UNIVERSITY OF
MODENA E REGGIO EMILIA

PhD Course in
Labour, Development and Innovation

Cycle XXXIV

Innovation, social cohesion and inequalities in Europe: a poset-based approach

Innovazione, coesione sociale e disuguaglianze in Europa: un approccio poset-based

Innovación, cohesión social y desigualdades en Europa: un método poset-based

Candidate: Filippo Damiani

Tutor: Professor Silvia Muzzioli

Co-Tutor: Professor Paula Rodríguez-Modroño

PhD Coordinator: Professor Tindara Addabbo

Ringraziamenti

Questa tesi di dottorato è dedicata a mio nonno Agostino Melotti ed al mio amico Giacomo Gnoli.

Mi è doveroso dedicare questo spazio del mio elaborato alle persone che hanno contribuito, con il loro fondamentale supporto, a rendere speciale la mia esperienza di dottorato.

Ringrazio la Prof.ssa Silvia Muzzioli (tutor di questa tesi) per aver accettato di esplorare nuovi ambiti di ricerca, per aver creduto nelle mie idee e per aver creato i presupposti affinché si realizzassero.

Un ringraziamento speciale va alla Prof.ssa Paula Rodríguez-Modroño (co-tutor di questa tesi) per avermi accolto nella splendida Universidad Pablo de Olavide di Siviglia e per contribuito significativamente a farmi appassionare a quella parte della ricerca che si batte incessantemente per il superamento delle disuguaglianze.

Grazie alla Prof.ssa Tindara Addabbo, per aver creduto in me coinvolgendomi nell’attività didattica e in altri numerosi progetti che hanno contribuito in maniera decisiva a rendere straordinaria l’esperienza del dottorato.

Ringrazio, inoltre, il Prof. Bernard De Baets per aver contribuito in modo decisivo ad alzare la qualità di questo lavoro e per avermi mostrato nuovi aspetti metodologici.

Un sentito grazie al mio collega e amico Dott. Filippo Ferrarini per essere stato un essenziale compagno di viaggio, sempre pronto ad accendere la luce nei momenti di difficoltà.

Un grazie va alle colleghe Dott.ssa Federica Pillo, Dott.ssa Shahin Manafi Varkiani e Dott.ssa Chiara Tasselli, e ai colleghi Dott. Andrea Marchioni, Dott. Massimo Bianchi, Dott. Luca Gambarelli, Dott. Giuseppe Caruso e Dott. Francesco Faenza, per la fondamentale condivisione di molti progetti e attività in questi anni.

Ringrazio la mia famiglia e la mia compagna di viaggio Maria Banari per il costante supporto e per darmi la forza di inseguire i miei sogni.

Infine, ringrazio la Prof.ssa Margherita Russo per essere stata la prima ad aver creduto in me e per essere stata una grande Maestra.

Acknowledgments

This doctoral thesis is dedicated to my grandfather Agostino Melotti and to my friend Giacomo Gnoli.

I feel the duty to mention the people who have contributed, with their crucial support, to making my PhD experience special.

I would like to thank Prof. Silvia Muzzioli (tutor of this thesis) for agreeing to explore new areas of research, for believing in my ideas and for creating the conditions for them to become reality.

Special thanks go to Prof. Paula Rodríguez-Modroño (co-tutor of this thesis) for welcoming me to the wonderful Pablo de Olavide University in Seville and for contributing significantly to attract me to that part of research that relentlessly strives to overcome inequalities.

Thanks to Prof. Tindara Addabbo, for believing in me and involving me in teaching activities and in numerous other projects that have contributed decisively to making the PhD experience extraordinary.

I would also like to thank Prof. Bernard De Baets for his decisive contribution to raising the quality of this work and for showing me new methodological aspects.

Heartfelt thanks to my colleague and friend Dr. Filippo Ferrarini for being an essential travelling companion, always ready to turn on the light in times of difficulty.

Thanks to the colleagues Dr. Federica Pillo, Dr. Shahin Manafi Varkiani, Dr. Chiara Tasselli, Dr. Andrea Marchioni, Dr. Massimo Bianchi, Dr. Luca Gambarelli, Dr. Giuseppe Caruso, and Dr. Francesco Faenza, for the fundamental sharing of many projects and activities in these years.

I wish to thank my family and my girlfriend Maria Banari for their constant support and for giving me the strength to pursue my dreams.

Finally, I thank Prof. Margherita Russo for being the first to believe in me and for being a great Master.

Abstract – ENG

This dissertation stems from a co-tutored doctoral project between the University of Modena and Reggio Emilia and the Pablo de Olavide University of Seville. The main objective consisted in the development of two complex topics related to innovation, social cohesion and inequalities in Europe in three different papers. The common thread consists in the methodology used: the partially ordered set (poset), a method based on the assumption that an object can be identified as "better" than another if and only if it has better results in all the indicators analysed in the comparison. We have chosen to use this methodology to propose a different data analysis capable of overcoming de issues of composite indicators, and to build rankings not based on the simple arithmetic mean of the normalised indicators.

In the first two papers we focused on the vast theme of regional innovation, analysing the performance at the regional level of 220 regions in the first paper, and of 60 regions (those of the four greatest countries of southern Europe, namely Greece, Italy, Portugal, and Spain) in the second paper. For both papers, we used data from the Regional Innovation Scoreboard 2019, through which we built a ranking, dividing the regions analysed into different performance levels. The creation of clusters of similar regions combined with poset analysis, allowed us to identify differences between the ranking presented in this thesis and the ranking proposed by the Regional Innovation Scoreboard. In particular, it was possible to identify the indicators that have the greatest impact in determining the results, and consequently the movements of the regions in the ranking, making it possible to propose targeted policies based also on the country or cluster of regions analysed. In the analyses conducted, one of the most impacting indicators is *Individual design applications per billion GDP (in purchasing power standards)*. Regarding the leaders of the analysis, we found that the majority are regions housing the capital city of the country.

The third paper addressed the issue of gender inequalities in the digital economy by using the data of the Women in Digital Scoreboard 2020 concerning the 27 countries of the European Union and the United Kingdom. Also in this case, a ranking of countries split into four performance levels was obtained. The impact analysis of the indicators revealed that the most meaningful are: *% of people with above basic digital skills in information, communication, problem solving and software for content creation*; *Graduates in STEM subjects per 1000 individuals aged 20-29*; *Gender pay gap in unadjusted form, considering all employees working in firms with ten or more employees*. In this case, the sensitivity analysis performed on the indicators, made it possible to identify strengths and weaknesses of the individual countries. At the same time, the findings helped us to propose areas of intervention aimed at improving the results in the most critical indicators in order to increase the position in the ranking. The results also highlighted important differences between the different European macro-regions; in particular, nations belonging to southern and eastern Europe are clearly behind to those belonging to the north and a large part of western Europe.

Keywords

social cohesion, regional innovation, gender equality, poset, ranking

Abstract – ITA

Questa tesi nasce da un progetto di dottorato in co-tutela tra l'Università di Modena e Reggio Emilia e la Universidad Pablo de Olavide di Siviglia, il cui obiettivo ha previsto lo sviluppo di due tematiche riguardanti l'innovazione, la coesione sociale, e le disuguaglianze in Europa attraverso tre distinti elaborati. Il filo rosso che li accomuna è rappresentato dalla metodologia utilizzata, il partially-orderded set (poset), il cui metodo di analisi fondamentale si basa sul fatto che un dato oggetto può essere identificato come "migliore" rispetto ad un altro se e solo se presenta risultati migliori in tutti gli indicatori analizzati nella comparazione. Abbiamo scelto di utilizzare questa metodologia per proporre una diversa analisi dei dati in grado di superare i problemi degli indicatori compositi e di costruire ranking basati non sulla semplice media aritmetica degli indicatori normalizzati.

Nei primi due paper ci si è focalizzati sul vasto tema dell'innovazione regionale, andando ad analizzare la performance a livello regionale di 220 regioni nel primo paper, e di 60 regioni (quelle dei quattro grandi Paesi del sud Europa, ovvero Grecia, Italia, Portogallo, e Spagna) nel secondo paper. Per entrambi i paper sono stati utilizzati i dati del Regional Innovation Scoreboard del 2019, attraverso i quali si è costruito un ranking, dividendo le regioni analizzate in diversi livelli di performance. La creazione di cluster di regioni simili combinata all'analisi poset, ci ha permesso di identificare differenze tra il ranking presentato in questa tesi e il ranking proposto dal Regional Innovation Scoreboard. In particolare, è stato possibile individuare gli indicatori che hanno un maggiore impatto nel determinare i risultati, e di conseguenza gli spostamenti delle regioni nel ranking, permettendo di avanzare proposte di policies mirate in base anche al Paese o al cluster di regioni analizzato. Nelle analisi effettuate, uno degli indicatori risultati tra i più impattanti è l'indicatore *Design applications individuali per miliardo di PIL (a parità di potere d'acquisto)*. Per quanto riguarda le regioni leader, troviamo specialmente le regioni localizzate nel nord-ovest d'Europa contenenti la capitale del Paese.

Nel terzo paper si è affrontato il tema delle disuguaglianze di genere nell'economia digitale attraverso l'analisi dei dati provenienti dal Women in Digital Scoreboard del 2020 riguardanti i 27 Paesi dell'Unione Europea e il Regno Unito. Anche in questo caso si è ottenuto un ranking dei Paesi in base ai livelli di performance. Lo studio ha rilevato che gli indicatori più impattanti sono: *% di persone con competenze digitali superiori a quelle di base in materia di informazione, comunicazione, problem solving e software per la creazione di contenuti*; *Laureate in materie STEM per 1000 individui di età compresa tra 20-29 anni*; *Divario* *retributivo di genere non corretto, considerando tutti i dipendenti che lavorano in imprese con dieci o più dipendenti*. In questo caso è stata effettuata un'analisi di sensitività sugli indicatori che ha permesso di identificare punti di forza e punti di debolezza dei singoli Paesi proponendo aree di intervento mirate al miglioramento dei risultati negli indicatori più critici per poter migliorare la posizione nel ranking. I risultati hanno inoltre evidenziato importanti differenze tra le diverse macro-regioni europee; in particolare, il sud e l'est Europa evidenziano un netto ritardo rispetto al nord e buona parte dell'ovest Europa.

Keywords

coesione sociale, innovazione regionale, uguaglianza di genere, poset, ranking

Abstract – ESP

Esta tesis doctoral es resultado de un proyecto de doctorado en co-tutela entre la Universidad de Módena y Reggio Emilia y la Universidad Pablo de Olavide de Sevilla. El objetivo principal consiste en el desarrollo de dos investigaciones relacionadas con los nuevos cambios que las innovaciones tecnológicas provocan en la cohesión social y las desigualdades en Europa en tres artículos diferentes. En todos ellos la metodología utilizada es la llamada “partially-ordered set” (poset), un método que parte del supuesto de que un objeto puede ser identificado como "mejor" que otro si y sólo si tiene mejores resultados en todos los indicadores analizados para compararlos. Se utiliza esta metodología para proponer un análisis alternativo de los datos, superando los problemas de los indicadores compuestos, y construir ranking no basados en la simple media aritmética de los indicadores normalizados.

Los dos primeros artículos de la tesis se centran en la medición de la innovación regional en Europa, analizando la performance a nivel regional de 220 regiones en el primer artículo, y de 60 regiones (las de los cuatro grandes países del sur de Europa, España, Grecia, Italia y Portugal) en el segundo artículo. Para ambos trabajos, hemos utilizado datos del Regional Innovation Scoreboard 2019, construyendo un ranking y dividiendo las regiones en diferentes niveles de resultados. La creación de clusters de regiones similares combinados con el análisis poset, nos permite identificar diferencias entre el ranking presentado en esta tesis y el ranking propuesto por el Regional Innovation Scoreboard. En particular, fue posible identificar los indicadores que tienen un mayor impacto en la determinación de los resultados y, en consecuencia, los movimientos de las regiones en el ranking, lo que permite proponer políticas específicas para el país o cluster de regiones. En el análisis, uno de los indicadores más impactantes es *Diseño de aplicaciones individuales por billón de PIB (en estándar de poder adquisitivo)*. En cuanto a las regiones líderes del ranking, encontramos que en su mayoría son las regiones donde se localiza la capital del país.

En el tercer artículo nos hemos centrado en analizar las desigualdades de género en la economía digital, utilizando los datos del Women in Digital Scoreboard 2020 relativos a los 27 países de la Unión Europea y el Reino Unido. También en este caso, se obtuvo un ranking de países dividido en cuatro niveles de resultados. El análisis de impacto de los indicadores reveló que los más significativos son: *% de personas con habilidades digitales superiores a las básicas en información, comunicación, resolución de problemas y software para la creación de contenidos*; *Graduadas en materias STEM por cada 1000 personas de 20 a 29 años*; *Brecha* *salarial de género en forma no ajustada, considerando a todos los empleados que trabajan en empresas con diez o más empleados*. En este caso, el análisis de sensibilidad realizado sobre los indicadores permitió identificar las fortalezas y debilidades de los países y proponer áreas de intervención dirigidas a mejorar los resultados en los indicadores más críticos. Los resultados también han destacado diferencias entre las diferentes macrorregiones europeas; en particular, las naciones pertenecientes a Europa meridional y oriental están claramente atrasadas con respecto a las pertenecientes al norte y a una gran parte de Europa occidental.

Keywords

cohesión social, innovación regional, igualdad de género, poset, ranking

Index

Introduction

It is generally acknowledged that innovation, social cohesion, and inequalities are complex phenomena, which must be supported by adequate structures of synthetic indicators, capable of rendering accurate pictures of the dynamics described, and to set targets policies towards them.

Synthesis allows the dissolution of complexity by making easier the analysis of phenomena. The traditional approach to synthesis "is the so-called composite indicator approach, a methodology that consists in the aggregation through a mathematical function of a set of basic indicators, which represent the different dimensions of a phenomenon" (Alaimo et al., 2021, p.79). According to Fattore (2017), this approach has significant limitations: first of all, the "synthesis-as-aggregation" paradigm does not allow the synthesis of ordinal attributes, since synthesizing through composite indicators flattens the dynamism of phenomena (Alaimo & Maggino 2020). Moreover, aggregated indicators, "collapse complexity into composite scores, hide the nuances and multi-faceted nature of social phenomena, are hard to interpret, and lose a great deal of information on the structure and the patterns of society, leading to policies that cannot be fine-tuned to the shape of social needs. Given its impact on public debates and policy making, it seems thus urgent to draw attention on the measurement problem, from both conceptual and practical points of view" (Arcagni et al., 2021, p.2).

To provide a concrete solution to the issues generated by adopting composite indicators and to perform data analysis computing synthetic indicators for policy and decision making, it has been proposed in the literature the systematic use of partial order theory (poset), a set of algebraic and combinatoric tools designed to describe and properly treat order relations and ordinal data (Fattore, 2016). This methodology is particularly suitable for the management of multidimensional systems of ordinal data. Several examples of employing partial order theory in social studies and other applied sciences can be found in the literature, such as in Brüggemann et al. (1999), Annoni (2007), Annoni & Brüggemann (2009), Brüggemann & Patil (2010), Annoni et al. (2011), Brüggemann & Patil (2011), De Loof et al. (2011), Freier et al. (2011), Carlsen & Brüggemann (2014), Badinger & Reuter (2015), Fattore et al. (2015), Bachtrögler et al. (2016); Carlsen (2017), Carlsen & Brüggemann (2017b), Iglesias et al. (2017), Caperna & Boccuzzo (2018), Cavalletti & Corsi (2018), di Bella et al. (2018), Fuhrmann et al. (2018), Arcagni et al. (2019), and Beycan et al. (2019). What these examples show is that the concepts and the tools from partial order theory allow for problems, such as ranking or evaluation, "to be consistently addressed in an ordinal setting, paving the way to deeper, more reliable and more effective representations of complex socio-economic traits" (Fattore & Arcagni, 2021, p.221).

In the poset theory, as stated in Fattore (2016), information is extracted directly from ordinal data and synthesis is achieved with no attribute aggregation, overcoming the limitations of composite indicators and counting approaches. Complexity refers to the impossibility of capturing a concept through aggregative-compensative approaches ultimately based on dimensional reduction tools. A complex concept comprises many different dimensions, logically related, but possibly too weakly interdependent to be statistically "composed" in a satisfactory way. The use of partial ordering has two different types of justification in inequality evaluation. First, complex systems may have enough ambiguity and fuzziness to make it a mistake to look for a complete ordering of either, this may be called the fundamental reason for incompleteness. Second, even if it is not a mistake to look for one complete ordering, it could be difficult, in practice, to identify it. Partial orders may indeed convey a great deal of information for evaluation, and the theory of partially ordered sets provides the right toolbox to exploit it (Fattore, 2016).

More specifically, many socio-economic problems are naturally conceptualized and formalized in terms of order relations and must then be addressed in ordinal terms, i.e. by using concepts and tools from the theory of partial orders. If the observed objects (in the case of this thesis we will refer to regions or countries) have so-called conflicting scores, and this is quite often the case, data can be ordered only partially, producing a partially ordered set (Carlsen & Brüggemann, 2017a). Moreover, partial orders are useful also "when numerical data systems are to be addressed and one does not want to, or cannot, mix variables through aggregated procedures, like those leading to composite indicators. The data structure and the tools adopted in any statistical analysis must be as faithful and consistent as possible with the phenomenon of interest and, in many situations, posets are the appropriate choice. Partial orders are ubiquitous in socio-economics and make their appearance whenever multi-criteria decision problems based on multi-indicator systems are to be addressed" (Fattore & Arcagni, 2021, pp.219-220).

The aim of this thesis was to study innovation, social cohesion, and inequalities in Europe by analysing two different data set that have been widely used in the literature: the Regional Innovation Scoreboard and the Women in Digital Scoreboard. By using the poset methodology, a different data analysis is proposed, with the aim of overcoming de issues of composite indicators, building rankings not based on the simple arithmetic mean of the normalised indicators. In the first paper – first chapter – the performance in innovation of 220 European regions (including British regions) is analysed; in the second paper – second chapter – social cohesion is studied, by focusing on the 60 regions of southern Europe in a two-steps analysis; in the third paper – third chapter – inequalities are studied by analysing the data about women digital skills across the 28 European countries (including UK).

References

Alaimo, L.S., Arcagni, A., Fattore, M., & Maggino, F. (2021). Synthesis of Multi-indicator System Over Time: A Poset-based Approach. *Social Indicators Research 157*, 77–99. https://doi.org/10.1007/s11205-020-02398-5.

Alaimo, L. S., & Maggino, F. (2020). Sustainable development goals indicators at territorial level: Conceptual and methodological issues—the Italian perspective. *Social Indicators Research, 147(2)*, 383–419. https://doi.org/10.1007/s1120 5-019-02162 -4.

Annoni, P. (2007). Different ranking methods: Potentialities and pitfalls for the case of European opinion poll. *Environmental and Ecological Statistics, 14*, 453–471. https://doi.org/10.1007/s10651-007-0041-0.

Annoni, P., & Brüggemann, R. (2009). Exploring partial order of European countries. *Social Indicators Research, 92*, 471–487. https://doi.org/10.1007/s11205-008-9298-4.

Annoni, P., Fattore, M., & Brüggemann, R. (2011). A multi-criteria fuzzy approach for analyzing poverty structure. *Statistica and Applicazioni, Special Issue*, 7–30.

Arcagni, A., di Belgiojoso, E. B., Fattore, M., & Rimoldi, S. M. L. (2019). Multidimensional analysis of deprivation and fragility patterns of migrants in Lombardy, using partially ordered sets and self-organizing maps. *Social Indicators Research, 141(2)*, 551–579. https://doi.org/10.1007/s11205-018-1856-9.

Arcagni, A., Fattore, M., Maggino, F., Vittadini, G. (2021). Some Critical Reflections on the Measurement of Social Sustainability and Well-Being in Complex Societies. *Sustainability 2021, 13*, 12679. https://doi.org/10.3390/su132212679.

Bachtrögler, J., Badinger, H., de Clairfontaine, A. F., & Reuter, W. H. (2016). Summarizing data using partially ordered set theory: An application to fiscal frameworks in 97 countries. *Statistical Journal of the IAOS, 32(3),* 383–402. https://doi.org/10.3233/SJI-160973.

Badinger, H., & Reuter, W. H. (2015). Measurement of fiscal rules: Introducing the application of partially ordered set (poset) theory. *Journal of Macroeconomics, 43,* 108–123. https://doi.org/10.1016/j.jmacro.2014.09.005.

Beycan, T., Vani, B. P., Brüggemann, R., & Suter, C. (2019). Ranking Karnataka districts by the multidimensional poverty index (MPI) and by applying simple elements of partial order theory. *Social Indicators Research, 143(1),* 173–200. https://doi.org/10.1007/s11205-0181966-4.

Brüggemann, R., & Patil, G. P. (2010). Multicriteria prioritization and partial order in environmental sciences. *Environmental and Ecological Statistics, 17,* 383–410. https://doi.org/10.1007/s10651-010-0167-3.

Brüggemann, R., & Patil, G. P. (2011). Ranking and prioritization for multi-indicator systems—Introduction to partial order applications. New York: Springer. ISBN: 978-1-4419-8477-7.

Brüggemann, R., Pudenz, S., Voigt, K., Kaune, A., & Kreimes, K. (1999). An algebraic/graphical tool to compare ecosystems with respect to their pollution. IV: Comparative regional analysis by Boolean arithmetics. *Chemosphere, 38,* 2263–2279. https://doi.org/10.1016/s0045-6535(98)00445-7.

Caperna, G., & Boccuzzo, G. (2018). Use of poset theory with big datasets: A new proposal applied to the analysis of life satisfaction in italy. *Social Indicators Research, 136(3),* 1071–1088. https://doi.org/10.1007/s11205-016-1482-3.

Carlsen, L. (2017). An alternative view on distribution keys for the possible relocation of refugees in the european union. *Social Indicators Research, 130(3),* 1147–1163. https://doi.org/10.1007/s11205-016-1234-4.

Carlsen, L., & Brüggemann, R. (2014). The ‘‘Failed State Index’’: Offers more than just a simple ranking. *Social Indicators Research, 115,* 525–530. https://doi.org/10.1007/s11205-012-9999-6.

Carlsen, L., & Brüggemann, R. (2017a). Partial Ordering and Metrology Analyzing Analytical Performance. In Partial Order Concepts in Applied Sciences, 49-70, https://doi.org/10.1007/978-3-319-45421-4_4.

Carlsen, L., & Brüggemann, R. (2017b). Fragile state index: Trends and developments. A partial order data analysis. *Social Indicators Research, 133(1)*, 1–14. https://doi.org/10.1007/s11205-016-1353-y.

Cavalletti, B., & Corsi, M. (2018). "Beyond GDP" effects on national subjective well-being of OECD countries. *Social Indicators Research, 136(3)*, 931–966. https://doi.org/10.1007/s11205-016-1477-0.

De Loof, K., De Baets, B., & De Meyer, H. (2011). Approximation of average ranks in posets. *MATCH Communications in Mathematical and in Computer Chemistry, 66*, 219–229. ISSN:0340-6253.

di Bella, E., Gandullia, L., Leporatti, L., Montefiori, M., & Orcamo, P. (2018). Ranking and prioritization of emergency departments based on multi-indicator systems. *Social Indicators Research, 136(3)*, 1089–1107. https://doi.org/10.1007/s11205-016-1537-5.

