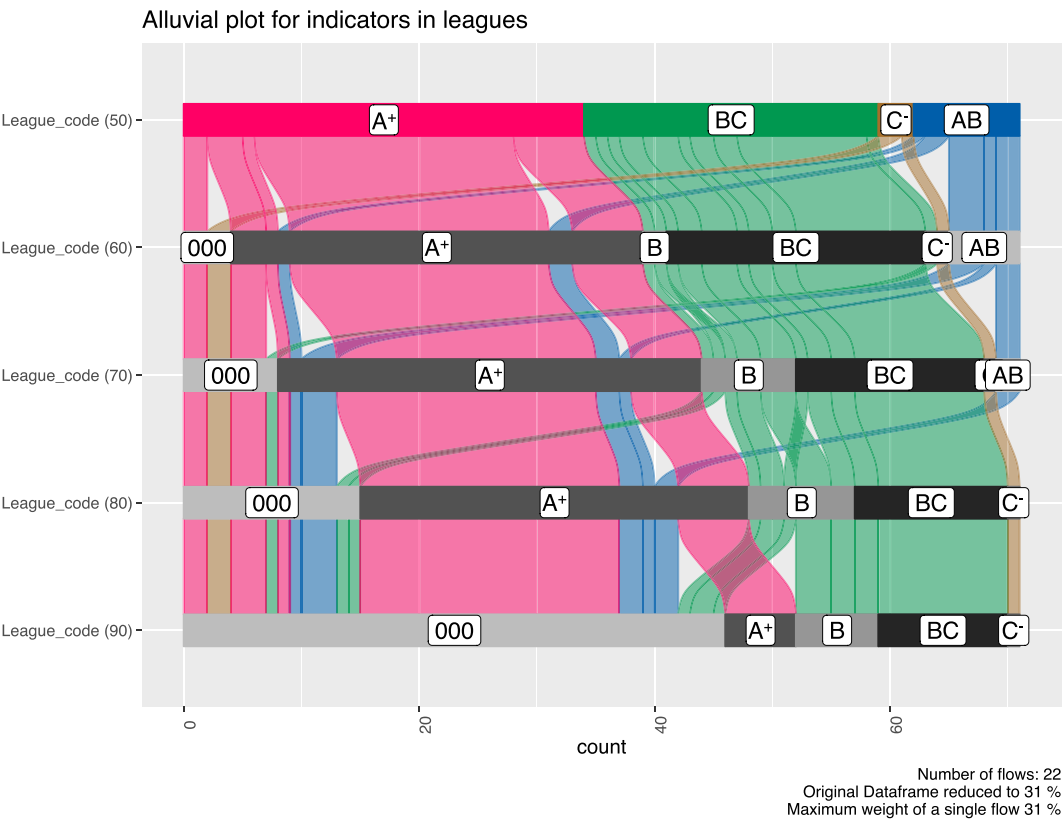


Fig. 6. Alluvial plot of indicator league membership across thresholds. Code 000 denotes indicators that do not belong to any league. [Source: Figure designed by the authors and generated in R.]

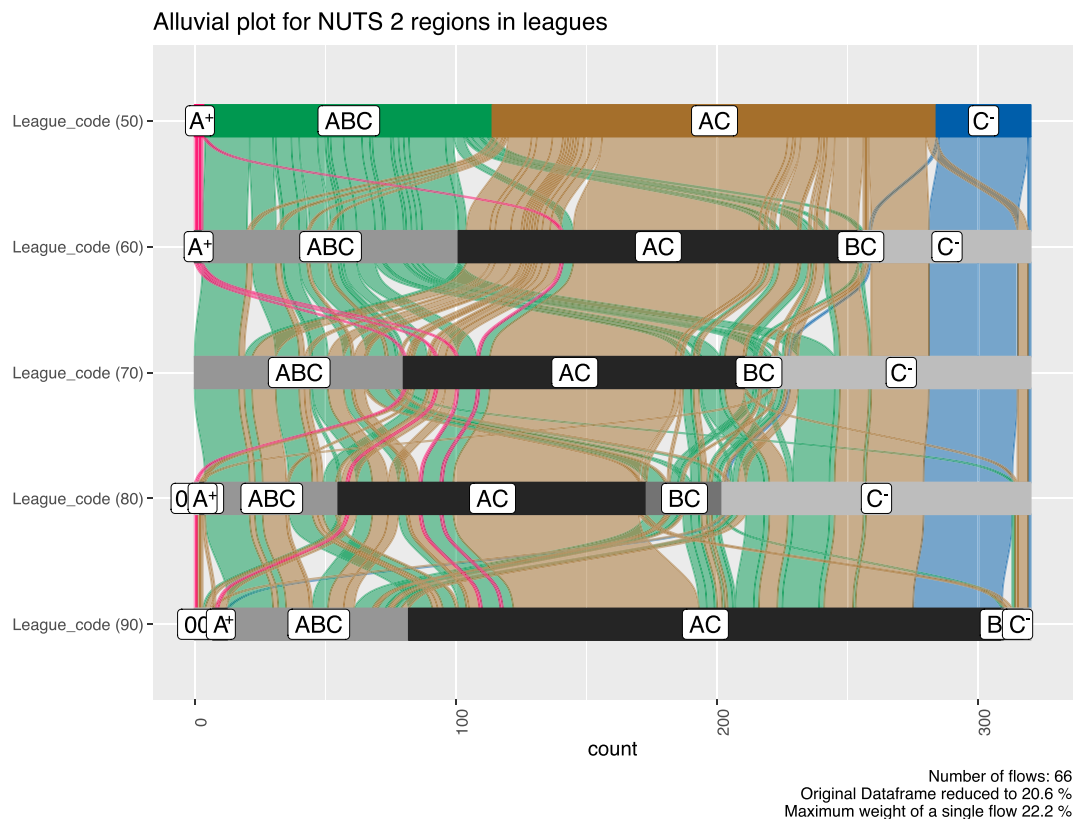


Fig. 7. Alluvial plot of regional league membership across thresholds. [Source: Figure designed by the authors and generated in R.]

in the last setting. Most of the regions in League C are in Turkey, Italy, Greece, and northern England. The supplementary settings for the other thresholds are shown in Fig. A.1.

Fig. 8 shows a map of league membership.

4.4. Spatial determination

Table 5 lists the five indicators with the highest global Moran's I statistics.

The five highest global Moran statistics are associated with population education (P01, P02, P03, and P05) and employment (E15), and all remain significant after false-discovery-rate correction. P01 was reversed after normalization to express higher educational readiness; this reversal leaves its global Moran statistic unchanged but makes its high–high and low–low map categories substantively consistent with those of the other readiness indicators. P01 and P02 are exact complements in the source definition; therefore, they become identical after the P01 reversal, explaining their tied Moran statistics.

The spatial patterns of two representative indicators with high Moran statistics (P01 and P05) and the two indicators with the lowest statistics (M01 and M04) are shown in Fig. 9. P02 was not mapped because its post-reversal values precisely duplicate those of P01. The gray regions do not satisfy the within-indicator Benjamini–Hochberg criterion ($q < 0.05$); the dark red (blue) regions are significant high–high (low–low) clusters, and light red (blue) regions are significant high–low (low–high) spatial outliers.

The high-Moran maps show a pronounced core–periphery pattern, whereas the lowest-Moran maps contain few or no local clusters after multiplicity correction. Across all 71 indicators, 69 global statistics remain significant at $q < 0.05$; the two nonsignificant results are M01 and M04. Thus, the corrected analysis supports spatial dependence for many indicators, but not for all of them, and avoids treating a positive statistic alone as evidence of significance.

5. Discussion

The media indicators (M01–M12) include both the highest and lowest normalized means and capture the amount and tone of regional industrial coverage. Media coverage can amplify favorable and unfavorable perceptions of regional economies [see, e.g., 30,90], influence investor confidence and public perceptions of I4.0 readiness [24], and draw attention to industries, skills, and labor-market trends [7,58]. Negative coverage may likewise foreground weaknesses in education, employability, and manufacturing competitiveness [65,69]. Therefore, the media indicators should be interpreted as measures of industrial discourse and visibility, not as direct measures of productive capacity.