Fattore, M. (2016). Partially Ordered Sets and the Measurement of Multidimensional Ordinal Deprivation. *Social Indicators Research 128*, 835-858. https://doi.org/10.1007/s11205-015-1059-6.

Fattore, M. (2017). Synthesis of indicators: The non-aggregative approach. In Complexity in society: From indicators construction to their synthesis, 193–212. Berlin: Springer. https://doi.org/10.1007/978-3-319-60595-1_8.

Fattore, M., & Arcagni, A. (2021). Posetic Tools in the Social Sciences: A Tutorial Exposition. In: Measuring and Understanding Complex Phenomena, Indicators and their Analysis in Different Scientific Fields, 219-241. https://doi.org/10.1007/978-3-030-59683-5_15.

Fattore, M., Maggino, F., & Arcagni, A. (2015). Exploiting ordinal data for subjective well-being evaluation. *Statistics in Transition New Series, 3(16)*, 409–428. https://doi.org/10.21307/stattrans-2015-023.

Freier, K. P., Brüggemann, R., Scheffran, J., Finckh, M., & Schneider, U. A. (2011). Assessing the predictability of future livelihood strategies of pastoralists in semi-arid Morocco under climate change. *Technological Forecasting and Social Change, 79*, 371–382. https://doi.org/10.1016/j.techfore.2011.07.003.

Fuhrmann, F., Scholl, M., & Brüggemann, R. (2018). How can the empowerment of employees with intellectual disabilities be supported? *Social Indicators Research, 136(3)*, 1269–1285. https://doi.org/10.1007/s11205-017-1666-5.

Iglesias, K., Suter, C., Beycan, T., & Vani, B. P. (2017). Exploring multidimensional well-being in switzerland: Comparing three synthesizing approaches. *Social Indicators Research, 134(3)*, 847–875. https://doi.org/10.1007/s11205-016-1452-9.

Chapter 1: A poset-based analysis of regional innovation at European level

Filippo Damiani[a], Silvia Muzzioli[b], Bernard De Baets[c]

Abstract

This paper examines the performance of regional innovation across 220 European regions. First, a cluster analysis is performed in order to detect patterns of comparable regions. Subsequently, a poset-based approach is adopted to obtain a ranking of the different clusters of European regions. The outcome is compared with the results described in the Regional Innovation Scoreboard 2019. Useful insights for policymakers are obtained.

Keywords

Regional innovation, partially ordered set, poset, Regional Innovation Scoreboard, European Union.

Acknowledgments

The work was supported by the FAR2019 project. We also wish to thank William Bromwich for his painstaking attention to the copy-editing of this paper.

[a] Marco Biagi Department of Economics, University of Modena and Reggio Emilia, Modena, Italy filippo.damiani@unimore.it
[b] Marco Biagi Department of Economics, University of Modena and Reggio Emilia, Modena, Italy silvia.muzzioli@unimore.it
[c] KERMIT, Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium bernard.debaets@ugent.be

1.1 Introduction

Innovation is a key driver of economic growth (Garud et al., 2013; Rondia et al., 2019), which is widely considered as conducive to improvements in standards of living (Acemoglu, 2012). Over the last decades, several innovation systems have been developed, at the national (Nelson, 1992) as well as regional level (Cooke et al., 1997). Such systems are thought to be the most reliable representation of the environment that is needed to create and develop innovation (Őnday, 2016). However, most policymakers and researchers agree that innovation is primarily determined at the regional level (Doloreux & Parto, 2004; Navarro et al., 2009; Lau & Lo, 2015). In fact, although the free movement of capital and labour is increasing, knowledge accumulation and exploitation remain spatially concentrated (European Union, 2014; Acs et al., 2017; Jang et al., 2017), and highly subordinated to socioeconomic and institutional conditions (Rodríguez-Pose & Crescenzi, 2008) as well as the capacity of a region to generate knowledge spillovers (Segarra-Blasco et al., 2018).

In the literature, the Regional System of Innovation (RSI), or Regional Innovation System (RIS) concept has been widely studied (Fernandes et al., 2020). It consists of actors, such as organisations, institutions, firms and stakeholders, and of the relationships between them (Tödtling & Trippl, 2005; Uyarra, 2010; McCann & Ortega-Argilés, 2015; Uyarra et al., 2017). These linkages should be encouraged and supported to generate positive results (Isaksen et al., 2018), as good regional innovation patterns positively influence regional economic performance (Capello & Lenzi, 2019). It is essential for policymakers to evaluate these patterns in a proper manner in order to better use the available resources to improve the results.

Measuring innovation at the regional level involves the choice of indicators. Some of the indicators that are considered to measure regional innovation performance are related to investment in research and development activities at both the public and private level, the support for public-private partnerships and the number of researchers employed in the region (Ponsiglione et al., 2018). In particular, technology development efficiency is higher in regions where R&D is more public-focues (Min et al., 2020). Moreover, knowledge creation, absorptive capacity and governance capacity are found to play a role in innovation (Navarro et al., 2009; Hajek et al., 2014). The number of patents per capita is also a main driver of innovation (Aghion et al., 2019) in cases in which the restrictions on intellectual property are not too strict (Moser, 2016). Other studies find that it is important to monitor the innovation activities through indicators such as *the percentage of SMEs that are innovating in-house for firms in high-*

concentration markets (Love & Roper, 2001; Doran et al., 2020), or *the number of collaborations among innovative SMEs aiming to enhance new-to-the-firm forms of innovation* (Sarpong & Teirlinck, 2018). Also the *sales of new-to-market and new-to-firm innovations as a percentage of total turnover* should be considered to evaluate the sales impact (Matras-Bolibok et al., 2017). These indicators are included in the most exhaustive available index that makes a comparative assessment of the innovation performance at the regional level: the Regional Innovation Scoreboard (RIS). It considers indicators subdivided into four pillars: framework conditions, investments, innovation activities and impacts (Hauser et al., 2018; European Union, 2019a).

From a classical point of view, as conceptualised by Schumpeter - one of the most influential economists of the twentieth century - innovation systems are complex systems (Schumpeter, 1935). The same concept has been adopted more recently by other authors (Kats, 2006; Asheim et al., 2013). An appropriate method is required to analyse such systems in the best possible way and better orient public policies for creating regional advantages in different contexts (Asheim et al., 2013). The method considered as reference point in the analysis of such complex systems is the Regional Innovation Scoreboard, that provides a final ranking of the European regions based on the average value of the indicators considered, making it more difficult to clearly detect whether a region is underperforming on some of the indicators. For this reason, in this paper, we aim to present an alternative approach borrowed from the theory of partially ordered sets (theory of posets, or poset theory, for short; see Subsection 1.3.2), that will permit us to obtain a ranking avoiding the use of aggregation methods (Fattore, 2016; Fattore & Arcagni, 2018; Ivaldi et al., 2020) and without pre-treatment of data: the performance can be evaluated considering all indicators simultaneously (Carlsen & Brüggemann, 2017). Therefore, the poset methodology is useful to overcome the curse of dimensionality without using a parametric model or introducing some subjective criteria. The poset-based approach is suitable in socioeconomics whenever multi-criteria decision problems based on multi-indicator systems are to be addressed (Fattore & Arcagni, 2021). It has been adopted for different purposes, including the calculation of new indices on the stringency of fiscal rules (Badinger & Reuter, 2015), the evaluation of multidimensional poverty (Fattore & Arcagni, 2014), the assessment of river water quality (Tsakovski et al., 2010), the synthetisation of multi-indicator systems over time (Alaimo et al., 2020), and various applications in chemistry (De Loof et al., 2008). The main strengths of the poset-based approach can be summarised as follows: it respects the ordinal nature of data, it maintains a high standard of objectivity (hence, reducing the need for subjective choices), and it fully exploits all information contained in the dataset (Badinger & Reuter, 2015). Through these characteristics, it is possible to identify relevant insights such as the impact of indicators in the construction of the ranking. We apply the poset-based approach to the regional data available from the RIS 2019 (the most recent ranking). The analysis consists of two steps: first, a clustering of the regions is carried out, and second, the poset-based approach is applied to establish a ranking of these clusters. The application of the poset-based approach is feasible even on a large dataset with thousands of observations. In this paper we propose a cluster analysis to create groups of similar regions and the attribute-related sensitivity analysis to find out the most impacting indicators of the four categories of the RIS. These two methods permit us also to reduce the incomparabilities (see Subsection 1.3.2). We also propose a poset analysis without the creation of clusters as a robustness test (Fattore & Maggino, 2014).

This paper is organised as follows. Section 1.2 analyses the methods applied to measure regional innovation performance and describes the Regional Innovation Scoreboard 2019. Section 1.3 presents the dataset and the methods used, with particular attention to the description of the poset theory and the various steps of the analysis. Section 1.4 presents the results of the study. The last section is dedicated to the discussion of findings, conclusions, limitations, and perspectives for future research. Appendix 1.A shows an example of data analysis using the poset-based approach illustrating the steps performed in this work. Appendix 1.B describes the regions analysed in the study and provides additional information about the clusters and the membership to a performance group.[1] Appendix 1.C presents the results of the poset analysis considering all the 220 regions and all the 17 indicators without the creation of clusters. We observe certain differences in the performance group membership for some European regions compared to the results of the Regional Innovation Scoreboard 2019. This comparison could be of interest to policymakers and innovation ecosystem actors to gain a better understanding of which regions are similar in terms of innovation performance and which indicators should be targeted in order to increase the position of a cluster of regions in the ranking.

[1] In Appendix 1.B, all 220 European regions analysed in this study are listed following the order of the ranking obtained after the data analysis from the poset-based approach, comparing their performance group both in the poset-based analysis and in the RIS 2019.

1.2 How regional innovation is measured

Since the beginning of the 1990s, studies about innovation have been on the rise, due to the need to understand the driving forces leading to a high innovation performance and the best methods to describe the phenomenon of innovation. In Subsection 1.2.1 we review the currently available indices and in Subsection 1.2.2 we describe the Regional Innovation Scoreboard, which is taken as our benchmark.

1.2.1 Currently available indices

Measuring the innovation performance of a region is a difficult task, due to the changing nature of innovation, in particular since the development of the global economy (Council of Competitiveness, 2005). Furthermore, even if finding data at regional level is harder than at national level, regional data availability is better than before, allowing for a more sophisticated assessment of innovation performance (Borrás & Jordana, 2016). Before considering the most effective way to measure regional innovation performance, it may be useful to provide a literature review of the state-of-the-art on the measurement of innovation performance in a broader context.

One of the most popular methods for measuring innovation is to consider a single indicator, such as patent statistics (Bottazzi & Peri, 2003; Bilbao-Osorio & Rodríguez-Pose, 2004); this method is frequently used to measure innovation at the firm level (Bock et al., 2012; Clauss, 2017). Another approach frequently adopted is the use of an extensive set of indicators (Hauser et al., 2018; Capello & Lenzi, 2013), which makes it possibile to construct different typologies of innovation processes. A third method considers a number of innovation indicators to create a composite index. This holds, for instance, for the Bloomberg Innovation Index, the Global Innovation Index and the European Innovation Scoreboard, the most frequently adopted indices to cast light on innovation and compare performances at the country level (Hogan & Gallaher, 2018). However, the three indices collect data just at country level; this is the reason why, starting from the European Innovation Scoreboard, the Regional Innovation Scoreboard has been developed, collecting data about all the regions of the European Union and the neighboring regions.

The discussion about which approach should be considered the best for the measurement of innovation performance is still open. However, for each of the three methods discussed above, some problems have been identified. For instance, the adoption of a single indicator is useful only in cases in which the focus of the analysis is on a specific aspect of innovation: to find more evidence of the innovation performance of a region, multiple indicators are required (Hauser et al., 2018). Moreover, as suggested by some researchers, policymakers should contemplate the results of different analyses to obtain a more comprehensive view of a regional innovation system (Zabala-Iturriagagoitia et al., 2007), as innovation is a complex phenomenon that cannot be entirely explained with the use of proxy statistics; as a result, linkages between input indicators and output (intended to describe innovation performance) could be fuzzy.

Despite the existence of various composite indicators to measure innovation, as discussed above, the most popular is the European Innovation Scoreboard (EIS), as it provides a comparative assessment of all European Member States, facilitating the understanding of which areas they should focus on in order to improve their results (European Union, 2019a; Ter Haar, 2018). The Regional Innovation Scoreboard (RIS) has been developed on the basis of the EIS and is considered the most important index at the regional level (Hauser at al., 2018). It assesses the innovation performance of European regions since 2009 and at present covers more than two hundred regions. For this reason, the RIS dataset is at the basis of this study and will be extensively discussed in the next section. Similar to the work of researchers in other fields (Caperna & Boccuzzo, 2018), our poset-based approach aims to provide an alternative analysis to offer more insight into the complex phenomenon of regional innovation by relying on ordinal data, avoiding the use of more synthetic measures such as ranks constructed simply on the basis of the average of indicators.

1.2.2 Regional Innovation Scoreboard

The Regional Innovation Scoreboard (RIS) is the regional extension of the European Innovation Scoreboard (EIS). The most recent EIS ranking, published in 2019, assesses the innovation performance of all 27 member states of the EU, in addition to other non-EU countries (including the United Kingdom), analysing the scores of 27 different indicators. As already mentioned, the regional availability of the data is more complex, in fact, the RIS 2019 is limited to the use of regional data for 17 of the 27 indicators included in the EIS. The RIS 2019 is the ninth annual ranking, and the regional coverage has increased compared to previous years. It now includes 238 NUTS 2 (Nomenclature of Territorial Units for Statistics) representing 22 European countries, including Norway, Serbia, Switzerland and the United Kingdom. It also includes five NUTS 1 (countries: Cyprus, Estonia, Latvia, Luxembourg, and Malta) that are considered in the same way as NUTS 2. Hence, the total number of objects (regions and countries) analysed is 243.

The 17 indicators are grouped into four different categories: framework conditions (population aged 30-34 with tertiary education, lifelong learning, international scientific co-publications, top 10% most cited publications), investments (R&D expenditure in public sector, R&D expenditures in business sector, non-R&D innovation expenditure), innovation activities (SMEs with product or process innovations, SMEs with marketing or organisational innovations, SMEs innovating in-house, innovative SMEs collaborating with others, public-private co-publications, PCT patent applications, trademark applications, design applications), and impacts (knowledge-intensive services exports, sales of new-to-market and new-to-firm innovations).

The regional data of the listed indicators are taken mostly from Eurostat; other sources include the OECD REGPAT database, Community Innovation Survey (CIS) data, National Statistical Offices, CWTS (Leiden University) as part of a contract with the European Commission (DG Research and Innovation). Before imputation, data availability is 90.9%, even if 10 out of 17 indicators have an availability of at least 95%. After the application of several imputation techniques (based on the availability of regional or national data referring to the previous year of observation), data availability increases to 98.9% and some data are still missing just for a few regions of 10 different countries, with Ireland and Serbia showing the lowest result: 94.1%. Data is then normalised adopting the min-max procedure. The minimum and the maximum are calculated based on the data of the last five biennial observations. The final index is obtained by applying a country correction factor (based on the results at the national level reported in the EIS) to the average of the normalised scores of the 17 indicators (European Union, 2019b).

Once the scores have been calculated, the regions are grouped into four different categories: innovation leaders with a relative performance higher than 120% of the EU average; strong innovators with a relative performance between 90% and 120% of the EU average; moderate innovators with a relative performance between 50% and 90% of the EU average; and modest innovators with a relative performance below 50% of the EU average. The RIS 2019 includes 38 regions in the group of innovation leaders, 170 in the middle groups (73 regions as strong innovators and 97 as moderate innovators) and 30 in the group of modest innovators. Each performance group is further divided into three subgroups. Regarding the performance, the leaders are mostly regions from Switzerland, Finland, Sweden, the United Kingdom, Denmark, the Netherlands, and Germany, whereas the modest innovators are mostly regions from Poland, Bulgaria, and Romania (European Union, 2019a).

An interesting aspect of the RIS 2019 is the average score of the indicators per regional performance group. Considering the EU average equal to 100, the report of the RIS 2019 shows that 15 out of 17 indicators have the best score in the leaders group and the worst score in the modest group. Just two indicators follow a different pattern. The first one is the indicator *innovative SMEs collaborating with others*, which has a slightly higher score in the strong innovators group than in the leader innovator group (126 vs 118); however, the difference is small, and in the moderate and modest innovator groups the score is much lower compared to the leader innovator group. The second one is the indicator *non-R&D innovation expenditures*, which shows the highest scores in the strong and moderate innovator groups, whereas in the innovation leaders group, it has an outcome similar to that of the modest innovator group. Hence, it seems that in this context, this indicator does not respect the outcomes of the innovation performance groups. The problems with the *non-R&D innovation expenditures* indicator have been discussed before in the literature, for instance, in Blažek & Kadlec, 2019 and Spescha & Woerter, 2019. For this reason, it was decided to exclude this indicator from our analysis.

A limitation of the Regional Innovation Index lies in the fact that the final regional score is simply computed as the average of all indicators, which could be affected by so-called compensation effects, as in the case of arithmetic addition (Munda, 2008; Carlsen, 2018). To be more precise, the low performance of a region for one indicator could be compensated by a high score for another one. Limitations of the RIS have been also identified in (Carayannis et al., 2018), where the ranking was revisited by using a multiple criteria decision analysis approach combining AHP and TOPSIS methods in the context of the Quadruple Innovation Helix framework. In this paper, we propose to use a poset-based approach that, unlike (Carayanni et al., 2018), permits to identify the indicators with the strongest impact, which are used to construct a ranking of the clusters.

1.3 Material and methods

In this section, we provide a description of the dataset adopted for the study of regional innovation (Subsection 1.3.1) and of the methods adopted in the different steps of our investigation (Subsection 1.3.2).

1.3.1 Material

The analysis has been performed on the dataset obtained from the website of the Regional Innovation Scoreboard 2019.[2] As explained in Section 1.2, we exclude the indicator *non-R&D innovation expenditures* from the dataset. The full dataset therefore contains data for the 16 indicators discussed in Section 1.2. In our analysis we considered 220 regions out of 238: the 208 regions of the EU Member States (involving in total 22 different countries), plus the 12 regions of the United Kingdom that were part of the EU in 2019. We excluded the non-EU regions (Norway, Switzerland, and the Republic of Serbia).[3] All regions are NUTS 2.

1.3.2 Methods

The initial data matrix is composed of 220 objects (regions) and 16 indicators (attributes), with 31 missing data. The first step of the analysis is the imputation of the missing data. To this end, we used the nearest neighbour imputation method, a commonly applied method (Jadhav et al., 2019). More precisely, we considered the five nearest neighbour values for computing each of the missing data. The imputation was done for each indicator separately. After imputation, the data matrix contains 3520 observations.

The application of the poset-based approach to a large dataset could generate results that are difficult to interpret. As a result, we reduced the number of objects (regions) through a cluster analysis by performing a hierarchical clustering with the default distance measure, namely the Euclidian distance measure; the function used is ‘hclust’ with the complete linkage method (using the software R). The scores of the clusters correspond (for each attribute) to the average of the scores of the objects that compose each cluster. The number of attributes (indicators) is reduced to two through the attribute-related sensitivity method (see Appendix 1.A, Table 1.A.4). After the reduction of both the number of objects and attributes, the last step corresponds to the application of the poset-based approach to the final data matrix (composed of the 11 clusters of regions and the two most impacting indicators for each of the four categories) to create a ranking (using the software PyHasse).

The main assumption of the poset-based approach is summarised as follows. Two objects (a and b) that are compared on the basis of two different attributes (q_1 and q_2) can be ordered (ranked) if and only if one of them has at least the same performance as the other one on both

[2] The database is available at the following link: https://bit.ly/3cc8PAP.

[3] We did not consider small European countries such as Cyprus, Estonia, Latvia, Luxembourg and Malta in the analysis. In the RIS 2019, their data at country level was used in the NUTS 2 analysis.

attributes (in the unlikely case of exactly the same performance, the objects would be tied). On the contrary, if, for instance, a has a higher performance than b on q_1 and b has a higher performance than a on q_2, then the two objects are called incomparable, and it is not possible to establish an order between them (Brüggemann & Patil, 2011).

The ordering of the objects can be represented graphically through a Hasse diagram, which makes it possible to visually display the most important characteristics of a partially ordered set (poset, for short): the relationships among objects, and the isolated elements (objects that are not comparable with any other object). To better understand the characteristics of this approach, we provide an example in Appendix 1.A.

1.4 Results

As explained in the previous section, after the imputation of the missing data, the dataset is composed of 220 regions belonging to 22 countries of the European Union plus the United Kingdom and includes 16 indicators. As 3520 data are too many to be analysed with the poset-based approach, it is necessary to create clusters of regions.

1.4.1 Cluster analysis

The first step is the computation of the distance matrix, showing for each pair of objects (regions) their Euclidean distance considering all the indicators. The clusters are then created based on the distance matrix according to the complete linkage method.

The choice of the number of clusters (k) is based on the inspection of the scree plot. One of the most popular methods for selecting the number of clusters is the 'elbow method' (Bholowalia & Kumar, 2014); however, as no elbow is visible in the scree plot in our case, we decided to select a number of clusters to be able to reduce the 'within group sum of squares' and at the same time obtain a sufficiently rich partial order. This is attained by choosing a number of clusters ranging from 8 to 11. We chose the maximum number of clusters (11) to limit the variability inside clusters. In this case, the matrix is formed by 11 rows (clusters of regions) and 16 columns (indicators).[4] The elements of the matrix represent the average value of the regions

[4] The score of a cluster is the averages of the scores, between 0 and 1 (normalised values), of the regions that compose the cluster.

that are included in the cluster, for each indicator. We also report the results considering eight clusters to validate the results obtained with 11 clusters.

According to the cluster analysis performed, the number of regions included in the different clusters is not homogeneous. In particular, we observe a large cluster consisting of 61 regions (cluster n°2) and another one that includes just one region (cluster n°11).[5] This is the first result of the analysis: the Finnish region of Åland (an archipelago) shows data that is incomparable with all other regions included in the dataset, and, with $k = 11$, it is impossible to include it in any cluster.

1.4.2 Attribute-related sensitivity analysis

In order to reduce the number of indicators from 16 to 8, in this step of the analysis we aim to select the two most impacting indicators for each of the four categories. Since the two categories investments and impacts are formed by two indicators each, it is not necessary to perform any reduction for them. As a result, we apply the attribute-related sensitivity analysis to the two remaining categories. We reduce the four indicators of the category *framework conditions* and the eight indicators of the category *innovation activities*.

Starting with *framework conditions*, we consider a data matrix consisting of the 11 clusters as objects and the four indicators of the category under analysis. After obtaining the Hasse diagram representing the relationships among the clusters for this category, it is important to compute the total number of incomparabilities as an estimate of the complexity of the poset, and then to find the pair of indicators that reproduces the closest number of incomparabilities. There are 24 incomparabilities in the Hasse diagram generated considering all four attributes of the category. The indicators *population aged 30-34 with tertiary education* and *lifelong learning* alone create 17 incomparabilities (71% of the total); thus, they are the ones with the strongest impact for the category and will be considered in the final data matrix.