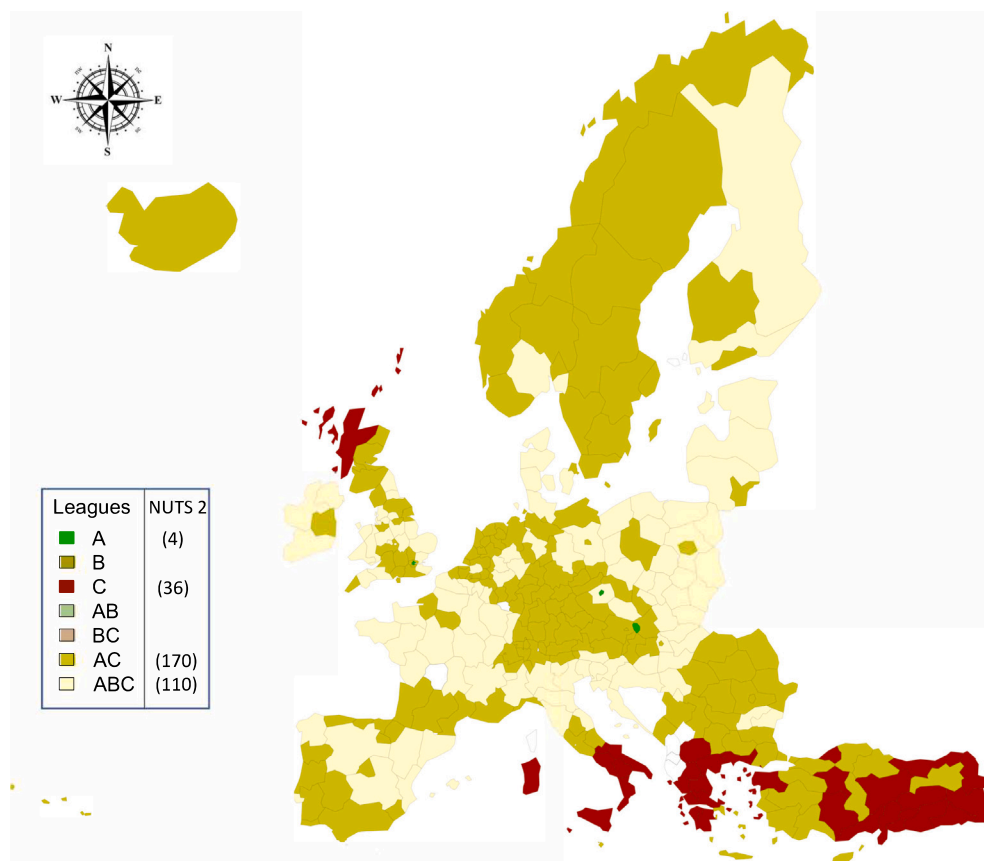


Fig. 8. Biclustered NUTS 2 regions and regional Venn diagram at $tr = 0.5$. [Source: figure designed by the authors and generated in R.]

Table 5

Indicators with the five highest global Moran statistics and their membership league. An asterisk indicates Benjamini–Hochberg-adjusted $q < 0.05$ within the 71-indicator family.

Rank	Indicators	Moran Index	Normalized mean value	League code (threshold setting)				
				(50)	(60)	(70)	(80)	(90)
1	P01	0.7716*	0.7158	A+	A+	A+	A+	000
2	P02	0.7716*	0.7158	A+	A+	A+	A+	000
3	P05	0.7173*	0.7687	A+	A+	A+	A+	000
4	P03	0.7052*	0.5241	A+	000	000	000	000
5	E15	0.6638*	0.7342	A+	A+	A+	A+	000

In relation to O_1 , the three league types reveal distinct patterns of regional competitiveness (Fig. 2; Tables 3 and 4). League A combines high employment and education values with strongly clustered region membership under the first four thresholds, which is consistent with the concentration of advantages in core European regions [1]. League C combines low values with diffuse region membership, particularly for science- and research-related indicators. These findings are consistent with documented disparities in I4.0 readiness across EU countries [25]. The alluvial plots further show that indicator assignments are more robust to increasing thresholds than regional assignments are. The spatial analysis qualifies the resulting interpretation: League-B membership is weakly but significantly clustered, whereas League-C membership is nonsignificant under the four nonconstant settings. At the indicator level, however, 69 of 71 global Moran tests remain significant after correction. Therefore, a diffuse league-membership pattern must not be interpreted as evidence that every component indicator is aspatial. This distinction supports differentiated place-based and broader policy responses [31].

Our analysis reveals a qualified dichotomy in the developmental landscape of European NUTS 2 regions, directly addressing our third research question (RQ₃). Most science, graduate, and Erasmus indicators are associated with League BC, whose mean values are low; however, League B membership has a small but significant positive spatial autocorrelation, and many component indicators are individually significant. The stronger evidence for a common challenge comes from League C: its regional membership statistic is nonsignificant under every nonconstant threshold, even though some of its constituent indicators retain spatial structure. This distinction between the geography of a league and the geography of an individual indicator prevents an overly broad “aspatial science”

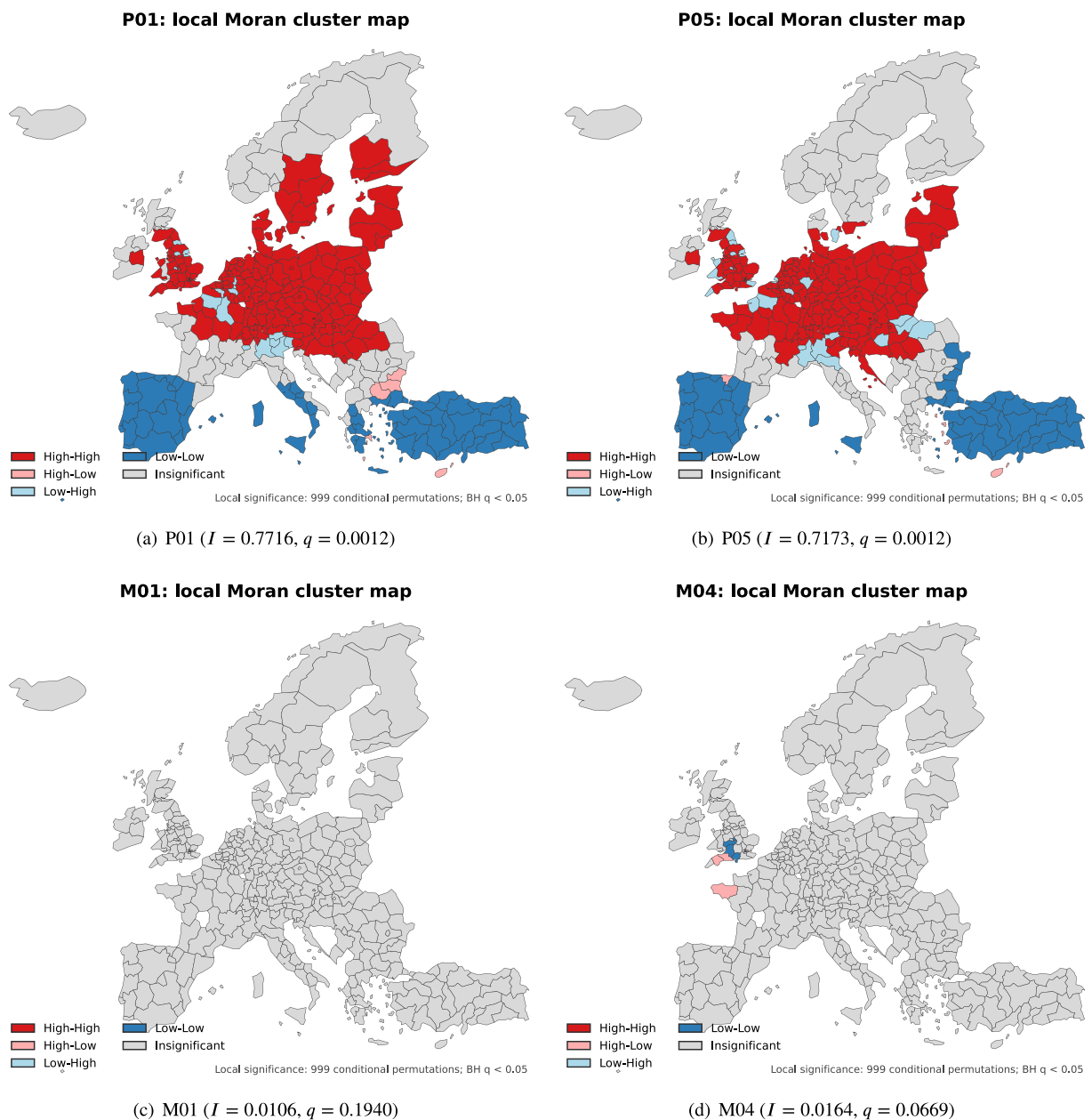


Fig. 9. Local Moran cluster maps for representative indicators with high global autocorrelation (a and b) and for the two indicators with the lowest global statistics (c and d). Local categories are displayed only for within-indicator Benjamini–Hochberg-adjusted $q < 0.05$. [Source: Figure designed by the authors and generated with the released reproduction code.]

conclusion. Conversely, most employment and population indicators are associated with Leagues AB and A+, and several of them have the highest global Moran statistics. These findings support a core–periphery pattern consistent with the research on regional inequality in Europe [49,74]. Therefore, the practical implication is conditional: regionally concentrated programs are supported where both league membership and the relevant indicator are strongly spatial, whereas diffuse league membership supports broader framework programs without ruling out targeted action for a particular indicator.

The spatial visualization of our derived industrial leagues (Fig. 5) and the analysis of their composition under varying biclustering conditions (Figs. 6–8 and A.1) offer a granular roadmap for policy intervention, directly addressing our second research objective (O₂). Science and education indicators consistently emerge as important development needs, predominantly populating League C and its derivatives. These findings align with Balland and Boschma [8], who noted that most European regions exhibit weak potential in I4.0 technologies. The nonsignificant spatial pattern of League C membership supports a broad policy component, whereas the significant Moran statistics of many individual science and education indicators show that this broad component may need to be supplemented

with indicator-specific place-based measures. This two-level interpretation, rather than a blanket aspatial classification, provides the empirical basis for separating regionally specific from sector-wide action plans (O_3).

Geographically, our analysis confirms a significant development lag in regions across southern Europe and Turkey, which are disproportionately represented in the low-performing leagues. This observation reinforces the well-documented core–periphery divide in European economic development [23,49]. The critical implication for policymakers is the need for a dual-pronged strategy: while geographically concentrated development programs are necessary for regions in southern Europe, they must be complemented by broader, nonspatial initiatives that increase the entire continent's scientific and educational capacity to foster a more globally competitive and equitable industrial landscape.

6. Conclusion

This research addresses the need for data-driven regional development planning in Europe by introducing an analytical framework that combines aspatial biclustering with spatial autocorrelation techniques. The primary objective of this research was to identify European industrial leagues—clusters of NUTS 2 regions with similar strengths and weaknesses in I4.0+ readiness—and to determine the development priorities associated with those leagues. Specifically, we identify common regional development challenges and opportunities (O_1), determine which indicators require improvement for regional advancement (O_2), and help decision-makers distinguish between regionally specific and common development priorities (O_3).

The empirical analysis of the 71 retained I4.0+ readiness indicators across 320 European NUTS 2 regions yields several findings relevant to the research questions. With respect to RQ_1 , seven industrial-league categories have distinct indicator profiles. League A+ contains the highest-performing regions (Inner London–West, Inner London–East, Bratislava, and Prague), primarily in terms of employment indicators. League C- contains lower-performing regions—predominantly in Turkey, Greece, and North Macedonia—with weaknesses concentrated in population and science indicators (S09, S13, and P07). These results are consistent with a core–periphery divide and identify region–indicator combinations relevant to targeted intervention (C_1).

With respect to RQ_2 , biclustering jointly identifies homogeneous sets of regions and indicators, whereas spatial autocorrelation reveals the geographical dependence relevant to the scale of intervention. The combined evidence provides a structured basis for the regional development priorities summarized in C_2 .

The distinction between regionally specific and common development needs (RQ_3) emerges at two levels. Employment and population include the indicators with the strongest global spatial autocorrelation, whereas League-A membership is strongly clustered under the first four thresholds. In contrast, League-C membership is spatially diffuse under every nonconstant setting, although many individual science and education indicators still have significant Moran statistics. These findings support a dual-pronged strategy that combines spatially targeted interventions with sector-wide initiatives and selects the scale of action from both league-membership and indicator-level evidence (C_3).

6.1. Implications for scholars

This research integrates spatial invariance techniques (seriation and biclustering) with a spatial autocorrelation analysis (Moran's I) in a single regional development framework. This combination addresses a methodological gap between approaches that identify homogeneous profiles without evaluating geography and spatial analyses that do not jointly select regions and indicators.

The industrial league concept contributes to the theoretical understanding of regional development by providing a systematic framework for categorizing regions on the basis of their development profiles rather than on that of geographical proximity or administrative boundaries alone. This approach aligns with contemporary theories of regional development that emphasize functional rather than territorial relationships, contributing to the growing literature on functional regionalism and place-based development strategies.

The qualified distinction between strong and weak spatial patterns provides theoretical insight into the mechanisms of regional development. Employment and population include several strongly clustered indicators; science and education indicators span a wider range of positive Moran statistics, whereas the low-performing League C membership pattern is diffuse. These multilevel results have implications for theoretical models that must distinguish the spatial concentration of a league from the spatial dependence of its constituent indicators.

Furthermore, this research contributes to the growing body of literature on the I4.0 readiness assessment by providing a comprehensive analytical framework that transcends simple indicator ranking by examining the complex interrelationships between multiple dimensions of technological readiness. The identification of indicator–region biclusters provides a more sophisticated understanding of how different aspects of I4.0 readiness manifest across European regions.

6.2. Methodological implications

Integrating spatial and aspatial techniques allows regional development analysis to consider profile similarity and spatial dependence separately. This separation addresses a limitation of approaches that use either type of evidence alone.

Robustness testing across different threshold levels demonstrates that the identified indicator patterns are more stable than regional memberships, suggesting that future research should prioritize analytical approaches focused on indicators rather than regions. These findings have important implications for the development of monitoring systems and evaluation frameworks that can provide consistent insights across different analytical specifications.

Biclustering identifies homogeneous groups of regions and indicators simultaneously, unlike conventional clustering approaches that treat these dimensions separately. Therefore, researchers can determine which indicators define each regional group and examine the multidimensional structure of development.

Spatial autocorrelation analyses provide a systematic framework for distinguishing between local and global development phenomena and offer researchers a robust empirical basis for determining the appropriate spatial scale for different types of interventions. This methodological contribution is particularly valuable for comparative regional studies, which must account for varying spatial dependencies across different development dimensions.

6.3. Implications for decision-makers

The findings provide actionable insights for policy-makers at multiple governance levels, from regional authorities to EU institutions. The identification of specific industrial leagues enables decision-makers to develop targeted intervention strategies that account for the unique development profiles of different regional clusters. Regions within the same league face similar development challenges and possess comparable strengths, which creates opportunities for interregional collaboration, knowledge sharing, and joint action plans.

The distinction between stronger and weaker spatial evidence offers guidance for resource allocation and policy design. With respect to indicators and league-membership patterns that exhibit a high degree of spatial autocorrelation with statistical significance, including many employment and population indicators as well as League A at the first four thresholds, decision-makers can prioritize place-based interventions that leverage proximity and regional clustering effects. These interventions might include regional innovation clusters, cross-border cooperation programs, and spatially concentrated infrastructure investments that can generate spillovers to neighboring regions.

Conversely, diffuse league membership, observed most clearly for League C, supports sector-wide initiatives addressing shared weaknesses across multiple regions. These interventions might include pan-European research and development programs, harmonized educational standards, and technology-transfer mechanisms. Because many science and education indicators also show significant individual spatial structures, such framework programs should be complemented by place-based components when the indicator-level evidence warrants them.

Taken together, the results imply a two-part decision rule. First, league membership identifies *which* regions and indicators require attention. Second, the corrected Moran evidence determines the appropriate spatial scale: high, significant indicator and membership statistics support place-based intervention; weak or nonsignificant membership statistics support a sector-wide component. When these two levels disagree—as can occur for League C and some of its science indicators—the appropriate response is mixed rather than exclusively regional or exclusively uniform.