Regarding the category *innovation activities*, there are eight indicators. As a result, the number of possible pairwise combinations is quite high. In this case, the pair of indicators with the strongest impact is formed by *innovative SMEs collaborating with others* and *design applications*, representing 31 incomparabilities out of a total of 36 (86%).

[5] As some clusters contain several regions, the variability inside these clusters could be quite high. As a result, some regions might be considered as outliers of such cluster (as in the case of Drenthe and Valle d'Aosta).

At this stage, we are able to construct the final data matrix that is shown in Table 1.1: 11 clusters and eight indicators, representing the two with the strongest impact for each category, listed as follows. 1. Framework conditions: 1a. *Percentage of population aged 30-34 having completed tertiary education*; 1b. *Lifelong learning, the share of population aged 25-64 enrolled in education or training aimed at improving knowledge, skills, and competences*. 2. Investments: 2a. *R&D expenditure in public sector as percentage of GDP*; 2b. *R&D expenditure in business sector as percentage of GDP*. 3. Innovation activities: 3a. *Innovative SMEs collaborating with others as percentage of SMEs*; 3b. *Individual design applications per billion GDP (in purchasing power standards)*. 4. Impacts: 4a. *Employment in medium-high and high-tech manufacturing and knowledge-intensive services*; 4b. *SMEs sales of new-to-market and new-to-firm innovations as percentage of total turnover*.

Table 1.1 – Final data matrix: 11 clusters and 8 indicators with the strongest impact (data normalised)

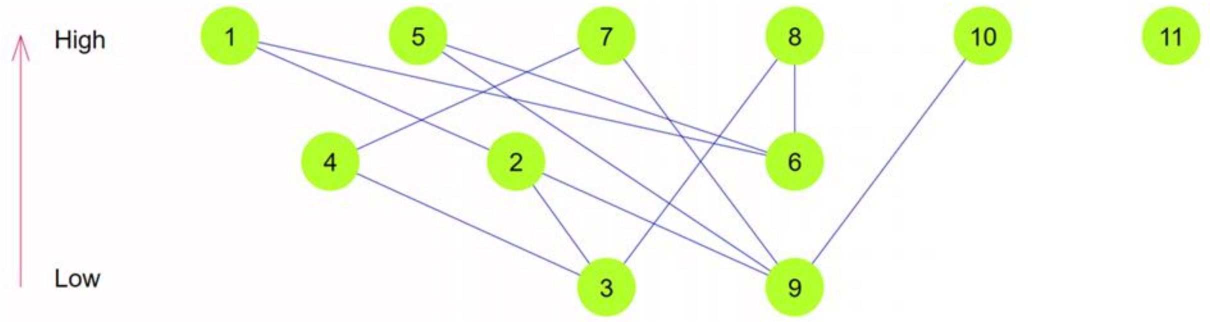
Cluster	***1a***	***1b***	***2a***	***2b***	***3a***	***3b***	***4a***	***4b***
1	0.560	0.551	0.605	0.704	0.592	0.458	0.511	0.563
2	0.329	0.329	0.467	0.476	0.352	0.455	0.476	0.545
3	0.289	0.054	0.199	0.212	0.090	0.399	0.293	0.284
4	0.526	0.151	0.379	0.392	0.164	0.517	0.409	0.369
5	0.731	0.284	0.514	0.526	0.339	0.344	0.691	0.530
6	0.355	0.210	0.265	0.317	0.208	0.231	0.394	0.503
7	0.725	0.898	0.824	0.745	0.506	0.597	0.688	0.500
8	0.398	0.266	0.772	0.663	0.276	0.605	0.683	0.531
9	0.282	0.145	0.181	0.445	0.313	0.163	0.236	0.497
10	0.541	0.276	0.352	0.510	0.918	0.306	0.360	0.877
11	0.293	0.724	0.251	0.078	0.838	0.146	0.436	0.110

The entries for cluster n°11 are just those of Åland since it is the only region in this cluster.

1.4.3 Poset-based analysis

The Hasse diagram obtained from the final data matrix is shown in Figure 1.1.

Figure 1.1 - 11 clusters (220 European regions), Hasse Diagram



The Hasse diagram clearly shows the relations between the clusters. For instance, it is evident that clusters 1, 5, 7, 8 and 10 do not have any links with clusters positioned at a higher level. At the same time, clusters 3, 6 and 9 do not have any links with clusters positioned at a lower level. Cluster 11 deserves special attention as it is the only one that is incomparable with all the other clusters. We already expected this result as cluster 11 consists of just one region, which is the Finnish archipelago of Åland; hence, as already commented, the scores for cluster 11 coincide with the data of Åland itself. To obtain and discuss the ranking of the clusters, we should look at the final score of each cluster, which is obtained by applying the Local Partial Order Model (LPOM).

The Local Partial Order Model highlights three levels of performance: the top level composed of, in order, clusters 1, 7, 5 and 8 (the last two have the same score); the middle level, formed by clusters 10, 2, 11 and 4; finally, the low level, containing clusters 6, 9 and 3.

The regions in the top and the bottom level are quite equally distributed (64 vs 67 regions), while the middle level is the one with the highest number of regions: 89. At this level, we find both cluster 2, consisting of 61 regions, and cluster 11, the one-of-a-kind cluster (Åland). Åland is incomparable with all the other clusters since it has a very good performance on some indicators such as *lifelong learning* and *innovative SMEs collaborating with others*, whereas it has a very low performance on other indicators, including *R&D expenditure in business sector*, *design applications* and *sales of new-to-market and new-to-firm innovations*. More detailed results are provided in Figure 1.2, which shows the composition of each cluster and gives information about the number of regions for each country.

Figure 1.2 - Composition of the 11 clusters (220 European regions)

Cluster	BE	BG	CZ	DK	DE	IE	EL	ES	FR	HR	IT	LT	HU	NL	AT	PL	PT	RO	SI	SK	FI	SE	UK	TOTAL
1	2			2	3				4					7	3						3	2	3	29
7				2																	1	3		6
8					17									1										18
5			1			2		4					1			1			1	1				11
10							5					1											8	14
2	1		6	1	18	1	1		8		12		1	3			4		1			3	1	61
11																					1			1
4		1						4								7		1						13
6			1					10		1	1		6	1						3				23
9							7		2	1	8	1					3							22
3		5						1								9		7						22
TOTAL	3	6	8	5	38	3	13	19	14	2	21	2	8	12	3	17	7	8	2	4	5	8	12	220

	BE	BG	CZ	DK	DE	IE	EL	ES	FR	HR	IT	LT	HU	NL	AT	PL	PT	RO	SI	SK	FI	SE	UK	TOTAL
Level 1	2	0	1	4	20	2	0	4	4	0	0	0	1	8	3	1	0	0	1	1	4	5	3	64
Level 2	1	1	6	1	18	1	6	4	8	0	12	1	1	3	0	7	4	1	1	0	1	3	9	89
Level 3	0	5	1	0	0	0	7	11	2	2	9	1	6	1	0	9	3	7	0	3	0	0	0	67

The clusters are ordered from top-performing to low-performing. It should be noted that clusters 5 and 8 are tied in the ranking. As regards the colours, blue represents the top-performing clusters, green identifies the middle-performing ones, and red the low-performing clusters. Looking at column DE, for instance, we can see that Germany has 20 regions in the top level clusters (three regions in cluster 1, which is the first in the ranking, and 17 regions in cluster 8) and 18 regions in cluster 2, which belongs to the middle level. To provide another example of how to read Figure 1.2, it may be said that cluster 1 is composed of regions from Austria, Belgium, Denmark, Finland, France, Germany, the Netherlands, Sweden, and the United Kingdom, while cluster 3 (the last in the ranking) contains regions from Bulgaria, Poland, Romania, and Spain.

The second part of Figure 1.2 shows that the only country that has all regions in top-performing clusters is Austria. Moreover, Belgium, Denmark, Finland, Germany, Ireland, the Netherlands, and Sweden have the majority of regions in top-level clusters; on the other hand, more than half of their regions of Bulgaria, Croatia, Greece, Hungary, Poland, Romania, Slovakia and Spain are in low-level clusters. The complete list of all regions and their respective cluster and level of performance is available in Appendix 1.B. For each region it is possible to identify the other regions that are grouped in the same cluster as well as the position of the cluster in the ranking.

1.4.4 Robustness

To validate the ranking obtained with 11 clusters, we now repeat the analysis considering just eight clusters, which represents the minimum number of clusters that makes it possible to reduce the 'within group sum of squares' and at the same time avoids the generation of many incomparable clusters in the Hasse diagram. In this case, the final data matrix is of size 8x8 and the most impacting indicators obtained from the attribute-related sensitivity analysis are the same eight indicators obtained in the study with 11 clusters except one: *lifelong learning* is substituted by *scientific publications among the top-10% most cited publications worldwide as percentage of total scientific publications of the country*.

Regarding the new composition of the groups of regions, clusters 7 and 1 of the analysis based on 11 clusters are now joined together into a single cluster. This is also the case for clusters 6 and 4, and for clusters 11 and 9.

The top-performing regions in the 8-cluster analysis are the same as in the 11-cluster analysis. The regions classified as low performing in the 8-cluster analysis are the same as in the 11-cluster analysis. The middle-level regions in the 8-cluster analysis are the same as in the 11-cluster analysis, except for Åland, which is now included in the low-performing regions. Given that it is a one-of-a-kind case, it is not important for the general ranking. As a result, we can conclude that the results obtained are robust with regard to the choice of the number of clusters.

To validate the whole analysis, we apply the poset-based approach to the entire dataset composed of the 220 regions and the 17 indicators (without creating clusters, and without excluding any indicator). Table 1.2 shows the regions in which we can observe the most important differences between the result of the poset analysis with clusters and the poset analysis without clusters. The complete ranking is listed in Appendix 1.C.

Table 1.2 – Regions with the most important differences between the result of the poset analysis with clusters and the poset analysis without clusters (RIS 2019)

Region	***Ranking poset (no clusters)***	***Level poset (with clusters)***
South West	12 out of 220	Middle^{+} (from 65^{th} to 78^{th})
Friuli-Venezia Giulia	22 out of 220	Middle (from 79^{th} to 139^{th})
Småland med öarna	22 out of 220	Middle (from 79^{th} to 139^{th})
Abruzzo	81 out of 220	Low (from 176^{th} to 198^{th})
East of England	81 out of 220	Top^{+} (from 1^{st} to 35^{th})
Basilicata	98 out of 220	Low (from 176^{th} to 198^{th})
Groningen	98 out of 220	Top^{+} (from 1^{st} to 35^{th})
Campania	110 out of 220	Low (from 176^{th} to 198^{th})
Cataluña	118 out of 220	Top^{-} (from 54^{th} to 64^{th})
Nordjylland	118 out of 220	Top^{+} (from 1^{st} to 35^{th})
Wielkopolskie	118 out of 220	Low^{-} (from 199^{th} to 220^{th})
Zahodna Slovenija	118 out of 220	Top^{-} (from 54^{th} to 64^{th})
Comunidad de Madrid	149 out of 220	Top^{-} (from 54^{th} to 64^{th})

Syddanmark	150 out of 220	Top$^+$ (from 1st to 35th)
Weser-Ems	185 out of 220	Middle (from 79th to 139th)
Friesland	186 out of 220	Middle (from 79th to 139th)
Mazowiecki regionalny	209 out of 220	Middle$^-$ (from 140th to 153rd)

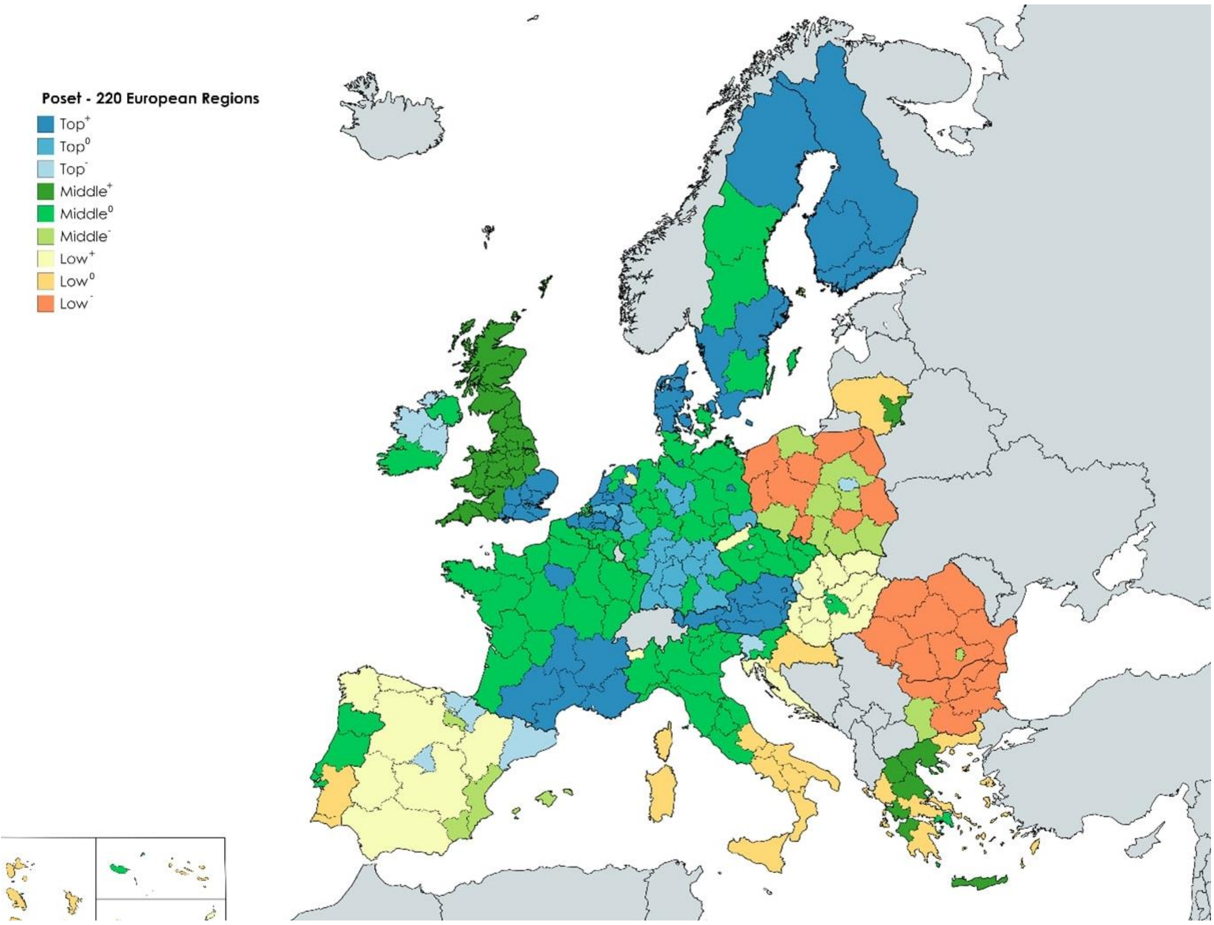
According to Table 1.2, only 17 out of 220 regions show important differences in the results between the poset analysis with clusters and the poset analysis without clusters. In particular, seven regions (four of them are Italian regions) are under-performing in the analysis with clusters (Abruzzo, Basilicata, Campania, Friuli-Venezia Giulia, Småland med öarna, South West, and Wielkopolskie), and ten regions are over-performing (Cataluña, Comunidad de Madrid, East of England, Frielsand, Groningen, Mazowiecki regionalny, Nordjylland, Syddanmark, Weser-Ems, and Zahodna Slovenija). The main reason that we can mention to explain the difference between the two analysis is the reduction of the observation in the final dataset used in the poset analysis with clusters due to the cluster analysis and the attribute-related sensitivity analysis; for instance, South West (UK) is under-performing in the poset analysis with clusters since it has the best scores in the attributes that are not included in the list of the most impacting ones; the opposite happened for Syddanmark (DK).

1.4.5 Sublevels and charts

In order to complete the analysis and display all the results in a political map of European regions, the last step is the definition of the three sublevels for each of the three performance groups already discussed. For the 64 regions in the top level, three sub-groups are identified as follows: the 35 regions in clusters 1 and 7 form the top$^+$ level, the 18 regions of cluster 8 compose the top^0 level, and finally the 11 regions of cluster 5 represent the top$^-$ level.

Regarding the middle level, it is easy to identify three subgroups, i.e. cluster 10 as middle$^+$, cluster 2 as middle^0, followed by cluster 4 as middle$^-$. In this level, we also find cluster 11 (the Finnish archipelago of Åland). Finally, concerning the low level, as it is formed by just three clusters already ranked after the first poset analysis, it is easy to assign cluster 6 as low$^+$, cluster 9 as low^0, and cluster 3 as low$^-$. In Figure 1.3, all regions are classified according to levels and sublevels.

Figure 1.3 - Final chart representing the result of the poset-based analysis of 220 European regions



In Figure 1.3, the top level is represented by regions in blue. The only country that has all its regions as top-performing is Austria. Moreover, the continental part of Finland belongs to this category as well as south-eastern England, southern France, southern Germany, some regions of north-eastern Spain, and most regions of Belgium, Denmark, Ireland, the Netherlands, and Sweden. A remarkably interesting aspect is the behaviour of the regions where the capital cities are located: 16 out of 23 of those regions belong to the top-level groups. Also, the capital cities of countries that are not considered as Innovation Leaders, such as the Czech Republic, Hungary, Poland, Slovakia, and Slovenia and are in top-level regions. The only countries that do not follow this pattern are Bulgaria, Croatia, Greece, Italy, Lithuania, Portugal, and Romania, but we have to keep in mind that for instance Lisbon, Rome and Zagreb are located in fairly large regions, whereas the majority of European capitals are in smaller regions and so they can concentrate all the assets in a high-density region.

The middle level (regions in green) is clearly represented by central Europe. In this category we can find most northern regions of several countries, such as France, Germany, Greece, Italy, Portugal, and the United Kingdom, as well as almost the entire Czech Republic and some regions of Poland. Furthermore, the capital cities of Bulgaria, Croatia, Greece, Italy, Lithuania, Portugal, and Romania belong to this level; as a result, no European capital is located in a low-level region.

The low level (from yellow, low$^+$, to red, low$^-$) is mostly composed of regions of southern Europe (most regions of Greece and Spain, southern Italy, southern Portugal) and central and eastern Europe (most regions of Bulgaria, Croatia, Hungary, Lithuania, Poland, and Romania). There are two isolated cases: one in northern Italy (Valle D'Aosta) and another one in the Netherlands (Drenthe). These two cases will be discussed in Section 1.4.6.

1.4.6 Poset-based ranking vs RIS 2019: a comparison

A comparison of the results obtained by the poset-based analysis with the outcomes obtained by the Regional Innovation Scoreboard 2019 is shown in Table 1.3.

Table 1.3 - A comparison between the composition of the performance groups in the RIS 2019 and in the poset-based analysis

Performance groups	***poset* Top level**	***poset* Middle level**	***poset* Low level**
RIS Innovation Leaders	100%	0%	0%
RIS Strong Innovators	40%	59%	1%
RIS Moderate Innovators	9%	51%	40%
RIS Modest Innovators	0%	3%	97%

Table 1.3 presents some significant findings. First, all regions in the Innovation Leaders group in the RIS 2019 belong to the top level in the poset-based analysis. Second, 99% of the regions belonging to the Strong Innovators group in the RIS 2019 are considered part of the top level or middle level in the poset-based analysis (just one region does not follow this pattern). Third, 91% of the Moderate Innovators regions of the RIS 2019 are placed in the middle or the low level in the poset-based analysis (only eight regions do not comply with this pattern). Finally, all regions except one belonging to the Modest Innovators group of the RIS 2019 are considered as low-level regions in the poset-based analysis as well. Hence, 210 out of 220 regions show the same classification (95.5% of the total). Only ten regions are ranked very differently. Among them, the most represented country is Spain, with four out of ten regions included.

As discussed above, the regions in which the capital city is located belong to the top or middle level, and in most cases show better results than the majority of the other regions in the same country. This aspect is also confirmed by looking at the ten regions mentioned above: half of them are regions in which the capital is located. Moreover, the Polish region of Mazowiecki is the region in which Warsaw is located. In addition, the Spanish region of Cataluña contains Barcelona, which is not the capital of Spain, but is a city with more than one and a half million inhabitants. Note that nine out of these ten regions improve their RIS 2019 ranking in the poset-based analysis, except for Drenthe in the Netherlands, which is classified as a Strong Innovator in the RIS 2019 but belongs to low-level cluster 6 according to the poset-based analysis. Considering the results of the poset analysis without clusters and without the reduction of the number of indicators, we can confirm the best results for the regions Bratislavský kraj, Budapest, Comunidad Foral de Navarra, País Vasco, and Warszawski stoleczny, and the worse performance for the region Drenthe, compared to the results described in the RIS 2019. However, the poset analysis without clusters seem to confirm the over-performance of the analysis with clusters for the regions Cataluña, Comunidad de Madrid, Mazowiecki regionalny, and Zahodna Slovenija, as already detected in Subsection 1.4.4.

1.5 Discussion and conclusions

The aim of this study was to provide an alternative analysis to measure the regional innovation performance of 220 European regions, starting from the data collected in the Regional Innovation Scoreboard 2019. As innovation is a complex issue, our main goal was to avoid the construction of the ranking of the analysed regions based on the simple arithmetic average of the normalised scores of the indicators and, thus, to provide a different point of view from the one suggested by the RIS 2019. First, the analysis presented shows that it is possible to adopt the poset-based approach in order to manage a large data matrix by reducing the number of objects through a cluster analysis and by considering only the indicators with the strongest impact detected through the attribute-related sensitivity analysis. Second, the poset-based approach implies that if a cluster is better ranked than another, it means that there are no indicators on which it has a lower score and, thus, that it provides a better performance. The information resulting from the cluster analysis could be interesting for stakeholders and policymakers to construct patterns of collaboration with other similar regions across Europe. In fact, the 220 European regions were grouped in 11 clusters, and the results revealed that the Innovation Leaders are regions located mostly in central and northern Europe, whereas the low-performing regions are located mainly in southern and eastern Europe. To facilitate an in-depth discussion of the results, we created nine different categories of outcomes (three for each performance level), identifying more detailed similarities among different European regions and clusters.

The attribute-related sensitivity analysis made it possible to detect the attributes with the strongest impact for categories with more than two indicators, namely ‘framework conditions’ and ‘innovation activities,’ which are *population aged 30-34 with tertiary education* and *lifelong learning* for the former, and *innovative SMEs collaborating with others* and *design applications* for the latter. Policymakers can therefore concentrate just on specific indicators in order to improve the ranking of the regions.

The innovation leaders identified in this analysis are the 35 regions classified as top^{+}, which is the combination of clusters 1 and 7, at the top of the ranking. They include two regions from Belgium (Région de Bruxelles Capitale and Vlaams Gewest), four from Denmark (Hovedstaden, Midtjylland, Nordjylland and Syddanmark), three from Germany (Berlin, Bremen and Hamburg), four from France (Auvergne–Rhône Alpes, Île de France, Languedoc-Roussillon–Midi-Pyrénées, Provence–Alpes–Côte d'Azur), seven from the Netherlands (Gelderland, Groningen, Limburg, Noord-Holland, Overijssel, Utrecht, Zuid-Holland), all the Austrian regions, all the Finnish regions (except for the one-of-a-kind archipelago of Åland), five regions from Sweden (Östra Mellansverige, Övre Norrland, Stockholm, Sydsverige, Västsverige) and three from the UK (East of England, London, and the South East).