This research provides specific guidance for EU cohesion policy by identifying which aspects of regional development benefit from concentrated geographical targeting and which require broader, more distributed interventions. This insight is particularly relevant for the design of structural funds and regional development programs that must balance efficiency considerations with equity objectives.

Regional development agencies can use the industrial league framework to benchmark their performance against comparable regions and identify specific areas requiring improvement. The methodology provides a systematic approach for regions to assess their current development status and develop evidence-based strategies for advancing to higher-performing leagues.

Taken together, the observed core-periphery pattern, the integration of spatial and aspatial evidence, and the resulting two-level decision rule contribute to both regional development research and policy analyses. The framework offers decision-makers a transparent way to connect region-indicator profiles with the geographical scale of potential interventions.

7. Limitations and future work

The framework has several limitations. First, the analysis provides a cross-sectional view of the European industrial landscape in 2018. The use of this year permitted data completeness and consistency to be maximized, and the original 101-indicator framework was screened against the 71 indicators used in the analysis. The resulting leagues nevertheless describe one historical period and cannot capture regional development trajectories. A longitudinal analysis is needed to determine whether league memberships remain stable or change following policy intervention.

Furthermore, 10,977 of the 34,340 values in the initial 2018 table (31.97%) were unavailable before processing. Temporally ordered substitution filled 6026 of these values, and after region and indicator screening, 740 of 22,720 cells (3.26%) remained unavailable and were encoded as zero in the archived analysis. Temporal substitution assumes that values are sufficiently stable across adjacent years, whereas residual-zero encoding may attenuate or exaggerate differences depending on an indicator's observed range. In the released audit, zero lay below the observed range for 35 of the 43 indicators with residual missingness, and the encoded cells comprised 0.99%–5.75% of the archived league blocks. Therefore, these choices may affect normalization, thresholding, and league membership, particularly for the most affected indicators and regions, and should be assessed with alternative missing-data procedures in future work.

The analysis is regional rather than cell based: the 320 regions are the statistical units, and the 22,720 region-indicator cells are not independent replicates. Country representation is unequal (1–41 retained regions), and spatial dependence means that even the regional observations do not provide the same amount of independent information for every indicator. Accordingly, the results describe the included regions and should not be interpreted as country-balanced estimates or as inferences based on 22,720 indepen-

dent observations. The unrecorded random-number seed of the original stochastic biclustering runs is an additional reproducibility limitation. Therefore, we archive the exact memberships used for the reported results and verify them deterministically without a seed. A separate stochastic verification uses Master Seed 20,260,814 and task-specific-derived seeds; it tests whether the released workflow is executable and repeatable but does not reconstruct the unknown historical random-number stream.

The spatial re-analysis removes ambiguity about implementation but does not make the selected weight specification unique. In particular, the 5-degree band leaves five geographically remote regions without a neighbor, and the deterministic nearest-neighbor fallback creates four connections longer than the nominal band, including a 60.047-degree FRY4–CY00 edge. These edges are transparent in the released weight file, and their influence is limited by row standardization, but alternative specifications that exclude remote territories, use k -nearest neighbors throughout, or employ contiguity weights should be assessed in future sensitivity analyses.

These constraints qualify, rather than erase, the observed distinction between spatially clustered employment strengths and more dispersed innovation weaknesses. They also underscore the value of longitudinal data, alternative imputation procedures, country-balanced sensitivity analyses, and spatial effective-sample-size assessments in future applications.

Future research should examine league stability and regional transitions over time. Longitudinal data would make it possible to evaluate whether observed changes are associated with specific intervention strategies and would provide a more complete account of regional development trajectories.

CRedit authorship contribution statement

Z.T. Kosztyán: Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Resources, Methodology, Data curation, Conceptualization. **Z. Banász:** Data curation. **M.T. Kurbucz:** Writing – review & editing, Writing – original draft, Data curation. **A.I. Katona:** Writing – review & editing, Writing – original draft, Visualization, Data curation. **A. Pribojszki-Németh:** Writing – review & editing, Data curation. **Á. Jakobi:** Writing – review & editing, Writing – original draft, Visualization, Data curation, Conceptualization.

Ethics statement

This research did not involve human participants, human tissue, or animal subjects. The analysis relies exclusively on secondary, publicly available regional statistical indicators; therefore, no ethical approval or informed consent was needed.

Submission declaration

This paper has not been published previously and is not under consideration for publication elsewhere.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Table A.1
Indicators examined. [Source: authors' construction].

Code	Indicator				
P	Population				
P01	Percentage of the population aged	25-64 with	less than primary or lower secondary education.		
P02			upper secondary, postsecondary nontertiary and tertiary education.		
P03			upper secondary and postsecondary nontertiary education.		
P04			tertiary education.		
P05		30-34 with	upper secondary, postsecondary nontertiary and tertiary education.		
P06			upper secondary and postsecondary nontertiary education.		
P07			tertiary education.		
E	Employment				
E01	Total employment rates of young people not in education and training, from less than primary education until postsecondary nontertiary education, for young people aged 15-34 years.				
E02	Employment rates of young people not in education and training.	with less than primary and lower secondary education,	more than 3	years since completion of the highest level of education.	
E03			more than 5		
E04		with nontertiary education,	from 1 to 3	years since completion of the highest level of education.	
E05			more than 3		
E06			more than 5		
E07			less than 3		
E08			less than 5		
E09					
E10		with upper secondary, postsecondary, till tertiary education,	from 1 to 3	years since completion of the highest level of education.	
E11			more than 3		
E12			more than 5		
E13			less than 3		
E14			less than 5		
E15					
E16		upper secondary to nontertiary education,	from 1 to 3	years since completion of the highest level of education.	
E17			more than 3		
E18			more than 5		
E19			less than 3		
E20			less than 5		
E21		with tertiary education,	from 1 to 3		
E22			more than 3		
E23			more than 5		
E24			less than 3		
E25			less than 5		
E26					
E27	General employment rates of young people not in education and training,	from 1-3 years.			
E28		for more than 3 years.			
E29		after more than 5 years.			
E30	Employment rates of young people not in education and training, educational less than	3	years since completion of the highest level of education.		
E31		5			
E32					
ER	Erasmus				
ER1	Count of students in the Erasmus program	in the year when the selected region is the source region.			
ER2		who selected the region as their destination.			
G	Graduates				
G01	Number of graduates in	natural sciences, mathematics, and statistics.			
G02		information and communication technologies.			
G03		engineering, manufacturing, and construction.			
S	Science				
S01	Persons with tertiary education	as a percentage of the total population.			
S02	Persons employed in science and technology				
S03	Persons with tertiary education and/or employed in science and technology				
S04	Scientists and engineers				
Code	Indicator				
S05	Employment in	high and medium high-technology manufacturing.			
S06		medium low-technology manufacturing.			
S07		high-technology sectors (high-technology manufacturing and knowledge-intensive high-technology services).			
S08		knowledge-intensive high-technology services.			
S09		knowledge-intensive sectors.			
S10		professional, scientific and technical activities			
S11		education.			
S12		Number of I4.0-related studies published in the year.			
S13		Total R&D personnel and researchers in the business enterprise sector as a percentage of total employment, measured in full-time equivalents.			
S14		Count of research institutes in the region.			
S15	Persons with tertiary education and employed in science and technology as a percentage of the total population.				
M	Media articles				
M01	Average tone of the media articles queried in the GDELT database matching the	competitive industries			
M02		education skills development and labor market			
M03		employability skills and jobs			
M04		industry policy			
M05		jobs			
M06		manufacturing			
M07		competitive industries			
M08		education skills development and labor market			
M09		employability skills and jobs			
M10		industry policy			
M11		jobs			
M12		manufacturing			