Finally, we compared the results of the poset-based analysis with the four performance categories presented in the RIS 2019 and it was possible to identify similarities: top regions in the poset-based analysis are either Innovation Leaders or Strong Innovators in the RIS 2019; middle-level regions in our analysis are either Strong or Moderate Innovators in the RIS 2019. Last, the low-level regions in the poset-based analysis are either Moderate or Modest Innovators in the RIS 2019.

Only 10 regions are ranked very differently in the poset-based analysis compared to the RIS 2019: the majority are regions in which the capital is located (such as Bratislava, Budapest, Ljubljana, Madrid, and Warsaw) and are better ranked in the poset-based analysis compared to the RIS 2019. The only region that is worse ranked in the poset-based analysis is Drenthe (the Netherlands), mainly due to the fact that it is low performing on the indicators that have the strongest impact.

As the analysis included 220 European regions, it is not possible to use the results to establish a ranking of the regions within the same country, which could be of interest for policymakers.

Hence, it could be interesting for future research to consider only the regions of a particular country adopting the approach discussed in this study. Another interesting investigation would be to perform the poset-based analysis at a country level by using the available data of the European Innovation Scoreboard. The results of the analysis at the regional level presented in this paper could be compared with the results of the analysis using national data of the 23 countries, which could be conducted adopting the method outlined in this study. It would be interesting to find further similarities.

Appendix 1.A

To better understand the characteristics of the theory of partially ordered sets, we provide a simple example as a guide for the analysis performed.

Consider four given objects a, b, c and d, and two attributes q_1 and q_2, as described in Table 1.A.1. We will call the set of objects X, and the set of attributes A. In the table below, we provide an example in which we consider two numerical attributes in which the higher the score, the better the outcome; however, in poset theory, attributes are just features and they could also be linguistic descriptions (i.e. high, medium, low) or ordinal attributes.

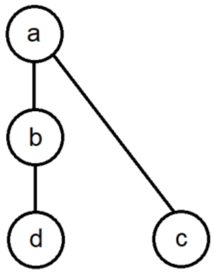
Table 1.A. 1 - Example: dataset

Objects	q_1	q_2
a	6	3
b	3	2
c	5	1
d	2	2

If we simply calculate the average of all indicators to determine the ranking, we will easily find that object a leads the ranking with a score of 4.5, followed by object c (3), and finally objects b and d (respectively with a score of 2.5 and 2). However, using the average may lead to wrong conclusions. In the poset this is avoided, since it is crucial to compare all the objects based on all attributes. Therefore, we could say that object a (6,3) is better than object b (3,2), object c (5,1) and object d (2,2) since it shows a higher score on both attributes. We could also say that object b is better than object d because even if the two objects tie on q_2 (2 for both objects b and d), object b has a higher score on q_1 compared to object d (3 for object b and 2 for object d). What is not possible to compare is object c with objects b and d: c shows a higher score on q_1 compared to both objects b and d (5 > 3 and 5 > 2), but a worse score on q_2 (1 < 2); hence, object c is comparable with object a only and incomparable with objects b and d.

Looking at the dataset, we could then write the relationships between the comparable objects: a > b > d, as well as a > c. At the same time, we know that c || b and c || d (where || is the sign to represent incomparability). The result can be also represented through a Hasse diagram, as in the figure below.

Figure 1.A. 1 - Example: Hasse diagram



Now it is possible to identify the downset and the upset of any of the objects. The downset of an object *x* consists of those objects *y* such that $y \leq x$; its cardinality is denoted as D(*x*). If $y < x$ for one or more indicators and $y > x$, then *x* and *y* are incomparable; the number of objects that are incomparable with an object *x* is denoted as I(*x*). We obtain Table 1.A.2.

Table 1.A. 2 - Example: downsets and incomparabilities of the objects, in numbers

Objects	D(*x*)	I(*x*)
a	4	0
b	2	1
c	1	2
d	1	1

In Table 1.A.2 it is possible to see, for instance, that the downset of object b consists of two objects (objects b and d).

We are now able to rank the objects of the poset. The method adopted is the so-called Local Partial Order Model (LPOM), where the 'final score' of an object is a function of D(*x*) and I(*x*). The formula to compute the 'final score' $\delta(x)$ of any object *x* is (Brüggemann & Patil, 2011):

$$\delta(x) = D(x)\,[(n + 1) / (n + 1 - I(x))] \quad (1)$$

where x is the object of interest and n indicates the total number of objects.

The number of the objects in this case is $n = 4$. For instance, the score of object a, applying the formula, is: 4 * (4 + 1) / (4 + 1 – 0) = 4 * 5 / 5 = 4. After having computed the score for all the objects, we obtain the following ranking: a, b, c, d, which is different from the ranking obtained by simply calculating the average of the indicators, which in this case yields a, c, b, d. Hence, the Hasse diagram highlights which objects are without any doubt better (or worse) than the others. With the LPOM it is possible to rank all the objects, even if some of them are incomparable.

Finally, in the poset-based analysis, it is possible to reduce the number of attributes through the so-called ‘attribute-related sensitivity analysis’. The aim is to examine how an attribute influences the position of the objects in the Hasse diagram by removing a column from the data matrix (Carayannis et al., 2018). To better understand how the attribute-related sensitivity analysis works, we could add a third attribute, namely q_3 (see Table 1.A.3) to the dataset we have analysed so far. The goal, now, is to find the pair of attributes (out of three) that permits to reproduce the original Hasse diagram of Figure 1.A.1.

Table 1.A. 3 - Example: dataset with three attributes

Objects	q_1	q_2	q_3
a	6	3	3
b	3	2	2
c	5	1	2
d	2	2	1

We first have to identify the downset of each object considering the whole data matrix (X, A). Then we compare these identified downsets with the ones of all objects (X) considering the same data matrix with the exclusion of one attribute at a time. To find, for instance, the impact of q_1, we have to look at the columns (X, A) and (X, A\{q_1}) of Table 1.A.4: for each object, we identify what are the downsets considering the two different data matrices. We can see in Table 1.A.4 that the downset of object b in (X, A) consists of two objects (b and d), but it consists of three objects in (X, A\{q1}) (objects b, c and d). The total difference in cardinality between the two data matrices (counting the number of objects that form the downsets) is 1, as indicated in the last row of Table 1.A.4. We then repeat the same exercise excluding indicators q_2 and q_3. The goal is to find the pair of attributes that allows to replicate the Hasse diagram of Figure 1.A.1.

Table 1.A. 4 - Example: attribute-related sensitivity analysis. Downsets of the objects in X for different subsets of attributes

Objects	(X, A)	(X, A\{q_1})	(X, A\{q_2})	(X, A\{q_3})
a	{a, b, c, d}	{a, b, c, d}	{a, b, c, d}	{a, b, c, d}
b	{b, d}	{b, c, d}	{b, d}	{b, d}
c	{c}	{c}	{b, c, d}	{c}
d	{d}	{d}	{d}	{d}
Total difference in cardinality		1	2	0

As shown in Table 1.A.4, q_3 has no impact on the results, while excluding attribute q_2 results in two differences; in fact, without q_2, object c is higher than both objects b and d, which is not the case in the original data matrix (in Table 1.A.4 the differences are marked in red). Finally, it is possible to conclude that the pair of attributes that best represents the original Hasse diagram is formed by q_1 and q_2, therefore if we want to simplify the data matrix, we can consider just the first two indicators.

Appendix 1.B

This appendix lists the 220 regions included in the study. Tables 1.B.1 to 1.B.4 represent the clusters of the top-performing level; Tables 1.B.5 to 1.B.8 represent clusters of the middle level; the clusters of the low- performing level are collected in Tables 1.B.9 to 1.B.11. The last column of the tables shows the performance category of the regions in the Regional Innovation Scoreboard 2019.

Table 1.B. 1 - Top⁺, Cluster n° 1 (1st position in the ranking)

Country	**Region**	**RIS 2019 classification**
Austria	Ostösterreich	Strong Innovator
	Südösterreich	Strong Innovator
	Westösterreich	Strong Innovator
Belgium	Région de Bruxelles Capitale	Innovation Leader
	Vlaams Gewest	Strong Innovator
Denmark	Nordjylland	Strong Innovator

	Syddanmark	Strong Innovator
Finland	Etelä-Suomi	Innovation Leader
	Länsi-Suomi	Innovation Leader
	Pohjois-ja Itä Suomi	Strong Innovator
France	Auvergne – Rhône Alpes	Strong Innovator
	Île de France	Strong Innovator
	Languedoc-Roussillon – Midi-Pyrénées	Strong Innovator
	Provence Alpes Côte d'Azur	Strong Innovator
Germany	Berlin	Innovation Leader
	Bremen	Strong Innovator
	Hamburg	Innovation Leader
Netherlands	Gelderland	Strong Innovator
	Groningen	Strong Innovator
	Limburg	Strong Innovator
	Noord-Holland	Innovation Leader
	Overijssel	Strong Leader
	Utrecht	Innovation Leader
	Zuid-Holland	Innovation Innovator
Sweden	Östra Mellansverige	Innovation Leader
	Övre Norrland	Strong Innovator
United Kingdom	East of England	Innovation Leader
	London	Innovation Leader
	South East	Innovation Leader

Table 1.B. 2 - Top^{+}, Cluster n° 7 (2^{nd} position in the ranking)

Country	**Region**	**RIS 2019 classification**
Denmark	Hovedstaden	Innovation Leader
	Midtjylland	Innovation Leader
Finland	Helsinki-Uusimaa	Innovation Leader
Sweden	Stockholm	Innovation Leader
	Sydsverige	Innovation Leader
	Västsverige	Innovation Leader

Table 1.B. 3 - Top^{0}, Cluster n° 8 (3rd position in the ranking, tie with cl. 5)

Country	Region	RIS 2019 classification
Germany	Braunschweig	Innovation Leader
	Darmstadt	Innovation Leader
	Dresden	Innovation Leader
	Düsseldorf	Strong Innovator
	Freiburg	Innovation Leader
	Gießen	Strong Innovator
	Hannover	Strong Innovator
	Karlsruhe	Innovation Leader
	Köln	Strong Innovator
	Mittelfranken	Innovation Leader
	Oberbayern	Innovation Leader
	Oberfranken	Strong Innovator
	Oberpfalz	Strong Innovator
	Rheinhessen-Pfalz	Innovation Leader
	Stuttgart	Innovation Leader
	Tübingen	Innovation Leader
	Unterfranken	Strong Innovator
Netherlands	Noord-Brabant	Innovation Leader

Table 1.B. 4 - Top^{-}, Cluster n° 5 (3th position in the ranking, tie with cl. 8)

Country	Region	RIS 2019 classification
Czech Republic	Praha	Strong Innovator
Hungary	Budapest	Moderate Innovator
Ireland	Eastern and Midland	Strong Innovator
	Northern and Western	Strong Innovator
Poland	Warszawski stoleczny	Moderate Innovator
Slovakia	Bratislavský kraj	Moderate Innovator
Slovenia	Zahodna Slovenija	Moderate Innovator

Spain	Cataluña	Moderate Innovator
	Comunidad de Madrid	Moderate Innovator
	Comunidad Foral de Navarra	Moderate Innovator
	País Vasco	Moderate Innovator

Table 1.B. 5 - Middle^{+}, Cluster n° 10 (5th position in the ranking)

Country	**Region**	**RIS 2019 classification**
Greece	Dytiki Ellada	Moderate Innovator
	Dytiki Makedonia	Moderate Innovator
	Kentriki Makedonia	Moderate Innovator
	Kriti	Strong Innovator
	Thessalia	Moderate Innovator
Lithuania	Sostinės regionas	Moderate Innovator
United Kingdom	East Midlands	Strong Innovator
	North East	Strong Innovator
	North West	Strong Innovator
	Scotland	Strong Innovator
	South West	Strong Innovator
	Wales	Strong Innovator
	West Midlands	Strong Innovator
	Yorkshire and The Humber	Strong Innovator

Table 1.B. 6 - Middle[0], Cluster n°2 (6th position in the ranking)

Country	Region	RIS 2019 classification
Belgium	Région Wallonne	Strong Innovator
Czech Republic	Jihovýchod	Moderate Innovator
	Jihozápad	Moderate Innovator
	Moravskoslezsko	Moderate Innovator
	Severovýchod	Moderate Innovator
	Strední Cechy	Moderate Innovator
	Strední Morava	Moderate Innovator
Denmark	Sjælland	Strong Innovator
France	Alsace Champagne Ardenne Lorraine	Strong Innovator
	Aquitaine Limousin Poitou Charentes	Strong Innovator
	Bourgogne - Franche Comté	Strong Innovator
	Bretagne	Strong Innovator
	Centre - Val de Loire	Strong Innovator
	NordPas de Calais - Picardie	Moderate Innovator
	Normandie	Moderate Innovator
	Pays de la Loire	Strong Innovator
Germany	Arnsberg	Strong Innovator
	Brandenburg	Strong Innovator
	Chemnitz	Strong Innovator
	Detmold	Strong Innovator
	Kassel	Strong Innovator
	Koblenz	Moderate Innovator
	Leipzig	Strong Innovator
	Lüneburg	Moderate Innovator
	Mecklenburg-Vorpommern	Strong Innovator
	Münster	Strong Innovator
	Niederbayern	Moderate Innovator
	Saarland	Strong Innovator
	Sachsen-Anhalt	Strong Innovator
	Schleswig-Holstein	Strong Innovator

	Schwaben	Strong Innovator
	Thüringen	Strong Innovator
	Trier	Strong Innovator
	Weser-Ems	Moderate Innovator
Greece	Attiki	Moderate Innovator
Hungary	Pest	Moderate Innovator
Ireland	Southern	Strong Innovator
Italy	Emilia-Romagna	Moderate Innovator
	Friuli-Venezia Giulia	Strong Innovator
	Lazio	Moderate Innovator
	Liguria	Moderate Innovator
	Lombardia	Moderate Innovator
	Marche	Moderate Innovator
	Piemonte	Moderate Innovator
	Provincia Autonoma Bolzano	Moderate Innovator
	Provincia Autonoma Trento	Moderate Innovator
	Toscana	Moderate Innovator
	Umbria	Moderate Innovator
	Veneto	Moderate Innovator
Netherlands	Flevoland	Strong Innovator
	Friesland	Moderate Innovator
	Zeeland	Moderate Innovator
Portugal	Centro	Strong Innovator
	Lisboa	Strong Innovator
	Norte	Strong Innovator
	Região Autónoma da Madeira	Moderate Innovator
Slovenia	Vzhodna Slovenija	Moderate Innovator
Sweden	Mellersta Norrland	Moderate Innovator
	Norra Mellansverige	Strong Innovator
	Småland med öarna	Strong Innovator
United Kingdom	Northern Ireland	Strong Innovator

Table 1.B. 7 - Middle⁻, Cluster n°11 (7th position in the ranking)

Country	Region	RIS 2019 classification
Finland	Åland	Moderate Innovator

Table 1.B. 8 - Middle⁻, Cluster n°4 (8th position in the ranking)

Country	Region	RIS 2019 classification
Bulgaria	Yugozapaden	Moderate Innovator
Poland	Dolnoslaskie	Moderate Innovator
	Lódzkie	Moderate Innovator
	Malopolskie	Moderate Innovator
	Mazowiecki regionalny	Modest Innovator
	Podkarpackie	Moderate Innovator
	Pomorskie	Moderate Innovator
	Slaskie	Moderate Innovator
Romania	Bucuresti - Ilfov	Moderate Innovator
Spain	Comunidad Valenciana	Moderate Innovator
	Illes Balears	Moderate Innovator
	La Rioja	Moderate Innovator
	Región de Murcia	Moderate Innovator

Table 1.B. 9 - Low⁺, Cluster n°6 (9th position in the ranking)

Country	Region	RIS 2019 classification
Croatia	Jadranska Hrvatska	Modest Innovator
Czech Republic	Severozápad	Moderate Innovator
Hungary	Dél-Alföld	Moderate Innovator
	Dél-Dunántúl	Moderate Innovator
	Észak-Alföld	Modest Innovator
	Észak-Magyarország	Moderate Innovator
	Közép-Dunántúl	Moderate Innovator

	Nyugat-Dunántúl	Moderate Innovator
Italy	Valle d'Aosta	Moderate Innovator
Netherlands	Drenthe	Strong Innovator
Slovakia	Stredné Slovensko	Moderate Innovator
	Východné Slovensko	Moderate Innovator
	Západné Slovensko	Moderate Innovator
Spain	Andalucía	Moderate Innovator
	Aragón	Moderate Innovator
	Canarias	Modest Innovator
	Cantabria	Moderate Innovator
	Castilla La Mancha	Modest Innovator
	Castilla y León	Moderate Innovator
	Ciudad Autónoma de Melilla	Modest Innovator
	Extremadura	Modest Innovator
	Galicia	Moderate Innovator
	Principado de Asturias	Moderate Innovator

Table 1.B. 10 - Low^{0}, Cluster n°9 (10th position in the ranking)

Country	Region	RIS 2019 classification
Croatia	Kontinentalna Hrvatska	Moderate Innovator
France	Corse	Moderate Innovator
	Régions ultrapériphériques françaises	Moderate Innovator
Greece	Anatoliki Makedonia, Thraki	Moderate Innovator
	Ionia Nisia	Moderate Innovator
	Ipeiros	Moderate Innovator
	Notio Aigaio	Modest Innovator
	Peloponnisos	Moderate Innovator
	Sterea Ellada	Moderate Innovator
	Voreio Aigaio	Moderate Innovator
Italy	Abruzzo	Moderate Innovator
	Basilicata	Moderate Innovator

	Calabria	Moderate Innovator
	Campania	Moderate Innovator
	Molise	Moderate Innovator
	Puglia	Moderate Innovator
	Sardegna	Moderate Innovator
	Sicilia	Moderate Innovator
Lithuania	Vidurio ir vakaru Lietuvos regionas	Moderate Innovator
Portugal	Alentejo	Moderate Innovator
	Algarve	Moderate Innovator
	Região Autónoma dos Açores	Moderate Innovator

Table 1.B. 11 - Low^{-}, Cluster n°3 (11th position in the ranking)

Country	**Region**	**RIS 2019 classification**
Bulgaria	Severe tsentralen	Modest Innovator
	Severoiztochen	Modest Innovator
	Severozapaden	Modest Innovator
	Yugoiztochen	Modest Innovator
	Yuzhen tsentralen	Modest Innovator
Poland	Kujawsko-Pomorskie	Modest Innovator
	Lubelskie	Modest Innovator
	Lubuskie	Modest Innovator
	Opolskie	Modest Innovator
	Podlaskie	Modest Innovator
	Swietokrzyskie	Modest Innovator
	Warminsko-Mazurskie	Modest Innovator
	Wielkopolskie	Moderate Innovator
	Zachodniopomorskie	Modest Innovator
Romania	Centru	Modest Innovator
	Nord-Est	Modest Innovator
	Nord-Vest	Modest Innovator
	Sud-Est	Modest Innovator

	Sud-Muntenia	Modest Innovator
	Sud-Vest Oltenia	Modest Innovator
	Vest	Modest Innovator
Spain	Ciudad Autónoma de Ceuta	Modest Innovator

Appendix 1.C

In this appendix, we present the analysis in which we consider all the 220 regions studied in this paper and all the 17 indicators of the Regional Innovation Scoreboard 2019. The Hasse diagram is obtained without the creation of clusters of regions. As the number of observations is very high, we have as result a messy Hasse diagram and a high number of incomparabilities. However, it is possible to obtain a ranking by applying the formula presented in the Appendix 1.A according to the Local Partial Order Method. We obtain the ranking described in Table 1.C.1.

Table 1.C. 1 - Comparison between the different rankings: poset without clusters, RIS 2019, and poset with clusters

Region	Rank Poset no clusters
Berlin	1
Helsinki-Uusimaa	2
Västsverige	3
Tübingen	4
Westösterreich	5
Karlsruhe	6
South East	6
Midtjylland	8
Südösterreich	8
Stockholm	10
Stuttgart	10
South West	12
Darmstadt	13
Utrecht	14
Auvergne - Rhône-Alpes	14
Oberbayern	16
Mittelfranken	16
Freiburg	16
Noord-Holland	16
Vlaams Gewest	16
Gießen	16
Zuid-Holland	22
Köln	22
Småland Med Öarna	22
Friuli-Venezia Giulia	22
Rheinhessen-Pfalz	26
Länsi-Suomi	26
Région De Bruxelles-Capitale / Brussels Hoofdstedelijk Gewest	26
Provence-Alpes-Côte D'azur	26
Hovedstaden	30

Östra Mellansverige	30
London	30
Gelderland	30
Languedoc-Roussillon - Midi-Pyrénées	30
Braunschweig	35
Etelä-Suomi	35
Limburg	35
Sydsverige	38
Noord-Brabant	38
West Midlands	38
Ostösterreich	38
East Midlands	38
Overijssel	38
Yorkshire And The Humber	38
Aquitaine - Limousin - Poitou-Charentes	38
Dresden	46
Île De France	46
Detmold	46
Emilia-Romagna	46
Veneto	46
Marche	46
Hamburg	52
Scotland	52
Eastern And Midland	52
Schwaben	52
Münster	52
Arnsberg	52
Lisboa	52
Norte	52
Alsace - Champagne-Ardenne - Lorraine	52
Lombardia	52
País Vasco	52
Leipzig	63
Bremen	63
Flevoland	63
Bretagne	63
Centro	63
Warszawski Stoleczny	63
Pohjois- Ja Itä-Suomi	69
Övre Norrland	69
Unterfranken	69
Oberfranken	69
North West	69
Thüringen	69
Northern And Western	69
Düsseldorf	69
Praha	69
Provincia Autonoma Trento	69

Toscana	69
Pest	69
East Of England	81
Oberpfalz	81
North East	81
Hannover	81
Région Wallonne	81
Chemnitz	81
Pays De La Loire	81
Northern Ireland	81
Bratislavský Kraj	81
Severovýchod	81
Piemonte	81
Umbria	81
Jihozápad	81
Comunidad Foral De Navarra	81
Lazio	81
Abruzzo	81
Comunidad Valenciana	81
Groningen	98
Wales	98
Trier	98
Kriti	98
Sjælland	98
Mellersta Norrland	98
Niederbayern	98
Normandie	98
Attiki	98
Budapest	98
Kentriki Makedonia	98
Basilicata	98
Southern	110
Centre - Val De Loire	110
Lüneburg	110
Moravskoslezsko	110
Malopolskie	110
Campania	110
Východné Slovensko	110
Kassel	117
Nordjylland	118
Brandenburg	118
Drenthe	118
Mecklenburg-Vorpommern	118
Bourgogne - Franche-Comté	118
Sachsen-Anhalt	118
Zahodna Slovenija	118
Sostinés Regionas	118
Cataluña	118

Strední Morava	118
Strední Cechy	118
Dytiki Makedonia	118
La Rioja	118
Kontinentalna Hrvatska	118
Wielkopolskie	118
Saarland	133
Norra Mellansverige	133
Koblenz	133
Liguria	133
Dolnoslaskie	133
Jihovýchod	138
Algarve	138
Vzhodna Slovenija	138
Região Autónoma Da Madeira	138
Provincia Autonoma Bolzano/Bozen	138
Vidurio Ir Vakaru Lietuvos Regionas	138
Puglia	138
Cantabria	138
Pomorskie	138
Molise	138
Jadranska Hrvatska	138
Comunidad De Madrid	149
Syddanmark	150
Schleswig-Holstein	150
Slaskie	150
Zeeland	153
Åland	153
Dytiki Ellada	153
Thessalia	153
Região Autónoma Dos Açores	153
Régions Ultrapériphériques Françaises	153
Sterea Ellada	153
Podkarpackie	153
Severozápad	153
Anatoliki Makedonia, Thraki	153
Sicilia	153
Západné Slovensko	153
Lódzkie	153
Illes Balears	153
Stredné Slovensko	153
Közép-Dunántúl	153
Aragón	169
Principado De Asturias	170
Nyugat-Dunántúl	171
Swietokrzyskie	171
Zachodniopomorskie	171
Región De Murcia	174

Yugozapaden	174
Nord-Pas De Calais - Picardie	176
Alentejo	176
Ipeiros	176
Ionia Nisia	176
Valle D'aosta/Vallée D'aoste	176
Peloponnisos	176
Észak-Magyarország	176
Severoiztochen	176
Vest	176
Weser-Ems	185
Friesland	186
Bucuresti - Ilfov	186
Dél-Dunántúl	186
Opolskie	189
Voreio Aigaio	190
Galicia	190
Castilla Y León	192
Sardegna	192
Észak-Alföld	194
Calabria	195
Lubelskie	195
Kujawsko-Pomorskie	195
Corse	198
Andalucía	199
Lubuskie	199
Severen Tsentralen	201
Dél-Alföld	202
Castilla-La Mancha	203
Canarias	204
Podlaskie	205
Yuzhen Tsentralen	206
Warminsko-Mazurskie	207
Extremadura	208
Mazowiecki Regionalny	209
Ciudad Autónoma De Melilla	210
Yugoiztochen	211
Ciudad Autónoma De Ceuta	212
Notio Aigaio	213
Severozapaden	214
Nord-Vest	215
Centru	216
Sud-Est	217
Sud - Muntenia	218
Nord-Est	219
Sud-Vest Oltenia	220

References

Acemoglu, D. (2012). Introduction to economic growth. *Journal of Economic Theory 147*, 545-550. doi:10.1016/j.jet.2012.01.023.