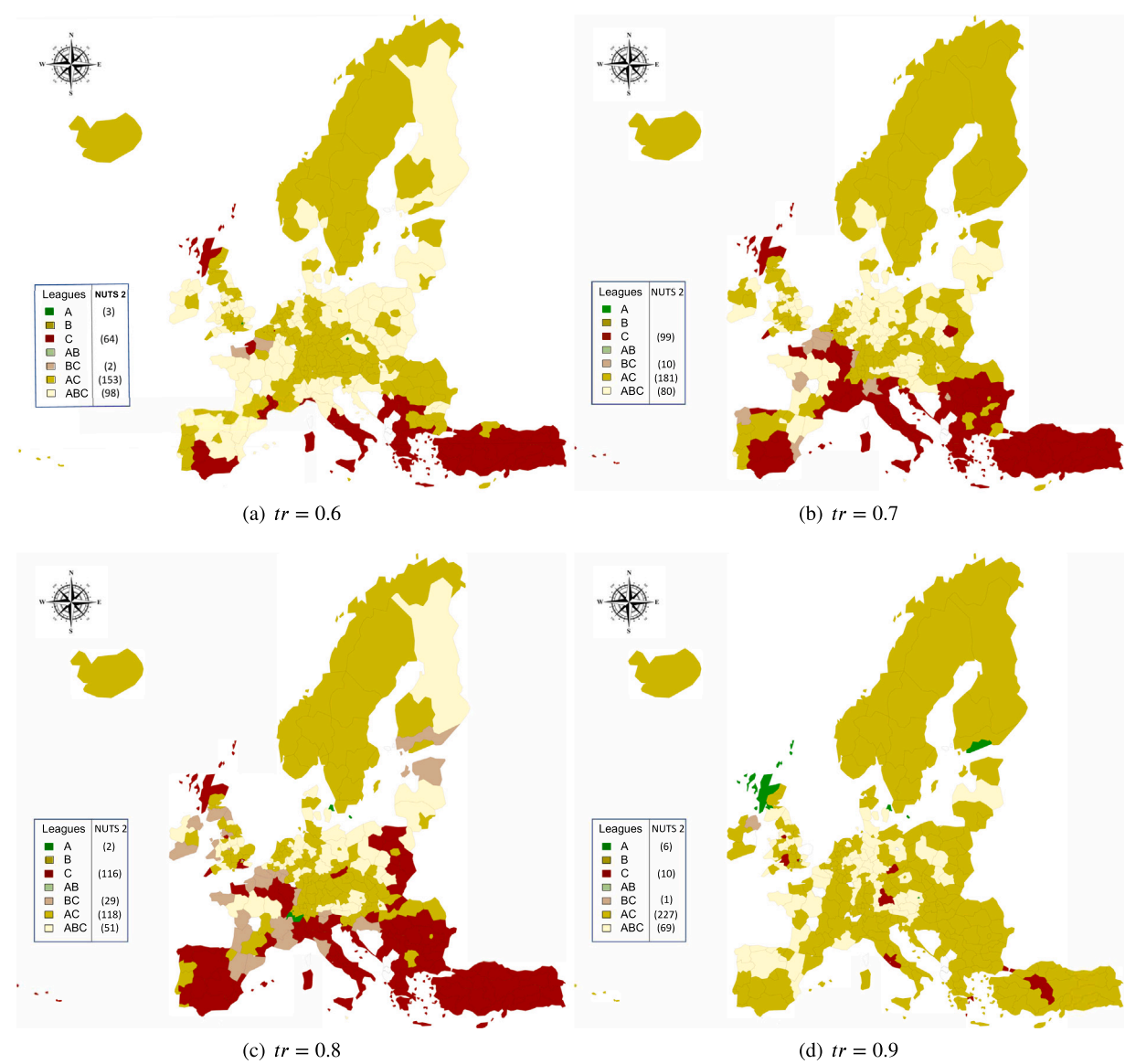


Fig. A.1. Biclustered NUTS 2 regions and regional Venn diagrams across thresholds. [Source: Figure designed by the authors and generated in R.]

Biclustered NUTS 2 regions and regional venn diagrams with different thresholds.

Global Moran statistics of indicators in descending order and their membership leagues .

Table A.2

Global Moran statistics of indicators in descending order and their membership league. An asterisk indicates Benjamini–Hochberg-adjusted $q < 0.05$ within the 71-indicator family. [Source: authors' calculations.]

Rank	Indicators	Moran Index	Normalized mean value	League code (threshold setting)				
				(50)	(60)	(70)	(80)	(90)
1.	P01	0.7716*	0.7158	A +	A +	A +	A +	000
2.	P02	0.7716*	0.7158	A +	A +	A +	A +	000
3.	P05	0.7173*	0.7687	A +	A +	A +	A +	000
4.	P03	0.7052*	0.5241	A +	000	000	000	000
5.	E15	0.6638*	0.7342	A +	A +	A +	A +	000
6.	E10	0.6557*	0.7118	AB	A +	A +	A +	000
7.	S02	0.6283*	0.4549	BC	B	B	B	000
8.	E32	0.6208*	0.7082	A +	A +	A +	A +	000
9.	E16	0.6086*	0.6972	A +	A +	A +	A +	000
10.	E27	0.5938*	0.6811	AB	A +	A +	A +	000
11.	E12	0.5905*	0.6862	AB	AB	AB	A +	000
12.	S09	0.5863*	0.4649	C-	000	000	000	000
13.	E13	0.5793*	0.6881	AB	AB	AB	A +	000
14.	S03	0.5536*	0.4112	BC	B	B	000	000
15.	P06	0.5522*	0.5550	A +	000	000	000	000
16.	E29	0.5481*	0.6733	AB	AB	A +	000	000
17.	E18	0.5397*	0.6567	AB	AB	A +	000	000
18.	E30	0.5352*	0.6744	AB	AB	A +	000	000
19.	E04	0.5253*	0.6670	A +	A +	A +	A +	000
20.	S04	0.5144*	0.3961	BC	BC	B	B	B
21.	E06	0.4883*	0.6458	A +	A +	A +	000	000
22.	E07	0.4777*	0.6413	AB	A +	A +	000	000
23.	E31	0.4717*	0.7432	A +	A +	A +	A +	000
24.	P04	0.4585*	0.3504	BC	C-	000	000	000
25.	S15	0.4479*	0.3622	BC	BC	B	B	B
26.	S01	0.4291*	0.3292	BC	BC	B	000	000
27.	E14	0.4144*	0.7320	A +	A +	A +	A +	000
28.	E19	0.4133*	0.8158	AB	AB	A +	A +	000
29.	E09	0.3816*	0.6979	A +	A +	A +	A +	000
30.	E11	0.3734*	0.7626	A +	A +	A +	A +	000
31.	E28	0.3654*	0.7404	A +	A +	A +	A +	000
32.	S05	0.3588*	0.2281	BC	BC	BC	B	000
33.	E08	0.3248*	0.6224	A +	A +	A +	A +	000
34.	E20	0.3192*	0.6328	A +	A +	A +	A +	000
35.	S10	0.3183*	0.3057	BC	BC	B	B	B
36.	S11	0.3165*	0.3167	BC	BC	B	B	000
37.	E21	0.3137*	0.7058	A +	A +	A +	A +	000
38.	P07	0.3082*	0.4816	C-	000	000	000	000
39.	E26	0.3069*	0.7852	A +	A +	A +	A +	000
40.	S06	0.3044*	0.3130	BC	BC	B	B	000
41.	E05	0.2856*	0.6180	A +	A +	A +	A +	000
42.	E17	0.2782*	0.6416	A +	A +	A +	A +	000
43.	E25	0.2645*	0.7255	A +	A +	A +	A +	000
44.	E23	0.2585*	0.8567	A +	A +	A +	A +	000
45.	S07	0.2468*	0.3091	BC	BC	BC	B	B
46.	E22	0.2372*	0.7378	A +	A +	A +	A +	000
47.	G02	0.1922*	0.1999	BC	BC	BC	BC	B
48.	S08	0.1857*	0.2538	BC	BC	BC	B	B
49.	G03	0.1847*	0.1873	BC	BC	BC	BC	B
50.	E24	0.1702*	0.8382	A +	A +	A +	A +	000
51.	G01	0.1678*	0.1153	BC	BC	BC	BC	BC
52.	S14	0.1487*	0.0681	BC	BC	BC	BC	BC
53.	ER2	0.1375*	0.0555	BC	BC	BC	BC	BC
54.	ER1	0.1342*	0.0863	BC	BC	BC	BC	BC
55.	E01	0.0931*	0.6324	A +	A +	000	000	000
56.	S13	0.0851*	0.0458	C-	C-	C-	C-	C-
57.	M08	0.0838*	0.0236	BC	BC	BC	BC	BC

(continued on next page)

Table A.2 (continued)

Rank	Indicators	Moran Index	Normalized mean value	League code (threshold setting)				
				(50)	(60)	(70)	(80)	(90)
58.	M09	0.0837*	0.0236	BC	BC	BC	BC	BC
59.	E02	0.0786*	0.5818	A +	A +	000	000	000
60.	S12	0.0758*	0.0663	BC	BC	BC	BC	BC
61.	E03	0.0694*	0.6106	A +	A +	000	000	000
62.	M11	0.0684*	0.0234	BC	BC	BC	BC	BC
63.	M10	0.0529*	0.0218	BC	BC	BC	BC	BC
64.	M07	0.0466*	0.0222	BC	BC	BC	BC	BC
65.	M12	0.0445*	0.0166	BC	BC	BC	BC	BC
66.	M06	0.0304*	0.9803	A +	A +	A +	A +	A +
67.	M03	0.0284*	0.9504	A +	A +	A +	A +	A +
68.	M02	0.0280*	0.9454	A +	A +	A +	A +	A +
69.	M05	0.0234*	0.9609	A +	A +	A +	A +	A +
70.	M04	0.0164	0.9766	A +	A +	A +	A +	A +
71.	M01	0.0106	0.9746	A +	A +	A +	A +	A +

Appendix B. Supplementary data

Supplementary data for this article can be found online at doi: [10.1016/j.heliyon.2026.e45501](https://doi.org/10.1016/j.heliyon.2026.e45501).

Data Availability

The I4.0+ indicator data used in this study were obtained from the publicly available dataset deposited by Honti et al. [47] at Mendeley Data (<https://data.mendeley.com/datasets/23gwn43ygp/1>, DOI: <https://doi.org/10.17632/23gwn43ygp.1>). The reproduction archive accompanying this revision contains the raw 2008–2019 panel used to audit temporal substitution, the archived pre-zero matrix and exact missingness mask, the complete list of 320 region codes and names, the dictionary of 71 indicators, the completed 320-by-71 analysis matrix, and league memberships at all five thresholds. It also contains the NUTS 2016 spatial input, the complete row-standardized weight matrix, global and local permutation results, generated figures, diagnostics, session information, seed manifest, analysis scripts, and parameter settings. The archived memberships reproduce the reported results without requiring a seed; Master Seed 20,260,814 is used only for the separately identified stochastic verification. No new primary data were generated for this research.

References

- [1] P. Annoni, L. Dijkstra, EU Regional Competitiveness Index. Technical Report. Joint Research Centre, 2010, <https://doi.org/10.2788/88040>
- [2] L. Anselin, Local indicators of spatial association—LISA, *Geogr. Anal.* 27 (1995) 93–115, <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
- [3] D. Anusil, M. Sushobhan, 6 A Comparative Analysis of Public Aid to Innovation vs Industry-Specific Skill Development, Taylor & Francis, 2024, <https://doi.org/10.4324/9781032613444-8>
- [4] F.R. Ashik, S.A. Mim, M.N. Neema, Towards vertical spatial equity of urban facilities: an integration of spatial and aspatial accessibility, *J. Urban Manag.* 9 (2020) 77–92, <https://doi.org/10.1016/j.jum.2019.11.004>
- [5] C. Bai, P. Dallasega, G. Orzes, J. Sarkis, Industry 4.0 technologies assessment: a sustainability perspective, *Int. J. Prod. Econ.* 229 (2020) 107776, <https://doi.org/10.1016/j.ijpe.2020.107776>
- [6] M.I. Baig, E. Yagdegaridehkordi, Industry 5.0 applications for sustainability: a systematic review and future research directions, *Sustain. Dev.* 32 (2023) 662–681, <https://doi.org/10.1002/sd.2699>
- [7] D. Bailey, L. De Propriis, 6. Industry 4.0, regional disparities and transformative industrial policy, in: *Regional Studies Policy Impact Books*, Taylor and Francis, vol. 1, 2019, pp. 67–78, <https://doi.org/10.1080/2578711X.2019.1621102>
- [8] P.A. Balland, R. Boschma, Mapping the potentials of regions in Europe to contribute to new knowledge production in Industry 4.0 technologies, *Reg. Stud.* (2021) 1–15, <https://doi.org/10.1080/00343404.2021.1900557>
- [9] Z. Banász, Z.T. Kosztyán, V.V. Csányi, A. Telcs, University leagues alongside rankings, *Qual. Quant.* 57 (2023) 721–736, <https://doi.org/10.1007/s11135-022-01374-0>
- [10] F. Barca, P. McCann, A. Rodríguez-Pose, The case for regional development intervention: place-based versus place-neutral approaches, *J. Reg. Sci.* 52 (2012) 134–152, <https://doi.org/10.1111/j.1467-9787.2011.00756.x>
- [11] G. Bentley, L. Pugalís, Shifting paradigms: people-centred models, active regional development, space-blind policies and place-based approaches, *Local Econ.* 29 (2014) 283–294, <https://doi.org/10.1177/02690942145413>
- [12] M. Boccia, A.M. Ferragina, New KIBS specialisation in the Italian provinces over 2007–2020: proximity, ICT and localisation, *Eurasian Bus. Rev.* (2025) 1–61, <https://doi.org/10.1007/s40821-024-00286-3>
- [13] A. Borst, Connectivity matrix seriation via relaxation, *PLOS Comput. Biol.* 20 (2024) e1011904, <https://doi.org/10.1371/journal.pcbi.1011904>
- [14] M. Bottazzi Schenone, V. Tomaselli, M. Vichi, Uncovering socioeconomic disparities in European regions: a Tucker 3 clustering approach, *Qual. Quant.* (2025) 1–19, <https://doi.org/10.1007/s11135-025-02133-7>
- [15] A. Bouët, D. Laborde, US trade wars in the twenty-first century with emerging countries: make America and its partners lose again, *World Econ.* 41 (2018) 2276–2319, <https://doi.org/10.1111/twec.12719>
- [16] S. Bourdin, The interplay of politics and space: how elected politicians shape place-based policies and outcomes, *Geogr. J.* 190 (2024) e12591, <https://doi.org/10.1111/geoj.12591>
- [17] S. Bourdin, N. Jacquet, Closing the loop at the local scale: investigating the drivers of and barriers to the implementation of the circular economy in cities and regions, *Ecol. Econ.* 231 (2025) 108542, <https://doi.org/10.1016/j.ecolecon.2025.108542>
- [18] G.E.P. Box, D.R. Cox, An analysis of transformations, *J. R. Stat. Soc.: B (Methodol.)* 26 (1964) 211–243, <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
- [19] R. Brisco, Understanding Industry 4.0 digital transformation, *Proc. Des. Soc.* 2 (2022) 2423–2432, <https://doi.org/10.1017/pds.2022.245>
- [20] A. Čajkova, E. Romanova, S. Tolstikova, B. Abushkin, Models of regional competitiveness: priority influencing factors, in: *E3S Web of Conferences*, EDP Sciences, 2021, 04005, <https://doi.org/10.1051/e3sconf/202130104005>