Acs, Z., Audretsch, D., Lehmann, E., & Licht, G. (2017). National systems of innovation. *The Journal of Technology Transfer 42*, 997-1008. doi:10.1007/s10961-016-9481-8.

Aghion, P., Akcigit, U., Bergeaud, A., Blundell, R., & Hemous, D. (2019). Innovation and top income inequality. *Review of Econoimc Studies 86 (1)*, 1-45. doi:10.1093/restud/rdy027.

Alaimo, L.S., Arcagni, A., Fattore, M., & Maggino, F. (2020). Synthesis of Multi-indicator Systems Over Time: a Poset-Based Approach. *Social Indicators Research (2020)*. doi: 10.1007/s11205-020-02398-5.

Asheim, B., Bugge, M., Coenen, L., & Herstad, S. (2013). What does Evolutionary Economic Geography Bring to the Table? Reconceptualising Regional Innovation Systems. CIRCLE *Working Paper 2013/05*. Retrieved from http://wp.circle.lu.se/upload/CIRCLE/workingpapers/201305_Asheim_et_al.pdf.

Badinger, H., & Reuter, W. (2015). Measurement of fiscal rules: Introducing the application of partially ordered set (POSET) theory. *Journal of Macroeconomics 43*, 108-123. doi:10.1016/j.jmacro.2014.09.005.

Bholowalia, P., & Kumar, A. (2014). EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. *International Journal of Computer Applications 105 (9)*, 17-24. Retrieved from https://bit.ly/3e9A2HU.

Bilbao-Osorio, B., & Rodríguez-Pose, A. (2004). From R&D to Innovation and Economic Growth in the EU. Growth and Change. *A Journal of Urban and Regional Policy 35 (4)*, 434-455. doi:10.1111/j.1468-2257.2004.00256.x.

Blažek, J., & Kadlec, V. (2019). Knowledge bases, R&D structure and socioeconomic and innovation performance of European regions. *Innovation: The European Journal of Social Science Research 32 (1)*, 26-47. doi:10.1080/13511610.2018.1491000.

Bock, A., Opsahl, T., George, G., & Gann, D. (2012). The Effects of Culture and Structure on Strategic Flexibility during Business Model Innovation. *Journal of Management Studies 49 (2)*, 279-305. doi:10.1111/j.1467-6486.2011.01030.x.

Borrás, S., & Jordana, J. (2016). When regional innovation policies meet policy rationales and evidence: a plea for policy analysis. *European Planning Studies 24 (12)*, 2133-2153. doi:10.1080/09654313.2016.1236074.

Bottazzi, L., & Peri, G. (2003). Innovation and spillovers in regions: Evidence from European patent data. *European Economic Review 47 (4)*, 687-710. doi:10.1016/S0014-2921(02)00307-0.

Brüggemann, R., & Patil, G. (2011). *Ranking and Prioritization for Multi-indicator Systems. Introduction to Partial Order Applications*. New York: Springer. doi:10.1007/978-1-4419-8477-7.

Capello, R., & Lenzi, C. (2013). Territorial patterns of innovation: a taxonomy of innovative regions in Europe. *The Annals of Regional Science 51 (1)*, 119-154. doi:10.1007/s00168-012-0539-8.

Capello, R., & Lenzi, C. (2019). Regional innovation evolution and economic performance. *Regional Studies 53 (9)*, 1240-1251. doi:10.1080/00343404.2018.1502421.

Caperna, G., & Boccuzzo, G. (2018). Use of Poset Theory with Big Datasets. A New Proposal Applied to the Analysis of Life Satisfaction in Italy. *Social Indicators Research 136*, 1071-1088. doi:10.1007/s11205-016-1482-3.

Carayannis, E.G., Goletsis, Y., & Grigoroudis, E. (2018). Composite innovation metrics: MCDA and the Quadruple Innovation Helix framework. *Technological Forecasting and Social Change 131*, 4-17. doi:10.1016/j.techfore.2017.03.008.

Carlsen, L. (2018). Happiness as a sustainability factor. The world happiness index: a posetic-based data analysis. *Sustainability Science 13*, 549–571. doi:10.1007/s11625-017-0482-9.

Carlsen, L., & Brüggemann, R. (2017). Partial Ordering and Metrology Analyzing Analytical Performance, in Partial Order Concepts in Applied Sciences, 49-70, doi:10.1007/978-3-319-45421-4_4.

Clauss, T. (2017). Measuring business model innovation: conceptualization, scale development, and proof of performance. *R&D Management 47 (3)*, 385-403. doi:10.1111/radm.12186.

Cooke, P., Gomez Uranga, M., & Etxebarria, G. (1997). Regional innovation systems: Institutional and organisational dimensions. *Research Policy 26 (4-5)*, 475-491. doi:10.1016/S0048-7333(97)00025-5.

Council of Competitiveness (2005). *Measuring Regional Innovation: A Guidebook for Conducting Regional Innovation Assessments*. United States of America: Council on Competitiveness. Retrieved from https://bit.ly/2NX1yhl.

De Loof, K., De Baets, B., De Meyer, H., & Brüggemann, R. (2008). A Hitchhikers Guide to Poset Ranking. *Combinatorial Chemistry & High Throughput Screening 11(9)*, 734-744. doi:10.2174/138620708786306032.

Doloreux, D., & Parto, S. (2004). Regional innovation systems: a critical review. *MERIT Working Paper*. Retrieved from https://bit.ly/382H0uI.

Doran, J., Ryan, G., Bourke, J., & Crowley, F. (2020). In-house or outsourcing skills. how best to manage for innovation. *International Journal of Innovation Management 24 (01)*, 1-25. doi:10.1142/S1363919620500103.

European Union (2014). *Regional Innovation Scoreboard 2014*. Belgium: Publications Office of the European Union. doi:10.2769/88893.

European Union (2019a). *Regional Innovation Scoreboard 2019*. Belgium: Publications Office of the European Union. doi:10.2873/89165.

European Union (2019b). *Regional Innovation Scoreboard 2019. Methodology Report*. Belgium: Publications Office of the European Union. Retrieved from http://bit.ly/3bZQ84y.

Fattore, M. (2016). Partially Ordered Sets and the Measurement of Multidimensional Ordinal Deprivation. *Social Indicators Research 128*, 835-858. doi:10.1007/s11205-015-1059-6.

Fattore M., & Arcagni A. (2014) PARSEC: An R Package for Poset-Based Evaluation of Multidimensional Poverty. In: Brüggemann R., Carlsen L., Wittmann J. (eds) Multi-indicator Systems and Modelling in Partial Order. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-8223-9_15.

Fattore, M., & Arcagni, A. (2018). A Reduced Posetic Approach to the Measurement of Multidimensional Ordinal Deprivation. *Social Indicators Research 136(3)*, 1053-1070. doi:10.1007/s11205-016-1501-4.

Fattore, M., & Arcagni, A. (2021). Posetic Tools in the Social Sciences: A Tutorial Exposition. In: Measuring and Understanding Complex Phenomena, Indicators and their Analysis in Different Scientific Fields, 219-241. doi:10.1007/978-3-030-59683-5_15.

Fattore, M., & Maggino, F. (2014). Partial Orders in Socio-economics: A Practical Challenge for Poset Theorists or a Cultural Challenge for Social Scientists? In: Multi-indicator systems and modelling in partial order. doi:10.1007/978-1-4614-8223-9_9.

Fernandes, C., Farinha, L., Ferreira, J. J., Asheim, B., & Rutten, R. (2020). Regional innovation systems: what can we learn from 25 years of scientific achievements? *Regional Studies*, 1–13. doi:10.1080/00343404.2020.1782878.

Garud, R., Tuertscher, P., & Van De Ven, A. (2013). Perspectives on Innovation Processes. *The Academy of Management Annals 7 (1)*, 775-819. doi:10.5465/19416520.2013.791066.

Hajek, P., Henriques, R., & Hajkova, V. (2014). Visualising components of regional innovation systems using self-organizing map. Evidence from European regions. *Technological Forecasting and Social Change 84*, 197-214. doi:10.1016/j.techfore.2013.07.013.

Hauser, C. M., Siller, M., Schatzer, T., Walder, J., & Tappainer, G. (2018). Measuring regional innovation: A critical inspection of the ability of single indicators to shape technological change. *Technological Forecasting and Social Change 129*, 43-55. doi:10.1016/j.techfore.2017.10.019.

Hogan, M., & Gallaher, M. (2018). Quantitative Indicators for Country-Level Innovation Ecosystems. *RTI Press*. doi:10.3768/rtipress.2018.op.0051.1805.

Isaksen, A., Tödtling, F., & Trippl, M. (2018). Innovation Policies for Regional Structural Change Combining Actor-Based and System-Based Strategies. *New Avenues for Regional Innovation Systems - Theoretical Advances, Empirical Cases and Policy Lessons*, 221-238. doi:10.1007/978-3-319-71661-9_11.

Ivaldi, E., Ciacci, A., & Soliani, R. (2020). Urban deprivation in Argentina: A POSET analysis. *Papers in Regional Science, 99(6)*, 1723-1724, doi:10.1111/pirs.12555.

Jadhav, A., Pramod, D., & Ramanathan, K. (2019). Comparison of Performance of Data Imputation Methods for Numeric Dataset. *Applied Artificial Intelligence 33 (10)*, 913-933. doi:10.1080/08839514.2019.1637138.

Jang, S., Kim, J., & von Zedtwitz, M. (2017). The importance of spatial agglomeration in product innovation: A microgeography perspective. *Journal of Business Research 78*, 143-154. doi:10.1016/j.jbusres.2017.05.017.

Kats, J. (2006). Indicators for complex innovation systems. *Research Policy 35 (7)*, 893-909. doi:10.1016/j.respol.2006.03.007.

Lau, A., & Lo, W. (2015). Regional innovation system, absorptive capacity and innovation performance: An empirical study. *Technological Forecasting & Social Change 92*, 99-114. doi:10.1016/j.techfore.2014.11.005.

Love, J., & Roper, S. (2001). Location and network effects on innovation success: evidence for UK, German and Irish manufacturing plants. *Research Policy 30 (4)*, 643-661. doi:10.1016/S0048-7333(00)00098-6.

Matras-Bolibok, A., Bolibok, P., & Kijek, T. (2017). Effectiveness of collaboration on innovation activity in EU regions. *12th European Conference on Innovation and Entrepreneurship (ECIE 2017)*. Paris, France: Loue, C.; Slimane, S.B.

McCann, P., & Ortega-Argiles, R. (2015). Smart Specialization, Regional Growth and Applications to European Union Cohesion Policy. *Regional Studies 49 (8)*, 1291-1302. doi:10.1080/00343404.2013.799769.

Min, S., Kim, J., & Sawng Y.W. (2020). The effect of innovation network size and public R&D investment on regional innovation efficiency. *Technological Forecasting and Social Change 155*, 119998. doi:10.1016/j.techfore.2020.119998.

Moser, P. (2016). Patents and Innovation in Economic History. *Annual Review of Economics 8*, 241-258. Retrieved from https://www.annualreviews.org/doi/pdf/10.1146/annurev-economics-080315-015136.

Muller, E., Héraud, J., & Zenker, A. (2017). Are innovation systems complex systems? *First Complex Systems Digital Campus World E-Conference 2015*, 167-173. doi:10.1007/978-3-319-45901-1_17.

Munda, G. (2008). *Social multi-criteria evaluation for a sustainable economy*. New York: Springer. Retrieved from https://link.springer.com/book/10.1007%2F978-3-540-73703-2.

Navarro, M., Gibaja, J., Bilbao-Osorio, B., & Aguado, R. (2009). Patterns of innovation in EU-25 regions: a typology and policy recommendations. *Environment and Planning C: Government and Policy 27*, 815-840. doi:10.1068/c0884r.

Nelson, R. (1992). National Innovation Systems: A Retrospective on a Study. *Industrial and Corporate Change 1 (2)*, 347-374. doi:10.1093/icc/1.2.347.

Őnday, Ő. (2016). National and Regional Innovation Systems, Industrial Policies and their Impacts on Firm Innovation Strategies and Performance - Economic Role of Knowledge. *International Journal of Contemporary Applied Sciences 3 (2)*, 1-35. Retrieved from http://ijcar.net/assets/pdf/Vol3-No2-February2016/01.pdf.

Ponsiglione, C., Quinto, I., & Zollo, G. (2018). Regional Innovation Systems as Complex Adaptive Systems: The Case of Lagging European Regions. *Sustainability 10 (8)*, 1-19. doi:10.3390/su10082862.

Rodriguez-Pose, A., & Crescenzi, R. (2008). Research and Development, Spillovers, Innovation Systems, and the Genesis of Regional Growth in Europe. *Regional Studies 42 (1)*, 51-67. doi:10.1080/00343400701654186.

Rondia, E., De Massis, A., & Kotlarb, J. (2019). Unlocking innovation potential: A typology of family business innovation postures and the critical role of the family system. *Journal of Family Business Strategy 10 (4)*. doi:10.1016/j.jfbs.2017.12.001.

Sarpong, O., & Teirlinck, P. (2018). The influence of functional and geographical diversity in collaboration on product innovation performance in SMEs. *The Journal of Technology Transfer 43*, 1667-1695. doi:10.1007/s10961-017-9582-z.

Schumpeter, J.A. (1935). The analysis of economic change. *Review of Economic Statistics 17 (4)*, 2-10. doi: 10.2307/1927845.

Segarra-Blasco, A., Arauzo-Carod, J., & Teruel, M. (2018). Innovation and geographical spillovers new approaches and empirical evidence. *Regional Studies 52 (5)*, 603-607. doi:10.1080/00343404.2018.1444273.

Spescha, A., & Woerter, M. (2019). Innovation and firm growth over the business cycle. *Industry and Innovation 26 (3)*, 321-347. doi:10.1080/13662716.2018.1431523.

Ter Haar, P. (2018). Measuring innovation: A state of the science review of existing approaches. *Intangible Capital 14 (3)*. doi:10.3926/ic.1254.

Tödtling, F., & Trippl, M. (2005). One size fits all?: Towards a differentiated regional innovation policy approach. *Research Policy 34 (8)*, 1203-1219. doi:10.1016/j.respol.2005.01.018.

Tsakovski, S., Astel, A., & Simeonov, V. (2010). Assessment of the water quality of a river catchment by chemometric expertise. *Journal of Chemometrics*, *24 (11–12)*, pp. 694-702. doi:10.1002/cem.1333.

Uyarra, E. (2010). What is evolutionary about “regional systems of innovation"? Implications for regional policy. *Journal of Evolutionary Economics 20*. doi:10.1007/s00191-009-0135-y.

Uyarra, E., Flanagan, K., Magro, E., Wilson, J., & Sotarauta, M. (2017). Understanding regional innovation policy dynamics: Actors, agency and learning. *Environment and Planning C: Politics and Space 35 (4)*, 559-568. doi:10.1177/2399654417705914.

Zabala-Iturriagagoitia, J., Voigt, P., Gutierrez-Gracia, A., & Jimenez-Saez, F. (2007). Regional Innovation Systems: How to Assess Performance. *Regional Studies 41 (5)*, 661-672. doi:10.1080/00343400601120270.

Chapter 2. Regional innovation in southern Europe: a poset-based analysis

Filippo Damiani[a], Silvia Muzzioli[b], Bernard De Baets[c]

Abstract

This paper examines the performance of regional innovation across the 60 southern European regions of Greece, Italy, Portugal and Spain. A poset-based analysis is carried out in two phases. The first phase establishes a ranking of the clusters in which regions are grouped to identify patterns of comparable regions. The second phase focuses on the country level, where the regions of each of the four countries are ranked into five different performance levels. The outcomes of the two phases are compared with the results described in the Regional Innovation Scoreboard 2019, with a view to providing insights for policymakers.

Keywords

Poset, European Union, Regional Innovation Scoreboard, Southern Europe, Greece, Italy, Portugal and Spain

Acknowledgments

The authors wish to thank William Bromwich for his painstaking attention to the editing of this paper.

[a] filippo.damiani@unimore.it, PhD Candidate
Marco Biagi Department of Economics, University of Modena and Reggio Emilia, Modena, Italy
[b] silvia.muzzioli@unimore.it, Associate Professor
Marco Biagi Department of Economics, University of Modena and Reggio Emilia, Modena, Italy
[c] bernard.debaets@ugent.be, Senior Full Professor
KERMIT, Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium

2.1 Introduction

Since the Great Recession of 2008-2009, inequalities among European regions have become more acute, causing a sharp increase in sovereign debt, especially in southern European countries, in particular, Portugal, Italy, Spain, and Greece (Perez & Matsaganis, 2018; Bekiros et al., 2018), where the overall unemployment rate and the youth unemployment rate have increased (Garofalo et al., 2018). The capacity for innovation of these countries has been severely impacted by the crisis, though some of them have been able to innovate even during the worst years of the recession (e.g. Donatiello & Ramella, 2017). Furthermore, the Covid-19 pandemic has made southern European economies even more fragile (Moreira et al., 2020). In this context, the analysis of regional innovation systems has become increasingly important (Rodil-Marzábal & Vence-Deza, 2020) to identify opportunities for economic growth and to secure regional resilience (Coenen et al., 2017).

The literature on regional innovation systems (RISs) has grown significantly in the last decades (Doloreaux & Porto Gomez, 2017). The interest in RISs is driven by the conceptualisation of innovation as a source of competitive advantage (Asheim et al., 2011) and by the linkages between innovation patterns and economic performance (Capello & Lenzi, 2019); arguably, innovation-driven economies can produce new jobs and new value-added products and services (Gabriel, 2019). Several authors consider innovation to be primarily determined at regional level (Doloreaux & Parto, 2004; Navarro et al., 2009; Lau & Lo, 2015), while others argue that regions still represent the basic territorial unit for organising the economy (Asheim et al., 2019).

Another relevant aspect related to innovation systems is the importance of cooperation in developing innovative processes to reduce duplication spillovers while sharing costs and risks (Nunes et al., 2013). Collaboration is widely considered a key element also to achieve higher regional innovation performance and achieve greater innovation synergy effects (Ponsiglione et al., 2018; Russell & Smorodinskaya, 2018). Moreover, interregional linkages have a positive effect on the likelihood of regions to diversify, especially for regions that showcase complementary capabilities (Balland & Boschma, 2021). For this reason, identifying peer regions is of crucial importance to create a powerful learning channel and positively influence innovation policies (Franco et al., 2020).

Among the different indices available in the literature to measure regional innovation performance, one of the most widely adopted is the Regional Innovation Scoreboard, which has been adopted by several scholars (Zabala-Iturriagagoitia et al., 2007; Arbolino et al., 2019;

Garcia-Bernabeu et al., 2020), some of whom consider it to be the most important innovation index at the regional level (Hauser et al., 2018). The ninth Regional Innovation Scoreboard (RIS, for short) was published in 2019 and provides a comparative assessment of the performance of innovation systems across 238 European regions (Hollanders et al., 2019a); the final score of the RIS is calculated as the unweighted average of the normalised scores of 17 indicators, which constitutes a severe limitation since such aggregation is subject to possible compensation effects (Carlsen, 2018). An attempt to revisit the RIS has been proposed, for instance, by Carayannis et al., 2018. In particular, these authors employ a Multiple-Criteria Decision Analysis (MCDA) approach combining the AHP and the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) methods.

In this paper, we adopt an approach borrowed from the theory of partially ordered set (theory of posets, or poset theory, for short). The poset-based approach can be used as an alternative to composite indicators and facilitates the process of ranking in an insightful manner (Brüggemann & Patil, 2011). Fuelled by powerful algorithms (De Loof et al., 2008), it has been adopted in a wide range of studies in the literature, including the calculation of new indices of stringency of fiscal rules (Badinger & Reuter, 2015), the synthetisation of multi-indicator systems over time (Alaimo et al., 2020), the evaluation of multidimensional poverty (Fattore & Arcagni, 2014), the quality assessment of river water (Tsakovski et al., 2010), and the statistical evaluation of socio-economic phenomena (Fattore et al., 2012). The main strengths of the poset-based approach can be summarised as follows: it respects the ordinal nature of data, it maintains a high standard of objectivity (reducing the need for subjective choices), and it fully exploits all the information in the dataset (Badinger & Reuter, 2015). We believe that the aforementioned characteristics enable it to capture and represent the complexity of the measurement of regional innovation performance. Through these characteristics, it is possible to identify the most impacting indicators and consider them as all relevant in the construction of the ranking.