- [21] R. Capello, C. Lenzi, Industry 4.0 and servitisation: regional patterns of 4.0 technological transformations in Europe, *Technol. Forecast. Soc. Change* 173 (2021) 121164, <https://doi.org/10.1016/j.techfore.2021.121164>
- [22] R. Capello, A. Rodríguez-Pose, Europe's quest for global economic relevance: on the productivity paradox and the draghi report, *Sci. Reg.* 24 (2025) 7–15, <https://doi.org/10.14650/116132>
- [23] S. Casagrande, B. Dallago, The European core–periphery divide: towards a new narrative, *Eur. J. Econ. Econ. Polic.* 20 (2023) 125–147, <https://doi.org/10.1080/13501763.2023.2290205>
- [24] D. Castellani, F. Lamperti, K. Lavoratori, Measuring adoption of Industry 4.0 technologies via international trade data: insights from European countries, *J. Ind. Bus. Econ.* 49 (2022) 51–93, <https://doi.org/10.1007/s40812-021-00204-y>
- [25] I. Castelo-Branco, F. Cruz-Jesus, T. Oliveira, Assessing Industry 4.0 readiness in manufacturing: evidence for the European Union, *Comput. Ind.* 107 (2019) 22–32, <https://doi.org/10.1016/j.compind.2019.01.007>
- [26] A. Cifolilli, A. Muscio, Industry 4.0: national and regional comparative advantages in key enabling technologies, *Eur. Plan. Stud.* 26 (2018) 2323–2343, <https://doi.org/10.1080/09654313.2018.1535573>
- [27] E. Commission, Cohesion in Europe towards 2050. Eighth report on economic, social and territorial cohesion, 2022, <https://doi.org/10.2776/624081>
- [28] J. Content, N. Cortinovis, K. Frenken, J. Jordaan, The roles of KIBS and R&D in the industrial diversification of regions, *Ann. Reg. Sci.* (2022) 1–36, <https://doi.org/10.1007/s00168-021-01068-9>
- [29] C. Corradini, E. Santini, C. Vecchiolini, The geography of Industry 4.0 technologies across European regions, *Reg. Stud.* (2021) 1–14, <https://doi.org/10.1080/00343404.2021.1884216>
- [30] C.J. Coyne, P.T. Leeson, Read all about it! Understanding the role of media in economic development, *Kyklos* 57 (2004) 21–44, <https://doi.org/10.1111/j.0023-5962.2004.00241.x>
- [31] T. Czvetkó, G. Honti, J. Abonyi, Regional development potentials of Industry 4.0: open data indicators of the Industry 4.0 + model, *PLoS One* 16 (2021) e0250247, <https://doi.org/10.1371/journal.pone.0250247>
- [32] S. Demin, A. Mikhaylova, S. Pyankova, Digitalization and its impact on regional economy transformation mechanisms, *Int. J. Syst. Assur. Eng. Manag.* 14 (2023) 377–390, <https://doi.org/10.1007/s13198-022-01806-y>
- [33] P. Di Caro, U. Fratesi, One policy, different effects: estimating the region-specific impacts of EU cohesion policy, *J. Reg. Sci.* 62 (2022) 307–330, <https://doi.org/10.1111/jors.12566>
- [34] A.M. Diah, J.L. Away, T. Kadang, Economic management in developing economies: strategies for sustainable growth and development, *Join: J. Soc. Sci.* 1 (2024) 363–372, <https://doi.org/10.59613/3knayg46>
- [35] J.C. Duque, R. Ramos, J. Surriñach, Supervised regionalization methods: a survey, *Int. Reg. Sci. Rev.* 30 (2007) 195–220, <https://doi.org/10.1177/0160017607301605>
- [36] S. Emili, F. Galli, Spatial and cross-sectoral input spillover effects: the case of the Italian tourism industry, *J. Product. Anal.* 59 (2023) 243–258, <https://doi.org/10.1007/s1123-023-00665-4>
- [37] I. Europe, Policy briefs from the policy learning platform on research and innovation. Policy Learning Platform on Research and innovation 2019. https://www.interreg-europe.eu/fileadmin/user_upload/plp_uploads/policy_briefs/INDUSTRY_4.0_Policy_Brief.pdf
- [38] M. Eva, A. Cehan, E. Corodescu-Roșca, S. Bourdin, Spatial patterns of regional inequalities: empirical evidence from a large panel of countries, *Appl. Geogr.* 140 (2022) 102638, <https://doi.org/10.1016/j.apgeog.2022.102638>
- [39] H. Floerkemeier, N. Spatafora, A. Venables, How to Achieve Inclusive Growth, Regional Disparities, Oxford University Press, 2021, <https://doi.org/10.1093/oso/9780192846938.001.0001>
- [40] J.M. Flórez-Parra, M.V. López-Pérez, A.M. López-Hernández, M.L. Arco-Castro, Determinants of the dissemination of economic, social and environmental information at the university level in the context of commitment to the sustainable development goals, *Sustain. Dev.* n/a (2023) 1–15, <https://doi.org/10.1002/sd.2760>
- [41] P. Gestraud, BiCARE: Biclustering Analysis and Results Exploration, 2008, <http://bioinfo.curie.fr>. R package version 1.38.0.
- [42] N. Ghafar, K. Harun, M. Lailawati, N. Rahman, Preliminary investigation of Industry 4.0 readiness among small and medium aviation companies in Malaysia, in: *Journal of Physics: Conference Series*, IOP Publishing, 2020, 012030. <https://doi.org/10.1088/1742-6596/1573/1/012030>
- [43] M. Ghobakhloo, M. Iranmanesh, B. Foroughi, M.L. Tseng, D. Nikbin, A.A. Khanfar, Industry 4.0 digital transformation and opportunities for supply chain resilience: a comprehensive review and a strategic roadmap, *Prod. plan. control* 36 (2025) 61–91, <https://doi.org/10.1080/09537287.2023.2252376>
- [44] D. Gusenleitner, E.A. Howe, S. Bentink, J. Quackenbush, A.C. Culhane, iBBiG: iterative binary bi-clustering of gene sets, *Bioinformatics* 28 (2012) 2484–2492, <https://doi.org/10.1093/bioinformatics/bts438>
- [45] A. Haleem, M. Javaid, R. Vaishya, Industry 4.0 and its applications in orthopaedics, *J. clin. orthop. trauma* 10 (2019) 615, <https://doi.org/10.1016/j.jcot.2018.09.015>
- [46] G. Honti, T. Czvetkó, J. Abonyi, Data describing the regional Industry 4.0 readiness index, *Data Br.* 33 (2020) 106464, <https://doi.org/10.1016/j.dib.2020.106464>
- [47] G. Honti, T. Czvetkó, J. Abonyi, Data describing the regional Industry 4.0 readiness index, 2020, <https://doi.org/10.17632/23gwn43ygp.1>
- [48] F. Huneus, R. Rogerson, Heterogeneous paths of industrialization, *Rev. Econ. Stud.* 91 (2024) 1746–1774, <https://doi.org/10.1093/restud/rdad066>
- [49] S. Iammarino, A. Rodríguez-Pose, M. Storper, Regional inequality in Europe: evidence, theory and policy implications, *J. Econ. Geogr.* 19 (2019) 273–298, <https://doi.org/10.1093/jeg/lby021>
- [50] Z.T. Kosztayán, Z. Banász, V.V. Csányi, A. Telcs, Rankings or leagues or rankings on leagues? - Ranking in fair reference groups, *Tert. Educ. Manag.* 25 (2019) 289–310, <https://doi.org/10.1007/s11233-019-09028-x>
- [51] Z.T. Kosztayán, T. Cizmádia, B.S.G. Pató, S. Berke, I. Neumann-Virág, A. Bencsik, Measuring organizational happiness in higher education-a data-driven approach, *Cogent Educ.* 10 (2023) 2190382, <https://doi.org/10.1080/2331186X.2023.2190382>
- [52] P. Krugman, Increasing returns and economic geography, *J. Polit. Econ.* 99 (1991) 483–499, <https://doi.org/10.1086/261763>
- [53] E. Lafuente, Y. Vaillant, F. Vendrell-Herrero, Territorial servitization and the manufacturing renaissance in knowledge-based economies, 2019, <https://doi.org/10.1080/00343404.2018.1542670>
- [54] J. Leendertse, M. Schrijvers, E. Stam, Measure twice, cut once: entrepreneurial ecosystem metrics, *Res. Policy* 51 (2022) 104336, <https://doi.org/10.1016/j.respol.2021.104336>
- [55] D. Lepore, F. Spigarelli, Integrating Industry 4.0 plans into regional innovation strategies, *Local Econ.* 35 (2020) 496–510, <https://doi.org/10.1177/0269094220937452>
- [56] I. Liiv, Seriation and matrix reordering methods: an historical overview, *Stat. Anal. Data Min.: ASA Data Sci. J.* 3 (2010) 70–91, <https://doi.org/10.1002/sam.10071>
- [57] Y. Liu, L. Zou, Y. Wang, Spatial-temporal characteristics and influencing factors of agricultural eco-efficiency in China in recent 40 years, *Land Use Policy* 97 (2020) 104794, <https://doi.org/10.1016/j.landusepol.2020.104794>
- [58] M. López-Golán, F. Campos-Freire, J.Á. Fernández-Holgado, Catalysts of change: technological innovations shaping Spanish public proximity media, *Journal. Media* 5 (2024) 444–455, <https://doi.org/10.3390/journalmedia5020029>
- [59] D. Lu, et al., *Regional Development and Its Spatial Structure*, Springer, 2024, <https://doi.org/10.1007/978-981-99-7683-6>
- [60] Y. Lu, The current status and developing trends of Industry 4.0: a review, *Inf. Syst. Front.* 27 (2025) 215–234, <https://doi.org/10.1007/s10796-021-10221-w>
- [61] D. MacKinnon, L. Kempton, P. O'brien, E. Ormerod, A. Pike, J. Tomaney, Reframing urban and regional 'development' for 'left behind' places, *Camb. J. Reg. Econ. Soc.* 15 (2022) 39–56, <https://doi.org/10.1093/cjres/rsab034>
- [62] S.C. Madeira, A.L. Oliveira, Biclustering algorithms for biological data analysis: a survey, *IEEE/ACM Trans. Comput. Biol. Bioinform.* 1 (2004) 24–45, <https://doi.org/10.1109/TCBB.2004.2>
- [63] W. Maisiri, L. van Dyk, R. Coetzee, Factors that inhibit sustainable adoption of Industry 4.0 in the South African manufacturing industry, *Sustainability* 13 (2021) 1013, <https://doi.org/10.3390/su13031013>