Our aim is to apply the poset-based approach to the RIS 2019 data of the 60 regions of southern Europe to identify similarities and differences with regard to their innovation performance. The findings of our analysis are intended to be of interest for policymakers considering that regional authorities need to tailor their own place-based policies (Grillitsch & Asheim, 2018; Morrison & Doussineau, 2019) as one-size-fits-all policies are not the solution, especially for regions lagging behind (Rodríguez-Pose & Ketterer, 2020). However, grouping similar regions into clusters can help connect policies to tackle challenges in a more focussed manner (Mazzucato, 2018). The analysis consists of three steps: first, a clustering of the 60 regions of the four countries (Greece, Italy, Portugal and Spain) is performed; second, the poset-based approach is applied to establish a ranking of these clusters (first phase); third, the first two steps are repeated considering only the data of the individual countries to find even more detailed evidence (second phase). The application of the poset-based approach is feasible even on a fairly large dataset, thanks to the cluster analysis and the attribute-related sensitivity analysis proposed in this paper. This enables us to identify the indicators with the greatest impact for each of the countries analysed. We finally compare our results with the ones described in the RIS 2019.

This paper is organised as follows. Section 2.2 presents the dataset and the methods used, with particular attention to the description of the various steps of the analysis. Section 2.3 presents the results of the study. The last section is dedicated to the discussion of findings, conclusions, limitations, and perspectives for future research. Appendix 2.A outlines the Regional Innovation Scoreboard 2019. Appendix 2.B provides a data analysis example using the poset-based approach. Appendix 2.C describes the 60 regions analysed in this study, listed in alphabetical order. It also provides additional information about the clusters and the assignment of the regions to a specific performance level resulting from the two phases of our analysis and from the RIS 2019 report. This description is intended to promote a better understanding of the regions that are similar in terms of innovation performance, and a clearer idea of which indicators should be prioritised to improve the position of a region in the ranking thanks to the attribute-related sensitivity analysis.

2.2 Material and methods

In this section, we describe the dataset adopted for the study of regional innovation in southern Europe and the methods adopted in the different steps of our investigation. The analysis was carried out adopting the dataset downloaded from the Regional Innovation Scoreboard 2019 website (https://bit.ly/3cc8PAP). The scores are already normalised for all indicators. We exclude the indicator *SMEs non-R&D innovation expenditures as percentage of GDP* from the dataset for the reasons explained in Appendix 2.A. As a result, in our analysis we consider the 60 regions (NUTS[1] 2) of Greece, Italy, Portugal and Spain and 16 indicators divided into four different frameworks. Only 21 values are missing, most of which (13 out of 21) belong to the

[1] NUTS stands for the Nomenclature of Territorial Units for Statistics and it is used for referencing the subdivisions of countries.

indicator *employment in medium-high and high-tech manufacturing* and *knowledge-intensive services as percentage of total employment*.

The first step of the analysis is the imputation of the 21 missing values. To this end, we adopted the nearest neighbour imputation method, a commonly applied method (Jadhav et al., 2019). More precisely, we considered the five nearest neighbour values to compute each item of missing data. The imputation was carried out for each indicator separately. After imputation, the data matrix contains 960 observations.

The application of the poset-based approach to a large dataset could generate results that are difficult to interpret. With this in mind, we reduced the number of objects (60 regions) through a cluster analysis by performing a hierarchical clustering with the default distance measure, namely the Euclidean distance measure; the function used is "hclust" with the complete linkage method (using the software R). The scores of the clusters correspond (for each attribute) to the average of the scores of the regions that compose each cluster. The number of attributes (indicators) is then reduced to two, for each of the four categories, through the attribute-related sensitivity method (see Appendix 2.B, Table 2.B.3). After the reduction of both the number of objects and attributes, the poset-based approach is applied to the final data matrix to create a ranking of the clusters of the 60 southern European regions (first phase of the analysis). The same procedure is then applied considering just one country at a time to detect even more differences among the regions of the same country (second phase of the analysis). This country-level focus can be considered as a robustness analysis to validate the ranking obtained in the first phase.

In the last step of the analysis, we provide a comparison between the performance levels of southern European regions obtained in the poset-based analysis and the performance groups described in the RIS 2019. Finally, we provide a comparison of the indicators with the greatest impact resulting from the attribute-related sensitivity analysis, which is carried out five times (first it is applied to the attributes of all 60 regions, then to the attributes of just the regions of the individual countries). We illustrate the characteristics of the poset-based approach in Appendix 2.B.

2.3 Results

In this section, we describe the results of the analysis conducted both at regional level for the 60 southern European regions, and at country level for Greece, Italy, Portugal and Spain. Subsections 2.3.1 and 2.3.2 contain the results for the cluster and the attribute-related sensitivity analyses. Subsections 2.3.3 and 2.3.4 contain the results of the poset-based analysis at regional and country level, respectively.

2.3.1 Cluster analysis

As explained in the previous section, after the imputation of the missing values, the dataset consists of the 60 regions of southern European countries and includes 16 indicators. As this is too large a dataset to be analysed with the basic poset-based approach, it is necessary to create clusters of regions.

The first step is the computation of the distance matrix, showing for each pair of objects (regions) their Euclidean distance considering all the indicators. The clusters are then created based on the distance matrix according to the complete linkage method.

We chose a number of clusters (k) equal to nine to be able to reduce the 'within group sum of squares' and at the same time obtain a sufficiently rich partial order, by which we can identify five different performance levels after the data analysis. As a result, the matrix consists of nine rows (clusters of regions) and 16 columns (indicators). The scores of a cluster are the averages of the scores, between 0 and 1 (normalised values), of the regions that compose the cluster.

According to the cluster analysis performed, the number of regions included in the different clusters is not homogeneous. In particular, we observe, on the one hand, larger clusters consisting of ten (clusters 4 and 8) or more regions (cluster 9), and on the other hand, one cluster that includes just one region (cluster 7).[2] Hence, the first result is that the Spanish autonomous city of Ceuta (located on the coast of north Africa) shows data that are incomparable with all other regions included in the dataset, and, with $k = 9$, it is impossible to include it in any cluster.

[2] As a number of clusters contain several regions, the variability inside these clusters is likely to be quite high. As a result, some regions might be considered as outliers of such clusters, as in the case of Dytiki Makedonia (cluster 1), Kriti (cluster 2), Notio Aigaio (cluster 3), and Região Autónoma da Madeira and Valle d'Aosta (cluster 9).

2.3.2 Attribute-related sensitivity analysis

In this step of the analysis, we aim to select the two indicators with the greatest impact, that is, the two most impacting factors for each of the four categories to reduce the number of indicators from 16 to 8. Since the two categories 'investments' and 'impacts' consist of two indicators each, it is not necessary to perform any reduction for them. As a result, we apply the attribute-related sensitivity analysis to the two remaining categories. We reduce the four indicators of the category 'framework conditions' and the eight indicators of the category 'innovation activities'.

Starting with 'framework conditions', we consider a data matrix consisting of the nine clusters as objects and the four indicators of the category under analysis. After obtaining the Hasse diagram representing the relations between the clusters for this category, it is important to compute the total number of incomparabilities as an estimate of the complexity of the poset, and then find the pair of indicators that reproduces the closest number of incomparabilities. There are 22 incomparabilities in the Hasse diagram generated considering all four attributes of the category. The indicators *percentage of population aged 30-34 with tertiary education* and alone create 15 incomparabilities (68% of the total); thus, since they are the ones with the greatest impact for the category, they will be considered in the final data matrix.

Regarding the category 'Innovation activities', there are eight indicators. As a result, the number of possible pairwise combinations is quite high. In this case, there are 31 incomparabilities and the pair of indicators with the strongest impact is formed by *SMEs introducing marketing or organisational innovations as percentage of SMEs* and *European design applications per billion GDP in PPS,* representing 24 incomparabilities (77% of the total).

The final data matrix in Table 2.1 shows the nine clusters and eight indicators, representing the two with the strongest impact for each category and listed as follows (the indicators are identified as indicated in Appendix 2.A: 1. Framework conditions: 1a. *percentage of population aged 30-34 with tertiary education*; 1d. *top-10% most cited publications worldwide as percentage of total scientific publications of the country*. 2. Investments: 2a. *R&D expenditure in public sector as percentage of GDP*; 2b. *R&D expenditure in business sector as percentage of GDP*. 3. Innovation activities: 3b. *SMEs introducing marketing or organisational innovations as percentage of SMEs*; 3h. *European design applications per billion GDP in PPS*. 4. Impacts: 4a. *Employment in medium-high and high-tech manufacturing and knowledge-*

intensive services; 4b. *SMEs sales of new-to-market and new-to-firm innovations as percentage of total turnover*.

Table 2. 1 - Final data matrix: nine clusters and eight indicators with the greatest impact (data normalised)

Cluster	*1a*	*1d*	*2a*	*2b*	*3b*	*3h*	*4a*	*4b*
1	0.349	0.356	0.389	0.140	0.544	0.141	0.184	0.669
2	0.492	0.440	0.601	0.274	0.664	0.247	0.283	0.683
3	0.339	0.335	0.418	0.102	0.602	0.118	0.158	0.342
4	0.469	0.376	0.407	0.241	0.300	0.206	0.271	0.556
5	0.650	0.486	0.513	0.523	0.371	0.316	0.641	0.667
6	0.440	0.421	0.438	0.256	0.313	0.570	0.257	0.547
7	0.226	0.000	0.104	0.007	0.105	0.388	0.215	0.384
8	0.285	0.505	0.491	0.465	0.511	0.697	0.505	0.666
9	0.188	0.520	0.446	0.278	0.463	0.251	0.354	0.631

Source: elaborated on the results obtained from software R.

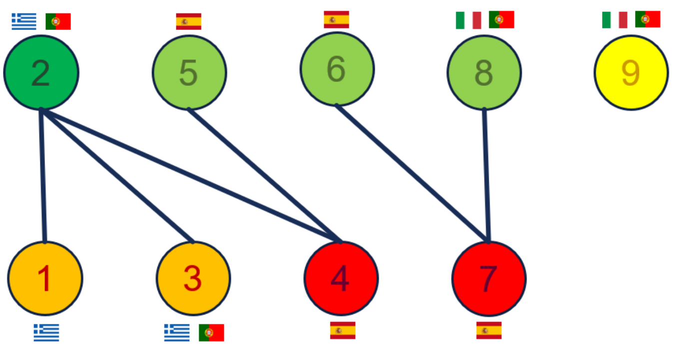
The entries for cluster 7 are just those for Ceuta since it is the only region in this sui generis cluster.

The results of the first phase of the analysis (the ranking of the clusters considering all 60 southern European regions) are presented in Subsection 2.3.3, whereas the results of the second phase (country-level analysis) are reported in Subsections 2.3.3.1 to 2.3.3.4.

2.3.3 Results of the poset-based analysis considering all the 60 southern European regions

The Hasse diagram obtained from the data matrix in Table 2.1 is shown in Figure 2.1.

Figure 2. 1 - Nine clusters (60 southern European regions), Hasse Diagram



The Hasse diagram clearly shows the relations between the clusters. In examining Figure 2.1, we can divide the clusters into five different levels according to the relations between them.

Cluster 2 (consisting of regions from Greece and Portugal) is better than three different clusters (1, 3, and 4), and we expect it to be the first cluster in the ranking; clusters 5, 6, and 8 are better than just one cluster each (cluster 5 is better than cluster 4, and clusters 6 and 8 are both better than cluster 7); cluster 9 is incomparable with all the other clusters; clusters 1 and 3 are worse than just one cluster, which is cluster 2 in both cases; finally, clusters 4 and 7 (Spanish regions only) are worse than two clusters each (cluster 4 is worse than clusters 2 and 5, and cluster 7 is worse than clusters 6 and 8). The ranking in Figure 2.2 is obtained by applying the Local Partial Order Model (LPOM).

Figure 2. 2 - Final scores of the clusters obtained by applying the LPOM

The Local Partial Order Model highlights five levels of performance: the top level, consisting of cluster 2; the middle-top level consisting of clusters 5, 6 and 8 (all of them with the same score); the middle level, consisting of cluster 9 (the cluster that is incomparable with all the others); the middle-bottom level, consisting of clusters 1 and 3 (with the same score); finally, the bottom level, consisting of clusters 4 and 7, again with the same score. More detailed results are provided in Figure 2.3, which shows the composition of each cluster and gives information about the number of regions for each country.

Figure 2. 3 - Composition of the nine clusters (60 southern European regions)

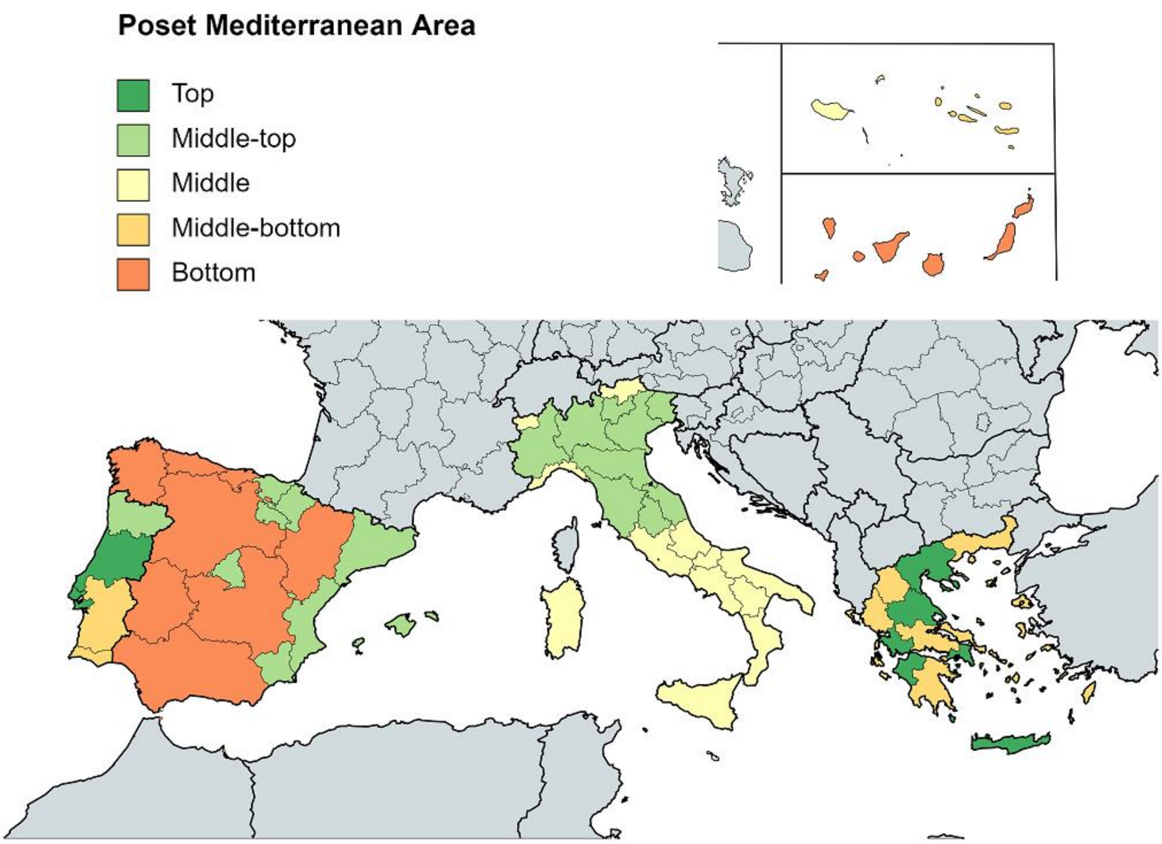
Cluster	EL	ES	IT	PT	TOTAL
2	5			2	7
5		4			4
6		4			4
8			9	1	10
9			12	1	13
1	5				5
3	3			3	6
4		10			10
7		1			1
TOTAL	**13**	**19**	**21**	**7**	**60**

	EL	ES	IT	PT	TOTAL
Top	5	0	0	2	**7**
Mid-Top	0	8	9	1	**18**
Middle	0	0	12	1	**13**
Mid-Bottom	8	0	0	3	**11**
Bottom	0	11	0	0	**11**

The left-hand panel in Figure 2.3 orders the clusters from top performing to bottom performing. As regards the colours, dark green represents the top-performing cluster (just cluster 2), light green represents the middle-top performing ones, yellow stands for the cluster in the middle level (cluster 9), orange represents the middle-bottom performing clusters, and red the bottom performing ones. The rows indicate the number of regions of each country that compose the clusters, whereas the columns indicate the clusters into which the regions of the different countries are divided. The right-hand panel of Figure 2.3 shows the number of regions (country by country) that compose each of the five levels: the number of regions is almost evenly distributed over the performance levels, with 25 regions placed in the first two levels, 22 regions forming the last two levels, and 13 in the middle. The top level is composed of five regions from Greece and two from Portugal, whereas the bottom level is composed only of Spanish regions. Moreover, Italian regions are the only ones that are neither in the top level, nor in the bottom level; in fact, nine Italian regions are in the middle-top level and the other 12 are in the middle level.

The results of the 60 regions analysed are displayed on a political map in Figure 2.4.

Figure 2. 4 - Political map representing the result of the poset-based analysis of 60 southern European regions



An examination of Figure 2.4 provides more insights. In Italy we observe a clear difference between the north and the south: the northern regions belong to the middle-top level (except for Valle d'Aosta, Provincia Autonoma di Bolzano, and Liguria), whereas the southern regions pertain to the middle level. The same pattern can be observed in Portugal: the northern regions are in the top or middle-top level, whereas the southern regions are in the middle-bottom level. More heterogeneous performances are visible in Spain and in Greece: in the case of Spain, it is evident that the best performing regions are located in the north-east and south-east of the country (including Madrid); in Greece, the top regions are equally distributed over the national territory. Last but not least, none of the regions in which the capital city is located belongs to a bottom (or middle-bottom) level cluster: Lisbon and Athens are both top-level regions, Madrid is a middle-top one, and Rome belongs to a middle-level region, namely Lazio (even if Rome is located in a large region compared to the aforementioned ones, which can concentrate all their resources in high-density areas).

If we intend to try to find more details about the differences in performance of regions of the same country, it is necessary to repeat the analysis considering just one country at a time. To do so, we performed a cluster analysis for each country to identify the number of clusters enabling us to obtain a sufficiently rich partial order to identify five performance levels for each country, as in the first phase of the analysis. In some cases, we have more clusters than performance levels (as in the first phase) since two or more clusters could have the same score, meaning that they will be assigned to the same performance level. At the same time, the attribute-related sensitivity analysis is also carried out for each country to find the indicators with the greatest impact. An interesting comparison of the different analyses to identify the most impacting indicators is presented in subsection 2.3.4. The country-level focus (presented in the following subsections) also serves as a robustness analysis.[3] In fact, the aim of the second phase of the analysis is not only to obtain more detailed results for each country, but also to check the consistency of the rankings obtained in the second phase with the ones obtained in the first phase, which means that, despite changes in the indicators with the greatest impact and the clusters in which regions are grouped, the order of regions in the ranking is not inverted.

[3] The robustness analysis has to be seen as a check of the order of the regions obtained in the first phase compared with the order of the same regions in the second phase. The comparison between the level of a given region in the two phases is not significant since the levels are constructed considering different scales in the two phases (60 regions in the first phase and just the regions of the country analysed in the second phase).

2.3.3.1 – A focus on Greece

The 13 regions of Greece were placed in three different clusters in the first phase of the analysis distributed in just two performance levels: five regions in the top level, and eight regions in the middle-bottom one. To obtain five performance levels the only possibility is to group the Greek regions into six clusters in the country-level analysis. The largest cluster consists of four regions (Attiki, Dytiki Ellada, Kentriki Makedonia, and Thessalia), whereas there is one cluster consisting of just one region (Kriti). The results of the second phase of the analysis for Greece are shown in Table 2.2, along with the results for the same regions in the first phase.

Table 2. 2 - Results for Greek regions in the two phases of the analysis

Region	First phase (level)	Second phase (level)
Kriti	Top	Top
Attiki	Top	Middle-top
Kentriki Makedonia	Top	Middle-top
Dytiki Ellada	Top	Middle-top
Thessalia	Top	Middle-top
Dytiki Makedonia	Middle-bottom	Middle
Ipeiros	Middle-bottom	Middle
Voreio Aigaio	Middle-bottom	Middle
Notio Aigaio	Middle-bottom	Middle-bottom
Sterea Ellada	Middle-bottom	Bottom
Ionia Nisia	Middle-bottom	Bottom
Anatoliki Makedonia, Thraki	Middle-bottom	Bottom
Peloponnisis	Middle-bottom	Bottom

Table 2.2 clearly shows the differences for the Greek regions between the first and the second phase. Comparing the two phases of the analysis, it is clear that the five regions that were placed in the same top-level cluster in the first phase are now divided into two clusters: Kriti is better ranked than the other four regions. Regarding the eight regions that in the first phase were at the middle-bottom level, they are now divided into three different performance levels in the country-level analysis: Dytiki Makedonia, Ipeiros, and Voreio Aigaio (middle level) are better ranked than Notio Aiagio (middle-bottom level), Sterea Ellada, Ionia Nisia, Anatoloki

Makedonia-Thraki, and Peloponnisis (bottom level). The results of the country-level analysis show for all regions of Greece that the ranking is consistent with the ranking obtained in the first phase, and more details about the relations between regions are found.

2.3.3.2 – A focus on Italy

To obtain five different levels of performance of the 21 Italian regions, it is necessary to divide them into eight clusters. The largest cluster consists of five regions (Abruzzo, Basilicata, Campania, Molise, and Puglia), whereas two clusters consist of just one region each (Bolzano and Valle d'Aosta), which are incomparable with all other clusters. The results of the second phase of the analysis for Italy are shown in Table 2.3, as well as the results for the same regions in the first phase.

Table 2. 3 - Results of Italian regions in the two phases of the analysis

Region	First phase (level)	Second phase (level)
Friuli-Venezia Giulia	Middle-top	Top
Provincia Autonoma Trento	Middle-top	Top
Toscana	Middle-top	Top
Lazio	Middle	Middle-top
Liguria	Middle	Middle-top
Emilia-Romagna	Middle-top	Middle
Lombardia	Middle-top	Middle
Marche	Middle-top	Middle
Piemonte	Middle-top	Middle
Umbria	Middle-top	Middle
Veneto	Middle-top	Middle
Provincia Autonoma Bolzano	Middle	Middle
Valle d'Aosta	Middle	Middle
Abruzzo	Middle	Middle-bottom
Basilicata	Middle	Middle-bottom
Campania	Middle	Middle-bottom

Molise	Middle	Middle-bottom
Puglia	Middle	Middle-bottom
Calabria	Middle	Bottom
Sardegna	Middle	Bottom
Sicilia	Middle	Bottom

In the first phase of the analysis, the Italian regions were placed in just two clusters: nine regions were in a cluster at the middle-top level and the other 12 regions in a cluster positioned at the middle level. The second phase of the analysis reveals more details about the differences in the performance of the Italian regions. On the one hand, the nine regions in the middle-top cluster of the first phase of the analysis are now divided into two different performance levels: Friuli-Venezia Giulia, Trento, and Toscana constitute the top level of the analysis at the country level, whereas Emilia-Romagna, Lombardia, Marche, Piemonte, Umbria, and Veneto are placed in the middle level. On the other hand, the 12 regions belonging to the middle level in the first phase are now divided into three different performance levels. In fact, Lazio and Liguria are now in the middle-top level, whereas the eight regions of the south are in two different clusters: Abruzzo, Basilicata, Campania, Molise, and Puglia are in a middle-bottom level cluster, and are ranked in a better position than Calabria, Sardegna, and Sicilia, which are placed in a bottom-level performance cluster. The results of the country-level analysis for the Italian regions demonstrate that the ranking obtained in the first phase is consistent for 19 regions out of 21. Liguria and Lazio were at a lower level than nine regions in the first phase of the analysis, whereas in the second phase they are ranked in a lower position than only three other regions, surpassing the regions of Emilia-Romagna, Lombardia, Piemonte (together these three regions form a cluster in the country-level analysis), Marche, Umbria, and Veneto (another cluster of three regions in the second phase of the analysis). These last two clusters are strongly penalised in the second phase due to a modest performance for just one indicator, namely *R&D expenditure in public sector as percentage of GDP*, which makes these clusters incomparable with all the others, even if for all the other indicators these two clusters are better performing than most of the other Italian clusters. This confirms that in the poset-based analysis it is not sufficient to obtain a ‘good mean’ score, but that it is fundamental not to have a low performance score on any indicator to avoid being downgraded in the ranking.

2.3.3.3 A focus on Portugal

Since Portugal has just seven regions, it is not necessary to conduct a cluster analysis to obtain five different performance levels. Table 2.4 shows the results of the analysis of the second phase (regarding Portugal), compared with the results of the same regions according to the analysis conducted in the first phase.

Table 2. 4 - Results of Portuguese regions in the two phases of the analysis

Region	First phase (level)	Second phase (level)
Lisboa	Top	Top
Centro	Top	Middle-top
Norte	Middle-top	Middle
Região Autónoma da Madeira	Middle	Middle
Algarve	Middle-bottom	Middle
Região Autónoma dos Açores	Middle-bottom	Middle-bottom
Alentejo	Middle-bottom	Bottom

In the first phase of the analysis, the regions of Portugal belonged to four different clusters (as shown in Figure 2.3) positioned on four different levels of the ranking (from top to middle-bottom). In the analysis with all countries, Lisbon and Centro were in the same top-level cluster. However, as shown in the last column of Table 2.4, in the country-level analysis they are divided into two different performance levels: Lisbon is still at the top level, whereas Centro is now at a middle-top level. Furthermore, in the second phase, the regions of Algarve, Madeira, and Norte are incomparable among themselves and with all the other Portuguese regions; hence, they constitute the middle level. Furthermore, Algarve, together with the region of the Açores and Alentejo, were all placed in the same middle-bottom level cluster in the analysis with all countries, and it was not possible to establish a clear relation between them. Thanks to the country-level analysis, it may be seen that Algarve is better ranked than the region of the Açores, which is better ranked than Alentejo. The results of the second phase, concerning Portuguese regions, show that the ranking is consistent with the ranking obtained in the first phase. Moreover, we are now able to identify relations among those regions that in the first phase were placed in the same clusters. However, if we consider the results of the second phase, it seems that Norte, Madeira, and Algarve are at the same performance level. In this case, the combination between the first and the second phases could be helpful in understanding the ranking among these regions.

2.3.3.4 A focus on Spain

The 19 regions of Spain were placed in four different clusters in the first phase of the analysis. To obtain five performance levels, it is sufficient to divide the Spanish regions into five clusters in the country-level analysis. This means that the clusters show a clear pattern considering the indicators with the greatest impact: each cluster corresponds to one performance level. The results of the second phase of the analysis for Spain are shown in Table 2.5, along with the results for the same regions in the first phase.

Table 2. 5 - Results for the Spanish regions in the two phases of the analysis

Region	First phase (level)	Second phase (level)
Cataluña	Middle-top	Top
Comunidad de Madrid	Middle-top	Top
Comunidad Foral de Navarra	Middle-top	Top
País Vasco	Middle-top	Top
Comunidad Valenciana	Middle-top	Middle-top
Islas Baleares	Middle-top	Middle-top
La Rioja	Middle-top	Middle-top
Murcia	Middle-top	Middle-top
Aragón	Bottom	Middle
Asturias	Bottom	Middle
Cantabria	Bottom	Middle
Galicia	Bottom	Middle
Andalucía	Bottom	Middle-bottom
Canarias	Bottom	Middle-bottom
Castilla-la Mancha	Bottom	Middle-bottom
Castilla y León	Bottom	Middle-bottom
Extremadura	Bottom	Middle-bottom

Ciudad Autónoma de Melilla	Bottom	Middle-bottom
Ciudad Autónoma de Ceuta	Bottom	Bottom

As in the case of the Greek regions, for the Spanish regions the differences between the first and the second phase are evident, as can be seen in Table 2.5. In the first phase of the analysis the 19 regions were grouped into four different clusters belonging to just two performance levels: eight regions in the middle-top level, and 11 regions in the bottom level. In the second phase more insights are available. The eight regions in the middle-top level in the first phase of the analysis are now divided into two different groups: Cataluña, Madrid, Navarra, and País Vasco compose the top level in Spain and are ranked higher than Comunidad Valenciana, Islas Baleares, La Rioja, and Murcia, which form the middle-top level in the second phase. Furthermore, the 11 regions grouped in the bottom level in the first phase, are divided into three different performance levels in the country-level analysis: Aragón, Asturias, Cantabria, and Galicia (middle level) rank higher than Andalucía, Canarias, Castilla-la Mancha, Castilla y León, Extremadura, Melilla (middle-bottom level), and Ceuta (bottom level). The results of the country-level analysis also show that for Spain the ranking is consistent with the ranking obtained in the first phase. Moreover, in line with the results for Greece, Italy, and Portugal, for the Spanish regions we are able to identify relations between those regions that in the first phase were placed in the same clusters or performance level.

2.3.4 Indicators with the greatest impact

During the analysis conducted in this paper in the application of the poset-based approach we adopted many times the attribute-related sensitivity analysis to reduce the number of indicators of the first category (framework conditions) and the third one (innovation activities). In Tables 2.6 and 2.7 we show respectively the indicators identified as having the greatest impact in the different analyses for the two categories mentioned above. Table 2.6 represents the indicators of the first category and Table 2.7 the indicators of the second category; 1 signifies that the indicator has the greatest impact, otherwise it is 0. The last row of Tables 2.6 and 2.7 shows the number of cases in which the indicator has the greatest impact.

Table 2. 6 - Indicators with the greatest impact resulting from the attribute-related sensitivity analysis for the category 'framework conditions'

Type of analysis	*1a*	*1b*	*1c*	*1d*
All 60 regions (first phase)	1	0	0	1
Greece	1	0	0	1
Italy	1	0	1	1
Portugal	0	1	0	1
Spain	0	1	0	0
Importance of the indicators[4]	**3**	**2**	**1**	**4**

The indicators with the greatest impact resulting from all the analyses conducted for the first category are 1a. *percentage of population aged 30-34 with tertiary education*, and 1d. *top-10% most cited publications worldwide as percentage of total scientific publications of the country*, as shown in Table 2.6. The number of indicators with the greatest impact resulting from the attribute-related sensitivity analysis could be higher (or lower) than two, as in the case of Italy (three indicators with the greatest impact) and Spain (one). In the case of Italy there are three indicators with the greatest impact since the comparison of two pairs of indicators (with only three indicators in the two pairs) resulted in an *ex aequo*. In the case of Spain there is just one impacting indicator. Indicator *1d* is most impacting in all the analyses, except for the Spanish analysis at the country-level. Spain and Portugal are the only two countries for which the indicator *lifelong learning* has a greater impact. Finally, the indicator *international scientific co-publications per million population* is most impacting for Italy only.

Table 2. 7 - Indicators with the greatest impact resulting from the attribute-related sensitivity analysis for the category 'innovation activities'

Type of analysis	*3a*	*3b*	*3c*	*3d*	*3e*	*3f*	*3g*	*3h*
All 60 regions (first phase)	0	1	0	0	0	0	0	1
Greece	0	1	0	0	0	0	1	1
Italy	0	1	0	0	1	0	1	1
Portugal	0	1	1	0	0	0	1	0
Spain	0	0	0	0	1	0	0	1
Importance of the indicators	**0**	**4**	**1**	**0**	**2**	**0**	**3**	**4**

[4] The importance is measured as the number of times the indicator is among the most impacting ones in the five analyses. Range of the impact: 0-5.

Table 2.7 shows that the most impacting indicators resulting from all the analyses conducted for the third category are 3b. *SMEs introducing marketing or organisational innovations as percentage of SMEs*, and 3h. *European design applications per billion GDP in PPS*. Indicator 3g. *trademark applications per billion GDP in PPS* also resulted often as most impacting. In this category Italy has four indicators that are all most impacting, Portugal and Greece have three, and Spain just two, as well as two that are the most impacting indicators resulting from the first phase of the analysis with all the 60 southern European regions. There are two more indicators that, in some cases, are among the most impacting ones: indicator 3e. *public-private co-publications per million population* (for Italy and Spain), and indicator 3c. *SMEs innovating in-house as percentage of SMEs* (for Portugal).

In the next, and last, subsection we provide a comparison to highlight the similarities and differences between the results of the data analysis of the poset-based approach conducted in this study with the one described in the Regional Innovation Scoreboard 2019 report.

2.3.5 Poset-based analysis vs RIS 2019: a comparison between the two rankings

We now compare the position of each region in the five performance levels of the first phase of our analysis with the position of the same region in the five performance groups of the RIS 2019, bearing in mind that the top, middle-top, middle, middle-bottom, and bottom levels of the poset-based analysis are comparable respectively to the strong, moderate+, moderate, moderate-, and modest groups indicated in the RIS 2019 report. The results are shown in Table 2.8 below.

Table 2. 8 - A comparison between the composition of the performance groups in the RIS 2019 and in the first phase of the poset-based analysis

Performance groups	**1. Poset Top**	**2. Poset Middle-Top**	**3. Poset Middle**	**4. Poset Middle-Bottom**	**5. Poset Bottom**
1. RIS Strong	**3**	**2**	0	0	0
2. RIS Moderate+	**2**	**10**	**0**	0	0
3. RIS Moderate	2	**5**	**8**	**7**	2
4. RIS Moderate-	0	1	**5**	**3**	**4**
5. RIS Modest	0	0	0	**1**	**5**

Table 2.8 presents some significant findings. First, 29 out of 60 regions are in the same performance level both in the RIS 2019 and in the poset-based analysis. Second, 26 regions are

in adjacent performance levels in both analyses. Third, all Strong Innovator and Moderate+ Innovator regions in the RIS 2019 analysis are in the top level or in the middle-top level in the poset-based analysis. Fourth, all Modest Innovator regions in the RIS 2019 are in the middle-bottom or the bottom level in our analysis. Finally, just five regions are two levels apart in the two analyses. These five regions are: Dytiki Ellada and Thessalia, two Greek regions that are classified as Moderate Innovators in the RIS 2019 but appear in the top level in our analysis; Aragón and Cantabria, two Spanish regions that are in the Moderate Innovators group in the RIS 2019, but appear in the bottom level in our analysis; finally, Isles Baleares (Spain), considered to be a Moderate-Innovator in the RIS 2019, appears in the middle-top level in our analysis.

The regions of Dytiki Ellada and Thessalia have the lowest scores of cluster 2, which is the highest ranked in the poset-based analysis. However, they are not outliers of the cluster, and their positive results are confirmed also in the country-level analysis, in which they have lower performance scores than just one Greek region (Kriti). Furthermore, they perform well on nearly all of the most impacting indicators. Hence, with a cluster analysis performed to obtain five performance levels in the poset-based analysis, Dytiki Ellada and Thessalia should be considered positive performers in terms of innovation.

Aragón and Cantabria have the highest scores of cluster 4, which is the lowest ranked (together with cluster 7) in the poset-based analysis. However, as in the case of the two Greek regions discussed above, they cannot be considered as outliers of their cluster, even if their scores are better compared to the average of the regions of the cluster. In fact, in the country-level analysis, the 10 regions of cluster 4 of the first phase are divided into three clusters, and Aragón and Cantabria (together with Asturias and Galicia) rank higher than the other six regions of cluster 4. Hence, in our analysis, Aragón and Cantabria should be included in a bottom-level cluster, but thanks to the country-level analysis we can conclude that they rank higher than the other regions of the cluster.

Lastly, Isles Baleares is included in a cluster that performs well on the indicators with the greatest impact, both in the first phase and in the country-level analysis, even if the mean of all indicators is not particularly high for the region, nor for the cluster. In particular, Isles Baleares (as well as its cluster) has one of the highest scores for the indicator *European design applications per billion GDP in PPS*, which is one with the greatest impact of the whole analysis.

2.4 Discussion and conclusions

The aim of this study was to provide a detailed analysis of the innovation performance of the 60 regions of southern Europe, providing more insights than the simple ranking based on just the unweighted average of the normalised scores, as suggested by the Regional Innovation Scoreboard. To achieve our goal, we analysed the data of the RIS 2019, which is the most recent version available of the Scoreboard, in two phases. In the first phase, we considered all the 60 regions together. In the second phase, we carried out four different country-level analyses, one for each country. The two phases consist of the same steps: the division of regions into clusters, the identification of the two most impacting indicators for each of the four categories of indicators, the computation of the ranking of the clusters of regions in order to identify five levels of performance (top, middle-top, middle, middle-bottom, and bottom level).

In the first phase, we divided the 60 southern European regions into nine clusters according to the similarity of their scores for the indicators. The information resulting from the cluster analysis is intended to be of interest for stakeholders and policymakers to design and implement forms of collaboration with similar regions across southern Europe that may have the same needs in terms of the enhancement of their innovation structures. Then we identified the two indicators with the greatest impact for each of the four categories of indicators, applying the attribute-related sensitivity analysis, according to poset theory. As two categories already consisted of two indicators, we conducted the analysis just for the most numerous categories, namely ‘framework conditions’ and ‘innovation activities’. Then we performed the poset-based analysis on the final dataset, which consists of the scores on the most impacting indicators of the nine clusters, and after the data analysis we obtained the ranking. The cluster that was found to be at the top level consists of the five Greek regions of Attiki, Dytiki Ellada, Kentriki Makedonia, Kriti, and Thessalia, and two Portuguese regions of Centro, and Lisboa. The indicators for which the top-level cluster is the best are *R&D expenditure in public sector as percentage of GDP* and *SMEs introducing marketing or organisational innovations as percentage of SMEs*. The positive results for Greece for public R&D investments also find support in the literature; in fact, Zoumpekas et al. (2021) analysing the data of the European Innovation Scoreboard from 2010 to 2017, demonstrated that Greece significantly enhanced its score for this indicator compared to the EU average. The middle-bottom and bottom levels of our analysis mostly consist of the remaining Greek regions and most regions of central and southern Spain. Regarding the Italian regions, they are placed at the middle-top level (most of northern and central Italy) and the middle level (mainly the southern regions).

The second phase of the analysis revealed more detailed information about the top-performing regions for each country. The leader in Portugal is Lisbon; the Italian top level consists of Friuli-Venezia Giulia, Provincia Autonoma di Trento, and Toscana; the leader in Greece is Kriti; four regions are top performing in Spain: Cataluña, Comunidad Valenciana, Madrid, and País Vasco. The results of the second phase also showed the ranking obtained in the first phase of the analysis to be consistent.

We then identified the most impacting indicators resulting from the combination of the five attribute-related sensitivity analyses that we conducted in our work (once in the first phase, and four times in the second phase). We found that in the category ‘framework conditions’, the indicator *top-10% most cited publications worldwide as percentage of total scientific publications of the country* was most impacting four times, followed by the indicator *percentage of population aged 30-34 with tertiary education* (three times); whereas for the category ‘innovation activities’, the indicators that resulted most impacting were *SMEs introducing marketing or organisational innovations as percentage of SMEs*, and *European design applications per billion GDP in PPS* (four times each). The attribute-related sensitivity analysis also enables us to ascertain that there are indicators that are most impacting only for some countries, as in the case of the indicator *SMEs innovating in-house as percentage of SMEs*, most impacting for Portugal, or the indicator *international scientific co-publications per million population*, most impacting for Italy. Regions can concentrate their efforts to improve the results for these indicators to improve their position in the ranking.

Finally, we compared the results of the first phase of the poset-based analysis with the performance categories presented in the RIS 2019, and we found that 29 out of 60 regions analysed are at the same performance level in both analyses, and that 26 other regions change position by just one level up or down. Three regions (Dytiki Ellada, Isles Baleares, and Thessalia), according to our analysis, improved their performance significantly; conversely, the outcomes for Aragón and Cantabria in the poset-based analysis compared well to the RIS 2019. The difference is explained by high (in the case of the first three regions) or low (in the case of the two Spanish regions) performances for the most impacting indicators compared to the average of all indicators, which is adopted by the RIS to construct the ranking.

For future research, it would be interesting to apply the poset-based analysis also to the data of the Regional Innovation Scoreboard relating to the previous years, to find the trend of innovation in southern Europe over the last decade. Furthermore, the same study could also be conducted on the national data of the European Innovation Scoreboard.

The coming years will be crucial for the whole of Europe, especially for the southern European regions, thanks to the opportunities of the Recovery Plan and of the Next Generation EU programmes. For this reason, understanding the patterns of innovation could help to relaunch the economies and to foster resilience.

Appendix 2.A Regional Innovation Scoreboard 2019

This is the ninth edition of the Regional Innovation Scoreboard (RIS) (the first one was published in 2009) and provides a comparative assessment of the performance of regional innovation systems across 238 regions of 23 EU Member States, together with Norway, Serbia, and Switzerland. The RIS is associated with the European Innovation Scoreboard (EIS), which assesses the performance of national innovation systems. The RIS assigns the European regions to four innovation performance groups: innovation leaders (i.e., regions with a relative performance greater than 120% of the EU average), strong innovators (i.e., regions with a relative performance between 90% and 120% of the EU average), moderate innovators (i.e., regions with a relative performance between 50% and 90% of the EU average), and modest innovators (i.e., regions with a relative performance below 50% of the EU average).

The RIS aims to measure innovation performance using the same 27 indicators adopted by the EIS, though regional data are not available for many indicators. As a result, the RIS assesses regions considering 17 indicators grouped into four different categories: 1. framework conditions (1a. *percentage of population aged 30-34 with tertiary education*, 1b. *lifelong learning* – the share of population aged 25-64 enrolled in education or training aimed at improving knowledge, skills and competences, 1c. *international scientific co-publications per million population*, 1d. *top-10% most cited publications worldwide as percentage of total scientific publications of the country*); 2. investments (2a. *R&D expenditures in public sector as percentage of GDP*, 2b. *R&D expenditures in business sector as percentage of GDP*, 2c. *SMEs non-R&D innovation expenditures as percentage of GDP*); 3. innovation activities (3a. *SMEs introducing product or process innovations as percentage of SMEs*, 3b. *SMEs introducing marketing or organisational innovations as percentage of SMEs*, 3c. *SMEs innovating in-house as percentage of SMEs*, 3d. *innovative SMEs collaborating with others as percentage of SMEs*, 3e. *public-private co-publications per million population*, 3f. *PCT patent applications per billion GDP in PPS*, 3g. *trademark applications per billion GDP in PPS*, 3h. *European design applications per billion GDP in PPS*); 4. impacts (4a. *employment in medium-high and high-tech manufacturing and knowledge-intensive services as percentage of total employment*, 4b. *SMEs sales of new-to-market and new-to-firm innovations as percentage of total turnover*).

Most of the data relating to the listed indicators are obtained from the Community Innovation Survey data with the help of National Statistical Offices, and from Eurostat. Other sources include the Centre for Science and Technology Studies of Leiden University, Science Metrics, and the OECD’s REGPAT database. The missing data are imputed adopting a range of techniques based on the availability of regional or national data referring to the previous years of observation. In relation to certain indicators of some regions for which data are difficult to find, there are still missing data even after imputation; however, data availability increases to approximately 99% (Hollanders et al., 2019b). The data are then normalised following the min-max procedure: the minimum normalised score is equal to 0 and the maximum normalised score is equal to 1. The final regional score is obtained as the unweighted average of the 17 indicators multiplied with a country correction factor (Hollanders et al., 2019b).

An interesting aspect of the RIS 2019 is the average score of the indicators per regional performance group. Considering the EU average equal to 100, the report of the RIS 2019 shows that 15 out of 17 indicators have the best score in the innovation leaders group and the worst score in the modest innovators group. Just two indicators follow a different pattern. The first one is the indicator related to *innovative SMEs collaborating with others*, which has a slightly higher score in the strong innovators group than in the innovation leaders group (126 vs 118). However, the difference is small, and in the moderate and modest innovator groups the score is much lower compared to the innovation leaders group. The second one is the indicator related to *non-R&D innovation expenditures*, which is the only one in which moderate innovator regions have a performance that is higher than 100% of the EU average, and the outcome of the innovation leaders group is similar to that of the modest innovators. Hence, it seems that in this context this indicator does not respect the outcomes of the innovation performance groups. As explained in the methodological report of the RIS 2019, the strong performance of both moderate and modest innovators on *non-R&D innovation expenditures* reflects the fact that in less innovative regions, it is more cost-effective for enterprises to innovate by purchasing advanced machinery and equipment, and knowledge developed elsewhere, than to invest in their own R&D activities as they are more expensive and at higher risk of failing to result in a useful product or process innovation (Hollanders et al., 2019b). The issues about the non-R&D innovation expenditures indicator have already been discussed in the literature, for instance, in Blažek & Kadlec (2019) as well as in Spescha & Woerter (2019). For this reason, we excluded this indicator from our analysis.

Table 2.A.1 shows the distribution of the 60 southern European regions over the performance groups according to the RIS 2019 report.

Table 2.A. 1 - The performance groups of the 60 regions of southern Europe, according to the RIS 2019 report

Performance Group (RIS 2019)	N° of regions
Strong Innovators	5
Moderate Innovators +	12
Moderate Innovators	24
Moderate Innovators -	13
Modest Innovators	6

Considering the regions represented in Table 2.A.1, five regions are categorised as strong innovators (Lisboa, Norte, and Centro in Portugal, Kriti in Greece, and Friuli-Venezia Giulia in Italy), six regions are categorised as modest innovators (Canarias, Castilla-la Mancha, Extremadura, Ceuta and Melilla in Spain, and Notio Aigaio – or Southern Aegean - in Greece), and all the other 49 regions are categorised as moderate innovators. Hence, the Moderate group is by far the largest one considering Greece, Italy, Portugal and Spain; however, three sub-categories are proposed for each performance group in the RIS 2019 report. By virtue of this, 12 regions are classified as "moderate innovators +", 24 regions are classified as simply "moderate innovators", and the remaining 13 regions are classified as "moderate innovators –". The same results can be shown by highlighting the number of regions of each country belonging to the different performance groups, as reported in Table 2.A.2.

Table 2.A. 2 - Performance group memberships of the southern European regions country by country, according to the RIS 2019 report

Country	Greece	Italy	Portugal	Spain
Strong	1	1	3	-
Moderate +	2	8	-	2
Moderate	6	7	4	7
Moderate -	3	5	-	5
Modest	1	-	-	5

According to the results of the RIS 2019, the 60 regions of Greece, Italy, Portugal and Spain are grouped in five different performance groups. On the one hand, Portugal and Italy have no regions classified as modest innovators; on the other hand, Spain has no regions considered as strong innovators; finally, most of the regions (24 out of 60) are in the moderate innovators group, in the middle of this ranking.