- [64] R. Martin, P. Sunley, The new economic geography and policy relevance, *J. Econ. Geogr.* 11 (2011) 357–369, <https://doi.org/10.1093/jeg/lbq042>
- [65] Y. Ménière, Patents and the fourth Industrial Revolution: the global technology trends enabling the data economy, in: *Platforms and Artificial Intelligence: The Next Generation of Competences*, Springer, 2022, pp. 103–109, https://doi.org/10.1007/978-3-030-90192-9_5
- [66] H.D. Mohammadian, IoT-Education technologies as solutions towards SMEs' educational challenges and 14. 0 readiness, in: *2020 IEEE Global Engineering Education Conference (EDUCON)*, IEEE, 2020, pp. 1674–1683, <https://doi.org/10.1109/EDUCON45650.2020.9125249>
- [67] P.A. Moran, Notes on continuous stochastic phenomena, *Biometrika* 37 (1950) 17–23, <https://doi.org/10.2307/2332142>
- [68] E. Moretti, Place-based policies and geographical inequalities, *Oxf. Open Econ.* 3 (2024) i625–i633, <https://doi.org/10.1093/ooec/odad049>
- [69] O. Murire, L. Cilliers, W.T. Chinyamurindi, Social media role in addressing employability challenges of graduates: an expert opinion analysis, *SA J. Hum. Resour. Manag.* 21 (2023) 2047, https://hdl.handle.net/10520/ejc-sajhrm.v21_n1_a2047.
- [70] D. Nigohosyan, A. Vutsova, The 2014–2020 European regional development fund indicators: the incomplete evolution, *Soc. Indic. Res.* 137 (2018) 559–577, <https://doi.org/10.1007/s11205-017-1610-8>
- [71] M.J. Nowak, Integrated development planning and local spatial policy tools, *J. Econ. Manag.* 41 (2020) 69–86, <https://doi.org/10.22367/jem.2020.41.04>
- [72] A. Omri, Technological innovation and sustainable development: does the stage of development matter? *Environ. Impact Assess. Rev.* 83 (2020) 106398, <https://doi.org/10.1016/j.eiar.2020.106398>
- [73] A. Patil, A. Dwivedi, M.A. Moktadir, et al., Big data-Industry 4.0 readiness factors for sustainable supply chain management: towards circularity, *Comput. Ind. Eng.* 178 (2023) 109109, <https://doi.org/10.1016/j.cie.2023.109109>
- [74] F.L. Pinheiro, P.A. Balland, R. Boschma, D. Hartmann, The dark side of the geography of innovation: relatedness, complexity and regional inequality in Europe, *Reg. Stud.* 59 (2025) 2106362, <https://doi.org/10.1080/00343404.2022.2106362>
- [75] M.E. Porter, *Competitive Advantage: Creating and Sustaining Superior Performance*, Simon and Schuster, 2008.
- [76] G. Pronesti, G. Pronesti, Cluster and smart specialization: different approaches to design innovation policy, *Life Cycle Clust. Des. Smart Spec. Polic.* (2019) 11–50, https://doi.org/10.1007/978-3-030-03780-2_2
- [77] V. Raponi, F. Martella, A. Maruotti, A biclustering approach to university performances: an Italian case study, *J. Appl. Stat.* 43 (2016) 31–45, <https://doi.org/10.1080/02664763.2015.1009005>
- [78] E.W. Solikhah, H.M. Asih, F.H. Astuti, I. Ghazali, E.B. Mohammad, Industry 4.0 readiness trends: a bibliometric and visualization analysis, *Int. J. Robot. Control Syst.* 4 (2024) 105–124, <https://doi.org/10.31763/ijrcs.v4i1.1247>
- [79] Ö. Sölvell, C. Ketels, G. Lindqvist, Industrial specialization and regional clusters in the ten new EU member states, *Compet. Rev.: Int. Bus. J.* 18 (2008) 104–130, <https://doi.org/10.1108/10595420810874637>
- [80] E.A. Stryabkova, M.V. Vladyka, J.V. Lyshchikova, A.Y. Rzaev, M.A. Kochergin, Smart specialization as a comprehensive territorial and sectoral approach to determining regional development priorities, *J. Environ. Manag. Tour.* 12 (2021) 1353–1370, [https://doi.org/10.14505/jemt.v12.5\(53\).20](https://doi.org/10.14505/jemt.v12.5(53).20)
- [81] J.E. Teixeira, A.T.C. Tavares-Lehmann, Industry 4.0 in the European Union: policies and national strategies, *Technol. Forecast. Soc. Change* 180 (2022) 121664, <https://doi.org/10.1016/j.techfore.2022.121664>, <https://www.sciencedirect.com/science/article/pii/S0040162522001962>.
- [82] S. Teske, K. Nagrath, Global sector-specific scope 1, 2, and 3 analyses for setting net-zero targets: agriculture, forestry, and processing harvested products, *SN Appl. Sci.* 4 (2022) 221, <https://doi.org/10.1007/s42452-022-05111-y>
- [83] W.R. Tobler, A computer movie simulating urban growth in the Detroit region, *Econ. Geogr.* 46 (1970) 234–240, <https://doi.org/10.2307/143141>
- [84] M.P. Todaro, S.C. Smith, *Economic Development*, Pearson UK, 2020.
- [85] B. Wang, Y. Miao, H. Zhao, J. Jin, Y. Chen, A biclustering-based method for market segmentation using customer pain points, *Eng. Appl. Artif. Intell.* 47 (2016) 101–109, <https://doi.org/10.1016/j.engappai.2015.06.005>
- [86] V.A. Wankhede, S. Vinodh, Benchmarking Industry 4.0 readiness evaluation using fuzzy approaches, *Benchmarking Int. J.* 30 (2023) 281–306, <https://doi.org/10.1108/BLJ-08-2021-0505>
- [87] R. Wisdawati, R. Nooraeni, B.C. Laksono, B.I. Fatah, Implementation of factor analysis and BiClustering in classifying multidimensional under-five poverty in East Nusa Tenggara, *IJCCS (Indones. J. Comput. Cybern. Syst.)* 16 (2022) 301–312, <https://doi.org/10.22146/ijccs.70433>
- [88] X. Wu, X. Hui, Economic dependence relationship and spatial stratified heterogeneity in the eastern coastal economic belt of China, *Complexity* 2021 (2021) 6645451, <https://doi.org/10.1155/2021/6645451>
- [89] J. Xie, A. Ma, A. Fennell, Q. Ma, J. Zhao, is time to apply biclustering: a comprehensive review of biclustering applications in biological and biomedical data, *Brief. Bioinform.* 20 (2018) 1450–1465, <https://doi.org/10.1093/bib/bby014>
- [90] Q. Zhang, P. Wu, R. Li, A. Chen, Digital transformation and economic growth efficiency improvement in the digital media era: digitalization of industry or digital industrialization? *Int. Rev. Econ. Finance* 92 (2024) 667–677, <https://doi.org/10.1016/j.iref.2024.02.010>