Appendix 2.B – Poset-based approach

To illustrate the characteristics of the theory of partially ordered sets, we provide a simple example as a guide for the analysis performed in this paper.

Consider four given objects a, b, c and d, three indicators q_1, q_2, and q_3, and the average of the indicators μ, as described in Table 2.B.1. We will call the set of objects X, and the set of indicators A. In the table below, we consider a scenario with three numerical indicators in which the higher the score, the better the outcome. However, in poset theory, indicators are just features and they could also be linguistic descriptions (i.e. high, medium, low, etc.) or ordinal indicators.

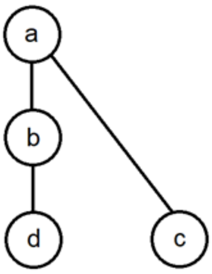
Table 2.B. 1 - Example: objects, indicators and the average of the indicators

Objects	q_1	q_2	q_3	μ
a	6	3	3	4
b	3	2	2	2.3
c	5	1	2	2.7
d	2	2	1	1.7

If we simply calculate the average of all indicators to determine the ranking, we will easily find that object a leads the ranking with a score of 4, followed by object c (2.7), and finally objects b and d (respectively with a score of 2.3 and 1.7). However, using the average may result in misleading conclusions. In the poset analysis this is avoided, since it is crucial to compare all the objects based on all attributes. Therefore, we could say that object a (6,3,3) is better than object b (3,2,2), object c (5,1,2) and object d (2,2,1) since it shows a higher score on all attributes. We could also say that object b is better than object d because even if the two objects tie on q_2 (2 for both objects b and d), object b has a higher score on both q_1 and q_3 compared to object d (3 for object b and 2 for object d on q1 and 2 for object b and 1 for object d on q_3). What is not possible to compare is object c with objects b and d: c shows a higher score on q_1 compared to both objects b and d ($5 > 3$ and $5 > 2$), but a lower score on q_2 ($1 < 2$); hence, object c is comparable with object a only, and incomparable with objects b and d.

Looking at the dataset, we could then establish the relations between the comparable objects: $a > b > d$, as well as $a > c$. At the same time, we know that $c \parallel b$ and $c \parallel d$ (where $\parallel$ is the sign to represent incomparability). The result can be also represented through a Hasse diagram, as in Figure 2.B.1.

Figure 2.B. 1 - Example: Hasse diagram



Now it is possible to identify the downset and the upset of any of the objects. The downset of an object *x* consists of those objects *y* such that *y* ≤ *x*; its cardinality is denoted as D(x). If *y* < *x* for one or more indicators and *y* > *x*, then *x* and *y* are incomparable; the number of objects that are incomparable with an object *x* is denoted as I(*x*). We obtain Table 2.B.2.

Table 2.B. 2 - Example: downsets and incomparabilities of the objects, in numbers

Objects	D(*x*)	I(*x*)
a	4	0
b	2	1
c	1	2
d	1	1

In Table 2.B.2 it is possible to see, for instance, that the downset of object b consists of two objects (objects b and d).

We are now able to rank the objects of the poset. The method adopted is the Local Partial Order Model (LPOM), where the "final score" of an object is a function of D(*x*) and I(*x*). The formula to compute the "final score" δ(x) of any object is as follows: (Brüggemann & Patil, 2011)

$$\delta(x) = \mathrm{D}(x)\,[(n+1)/\,(n+1-\mathrm{I}(x))] \qquad (1)$$

where *x* is the object of interest and n indicates the total number of objects, which, in this case, is *n* = 4. For instance, the score of object a, applying the formula, is: 4 * (4 + 1) / (4 + 1 – 0) = 4 * 5 / 5 = 4. After computing the score for all the objects, we obtain the following ranking: a, b, c, d; which is different from the ranking obtained by simply calculating the average of the indicators, which in this case yields a, c, b, d. Hence, the Hasse diagram highlights which objects are without doubt better (or worse) than the others. With the LPOM it is possible to rank all the objects, even if some of them are incomparable.

Finally, in the poset-based analysis, it is possible to reduce the number of attributes through the "attribute-related sensitivity" analysis. The aim is to examine how an attribute influences the position of the objects in the Hasse diagram by removing a column from the data matrix (Brüggemann & Patil, 2011). The goal, now, is to find the pair of attributes (out of three) that makes it possible to reproduce the original Hasse diagram of Figure 2.B.1.

We first have to identify the downset of each object considering the whole data matrix (X, A). Then we compare these identified downsets with the ones of all objects (X) considering the same data matrix with the exclusion of one attribute at a time. To find, for instance, the impact of q_1, we have to look at the columns (X, A) and (X, A\{q_1}) of Table 2.B.3: for each object, we identify what are the downsets considering the two different data matrices. We can see in Table 2.B.3 that the downset of object b in (X, A) consists of two objects (b and d), but it consists of three objects in (X, A\{q_1}) (objects b, c and d). The total difference in cardinality between the two data matrices (counting the number of objects that form the downsets) is 1, as indicated in the last row of Table 2.B.3. We then repeat the same exercise excluding indicators q_2 and q_3.

Table 2.B. 3 - Example: attribute-related sensitivity analysis. Downsets of the objects in X for different subsets of attributes

Objects	(X, A)	(X, A\{q_1})	(X, A\{q_2})	(X, A\{q_3})
a	{a, b, c, d}	{a, b, c, d}	{a, b, c, d}	{a, b, c, d}
b	{b, d}	{b, c, d}	{b, d}	{b, d}
c	{c}	{c}	{b, c, d}	{c}
d	{d}	{d}	{d}	{d}
Total difference in cardinality		1	2	0

As shown in Table 2.B.3, q_3 has no impact on the results, while excluding attribute q_2 results in two differences. In fact, without q_2, object c is higher than both objects b and d, which is not the case in the original data matrix (in Table 2.B.3 the differences are marked in red). Finally, it is possible to conclude that the pair of attributes that best represents the original Hasse diagram is formed by q_1 and q_2, therefore if we intend to simplify the data matrix, we can consider just the first two indicators.

Appendix 2.C

This appendix lists the 60 regions included in the study. All the regions are listed in Table 2.C.1 following the alphabetic order of their NUTS 2 code (first column). The second column presents the name of the region, whereas the columns from the third to the sixth show respectively the number of clusters (according to the first phase of the analysis), the level of the first phase of the analysis, the level of the second phase of the analysis (country-level focuses), and the group in the Regional Innovation Scoreboard 2019.

Table 2.C. 1 - Results of the 60 southern European regions, according to the poset-based analysis and the Regional Innovation Scoreboard 2019 report

Region	Cluster	Level (1st phase)	Level (2nd phase)	Group RIS 2019
Abruzzo	9	Middle	Middle-bottom	Moderate
Alentejo	3	Middle-bottom	Bottom	Moderate
Algarve	3	Middle-bottom	Middle	Moderate
Anatoliki Makedonia, Thraki	1	Middle-bottom	Bottom	Moderate-
Andalucía	4	Bottom	Middle-bottom	Moderate-
Aragón	4	Bottom	Middle	Moderate
Attiki	2	Top	Middle-top	Moderate+
Basilicata	9	Middle	Middle-bottom	Moderate
Calabria	9	Middle	Bottom	Moderate-
Campania	9	Middle	Middle-bottom	Moderate
Canarias	4	Bottom	Middle-bottom	Modest+
Cantabria	4	Bottom	Middle	Moderate
Castilla y León	4	Bottom	Middle-bottom	Moderate-
Castilla-la Mancha	4	Bottom	Middle-bottom	Modest+
Cataluña	5	Middle-top	Top	Moderate+
Centro	2	Top	Middle-top	Strong-
Ciudad Autónoma de Ceuta	7	Bottom	Bottom	Modest-
Ciudad Autónoma de Melilla	4	Bottom	Middle-bottom	Modest

Comunidad de Madrid	5	Middle-top	Top	Moderate
Comunidad Foral de Navarra	5	Middle-top	Top	Moderate
Comunidad Valenciana	6	Middle-top	Middle-top	Moderate
Dytiki Ellada	2	Top	Middle-top	Moderate
Dytiki Makedonia	1	Middle-bottom	Middle	Moderate
Emilia-Romagna	8	Middle-top	Middle	Moderate+
Extremadura	4	Bottom	Middle-bottom	Modest+
Friuli-Venezia Giulia	8	Middle-top	Top	Strong-
Galicia	4	Bottom	Middle	Moderate-
Ionia Nisia	1	Middle-bottom	Bottom	Moderate
Ipeiros	3	Middle-bottom	Middle	Moderate
Isles Baleares	6	Middle-top	Middle-top	Moderate-
Kentriki Makedonia	2	Top	Middle-top	Moderate+
Kriti	2	Top	Top	Strong-
La Rioja	6	Middle-top	Middle-top	Moderate
Lazio	9	Middle	Middle-top	Moderate
Liguria	9	Middle	Middle-top	Moderate
Lisboa	2	Top	Top	Strong-
Lombardia	8	Middle-top	Middle	Moderate+
Marche	8	Middle-top	Middle	Moderate+
Molise	9	Middle	Middle-bottom	Moderate-
Norte	8	Middle-top	Middle	Strong-
Notio Aigaio	3	Middle-bottom	Middle-bottom	Modest
País Vasco	5	Middle-top	Top	Moderate+
Peloponnisis	1	Middle-bottom	Bottom	Moderate-
Piemonte	8	Middle-top	Middle	Moderate+
Principado de Asturias	4	Bottom	Middle	Moderate-

Provincia Autonoma Bolzano	9	Middle	Middle	Moderate
Provincia Autonoma Trento	8	Middle-top	Top	Moderate+
Puglia	9	Middle	Middle-bottom	Moderate
Região Autónoma dos Açores	3	Middle-bottom	Middle-bottom	Moderate
Região Autónoma da Madeira	9	Middle	Middle	Moderate
Región de Murcia	6	Middle-top	Middle-top	Moderate
Sardegna	9	Middle	Bottom	Moderate-
Sicilia	9	Middle	Bottom	Moderate-
Sterea Ellada	1	Middle-bottom	Bottom	Moderate
Thessalia	2	Top	Middle-top	Moderate
Toscana	8	Middle-top	Top	Moderate+
Umbria	8	Middle-top	Middle	Moderate+
Valle d’Aosta	9	Middle	Middle	Moderate-
Veneto	8	Middle-top	Middle	Moderate+
Voreio Aigaio	3	Middle-bottom	Middle	Moderate-

References

Alaimo, L.S., Arcagni, A., Fattore, M. & Maggino, F. (2020). Synthesis of Multi-indicator System Over Time: A Poset-based Approach. *Social Indicators Research*. doi:10.1007/s11205-020-02398-5.

Arbolino, R., Boffardi, R. & De Simone, L. (2019). Which are the Factors Influencing Innovation Performances? Evidence from Italian Cohesion Policy. *Social Indicators Research 146*, 221-247. doi:10.1007/s11205-018-1904-5.

Asheim, B.T., Lawton Smith, H. & Oughton, C. (2011). Regional Innovation Systems: Theory, Empirics and Policy. *Regional Studies 45 (7)*, 875-891. doi:10.1080/00343404.2011.596701.

Asheim, B.T., Isaksen, A. & Trippl, M. (2019). Advanced Introduction to Regional Innovation Systems. Edward Elgar ISBN: 978 1 78536 197 5.

Badinger, H. & Reuter, W.H. (2015). Measurement of fiscal rules: Introducing the application of partially ordered set (POSET) theory. *Journal of Macroeconomics 43*, 108-123. doi:10.1016/j.jmacro.2014.09.005.

Balland, P.A. & Boschma, R. (2021). Complementary interregional linkages and Smart Specialisation: an empirical study on European regions. Regional Studies (online). doi:10.1080/00343404.2020.1861240. https://www.tandfonline.com/doi/full/10.1080/00343404.2020.1861240, last accessed 25 January 2021.

Bekiros, S., Hammoudeh, S., Jammazi, R. & Nguyen, D.K. (2018). Sovereign bond market dependencies and crisis transmission around the eurozone debt crisis: a dynamic copula approach. *Applied Economics 50 (47)*, 5031-5049. doi:10.1080/00036846.2018.1470313.

Blažec, J. & Kadlec, V. (2019). Knowledge bases, R&D structure and socio-economic and innovation performance of European regions. *Innovation: The European Journal of Social Science Research 32 (1)*, 26-47. doi:10.1080/13511610.2018.1491000.

Brüggemann, R. & Patil, G.P. (2011). Ranking and Prioritization for Multi-Indicator Systems. Introduction to Partial Order Applications. Springer-Verlag New York. ISBN: 978-1-4419-8476-0. doi:10.1007/978-1-4419-8477-7.

Capello, R. & Lenzi, C. (2019). Regional innovation evolution and economic performance. *Regional Studies 53 (9)*, 1240-1251. doi:10.1080/00343404.2018.1502421.

Carayannis, E.G., Goletsis, Y. & Grigoroudis, E. (2018). Composite innovation metrics: MCDA and the Quadruple Innovation Helix framework. *Technological Forecasting and Social Change 131*, 4-17. doi:10.1016/j.techfore.2017.03.008.

Carlsen, L. (2018). Happiness as a sustainability factor. The world happiness index: a posetic-based data analysis. *Sustainability Science 13*, 549-571. doi:10.1007/s11625-017-0482-9.

Coenen, L., Asheim, B., Bugge, M.M. & Herstad, S.J. (2017). Advancing regional innovation systems: What does evolutionary economic geography bring to the policy table? *Environment and Planning C: Politics and Space 35 (4)*, 600-620. doi:10.1177/0263774X16646583.

De Loof, K., De Baets, B., De Meyer, H. & Brüggemann, R. (2008). A hitchhiker's guide to poset ranking. *Combinatorial chemistry & high throughput screening 11 (9)*, 734-744. doi:10.2174/138620708786306032.

Doloreux, D. & Parto, S. (2004). Regional Innovation Systems: A Critical Review. *MERIT Working Paper*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.477.377&rep=rep1&type=pdf, last accessed 25th January 2021.

Doloreux, D. & Porto Gomez, I. (2017). A review of (almost) 20 years of regional innovation systems research. *European Planning Studies 25(3)*, 371-387. doi:10.1080/09654313.2016.1244516.

Donatiello, D. & Ramella, F. (2017). The Innovation Paradox in Southern Europe. Unexpected Performance During the Economic Crisis. *South European Society and Policies 22 (2)*, 157-177. doi:10.1080/13608746.2017.1327339.

Fattore, M. & Arcagni, A. (2014). PARSEC: An R Package for Poset-Based Evaluation of Multidimensional Poverty. In: Multi-indicator Systems and Modelling in Partial Order 317-330. Springer, New York, NY. ISBN: 978-1-4614-8222-2. doi:10.1007/978-1-4614-8223-9_15.

Fattore, M., Maggino, F. & Colombo, E. (2012). From Composite Indicators to Partial Orders: Evaluating Socio-Economic Phenomena Through Ordinal Data. In: *Quality of life in Italy*.

Social Indicators Research Series 41-68. Springer, Dordrecht. ISBN: 978-94-007-3897-3. doi:10.1007/978-94-007-3898-0_4.

Franco, S., Gianelle, C., Kleibrink, A. & Murciego, A. (2020). Learning from similar regions: how to benchmark innovation systems beyond rankings. *Chapter 7* 162-194 in Capello et al., Quantitative Methods for Place-Based Innovation Policy. doi:10.4337/9781789905519. ISBN: 9781789905502, pp 256.

Gabriel, K. (2019). Accelerating innovation in a changing world. *International Journal of Research, Innovation and Commercialisation 2 (2)*, 105. doi:10.1504/IJRIC.2019.104014.

Garcia-Bernabeu, A., Cabello, J.M. & Ruiz, F. (2020). A Multi-Criteria Reference Point Based Approach for Assessing Regional Innovation Performance in Spain. *Mathematics 8 (5)*, 797. doi:10.3390/math8050797.

Garofalo, A., Castellano, R., Punzo, G. & Musella, G. (2018). Skills and labour incomes: how unequal is Italy as part of the Southern European countries? *Quality & Quantity 52*, 1471-1500. doi:10.1007/s11135-017-0531-6.

Grillitsch, M. & Asheim, B. (2018). Place-based innovation policy for industrial diversification in regions. *European Planning Studies 26 (8)*, 1638-1662. doi:10.1080/09654313.2018.1484892.

Hauser, C., Siller, M., Schatzer, T., Walde, J. & Tappainer, G. (2018). Measuring regional innovation: A critical inspection of the ability of single indicators to shape technological change. *Technological Forecasting and Social Change 129*, 43-55. doi:10.1016/j.techfore.2017.10.019.

Hollanders, H., Es-Sadki, N. & Merkelbach, I. (2019a). Regional Innovation Scoreboard 2019. *Publications Office of the European Union*. ISBN:978-92-76-08723-6. doi:10.2873/89165.

Hollanders, H., Es-Sadki, N. & Merkelbach, I. (2019b). Regional Innovation Scoreboard 2019. Methodology Report. *European Union*. Available at https://ec.europa.eu/docsroom/documents/37783, last accessed 4 February 2021.

Jadhav, A., Pramod, D. & Ramanathan, K. (2019). Comparison of Performance of Data Imputation Methods for Numeric Dataset. *Applied Artificial Intelligence 33 (10)*, 913-933. doi:10.1080/08839514.2019.1637138.

Lau, A.K.W. & Lo, W. (2015). Regional innovation system, absorptive capacity and innovation performance: An empirical study. *Technological Forecasting and Social Change 92*, 99-114. doi:10.1016/j.techfore.2014.11.005.

Mazzucato, M. (2018). Mission-Oriented Research & Innovation in the European Union. A problem-solving approach to fuel innovation-led growth. *Publications Office of the European Union*. ISBN: 978-92-79-79918-1. doi:10.2777/36546.

Moreira, A., Léon, M., Coda Moscarola, F. & Roumpakis, A. (2020). In the eye of the storm…again! Social policy responses to COVID19 in Southern Europe. *Social & Policy Administration (online)*. doi:10.1111/spol.12681. https://onlinelibrary.wiley.com/doi/full/10.1111/spol.12681, last accessed: 25 January 2021.

Morrison, A., Doussineau, M. (2019). Regional innovation governance and place-based policies: design, implementation and implications. Regional Studies, *Regional Science 6 (1)*, 101-116. doi:10.1080/21681376.2019.1578257.

Navarro, M., Gibaja, J., Bilbao-Osorio, B. & Aguado, R. (2009). Patterns of innovation in EU-25 regions: a typology and policy recommendations. *Environment and Planning C: Government and Policy 27 (5)*, 815-840. doi:10.1068/c0884r.

Nunes, S., Carvalho, L. & Costa, T. (2013). Cooperation for innovation: evidence from southern European countries. *International Journal of Innovation and Regional Development 5 (2)*, 226-241. doi:10.1504/IJIRD.2013.055250.

Perez, S.A. & Matsaganis, M. (2018). The Political Economy of Austerity in Southern Europe. *New Political Economy 23 (2)*, 192-207. doi:10.1080/13563467.2017.1370445.

Ponsiglione, C., Quinto, I. & Zollo, G. (2018). Regional Innovation Systems as Complex Adaptive Systems: The Case of Lagging European Regions. *Sustainability 10 (8)*, 2862. doi:10.3390/su10082862.

Rodil-Marzábal, O. & Vence-Deza, X. (2020). Regional Innovation Systems and regional disparities in the Euro area: insights for regional innovation policy. *Chapter 7*, 139-161, in González-López, M. & Asheim, B.T., Regions and Innovation policies in Europe, learning form the margins. New horizons in regional sciences series. doi:10.4337/9781789904161.00012. ISBN:9781789904154.

Rodríguez-Pose, A., Ketterer, T. (2020). Institutional change and the development of lagging regions in Europe. *Regional Studies 54 (7)*, 974-986. doi:10.1080/00343404.2019.1608356.

Russell, M.G. & Smorodinskaya, N.V. (2018). Leveraging complexity for ecosystemic innovation. *Technological Forecasting and Social Change 136*, 114-131. doi:10.1016/j.techfore.2017.11.024.

Spescha, A. & Woerter, M. (2019). Innovation and firm growth over the business cycle. *Industry and Innovation 26 (3)*, 321-347. doi:10.1080/13662716.2018.1431523.

Tsakovski, S., Astel, A. & Simeonov, V. (2010). Assessment of the water quality of a river catchment by chemometric expertise. *Journal of Chemometrics 24 (11-12)*, 694,702. doi:10.1002/cem.1333.

Zabala-Iturriagagoitia, J.M., Voigt, P., Gutiérrez-Gracia, A. & Jiménez-Sáez, F. (2007). Regional Innovation Systems: How to Assess Performance. *Regional Studies 41 (5)*, 661-672. doi:10.1080/00343400601120270.

Zoumpekas, T., Vavalis, M., & Houstis, E. (2021). Analysis of innovation with data science: The case of Greece. *International Journal of Data Science and Big Data Analytics 1 (1)*, 20-42. doi:10.51483/IJDSBDA.1.1.2021.20-42.

Chapter 3: Measuring Women's Digital Inclusion. A poset-based approach to the Women in Digital Scoreboard

Filippo Damiani[i] and Paula Rodríguez-Modroño[ii]

Abstract

Women's participation in digital society is integral to achieving Agenda 2030 and an essential component in the EU strategy for digital transition. This article applies a poset-based approach to the European Women in Digital (WiD) Scoreboard, to examine women's digital inclusion in European countries. The poset methodology allows us to identify the most significant indicators for each of the dimensions that compose the WiD, considering the whole EU-28 as well as different clusters of countries, and to construct a new ranking that avoids the shortcomings of the aggregative approaches and the pre-treatment of data. Our results show that two indicators, STEM graduates and the unadjusted pay gap, are the most relevant ones in attaining women's digital inclusion. Our research contributes to better understand the dynamics and the underlying causes of women's digital inclusion in the EU-28 countries, providing a clustering of EU countries into four performance groups depending on their women's digital inclusion and contributes to the design of more targeted and effective policies for integrating gender equality in the EU digital transition strategy.

Keywords

Digital Economy, women's digital inclusion, digital transition, gender digital divide, poset

[i] filippo.damiani@unimore.it, PhD Candidate
Marco Biagi Department of Economics, University of Modena and Reggio Emilia, Modena, Italy.
PhD programme in Social Sciences, Department of Economics, Quantitative Methods and Economic History, Pablo de Olavide University, Seville, Spain.
[ii] prodmod@upo.es, Senior Lecturer
Department of Economics, Quantitative Methods and Economic History, Pablo de Olavide University, Seville, Spain.

3.1 Introduction

Recent decades have seen how digital technologies transform the world of work (JRC, 2019). Using digital technologies for professional purposes has become a prerequisite for successful integration of workers into the digitalised economy. The COVID-19 pandemic has accelerated the pace of digitalisation in our societies and economies. This digital transformation of the labour market creates both opportunities for and risks to gender equality.

Although gender differences in digital skills and use of digital devices are gradually levelling out in the EU, particularly among young people, still women are behind men in the use of various ICT technologies at work (EIGE, 2020). Studies show that gender inequalities continue to prevent women from reaching their full potential and hinder EU societies from taking full advantage of women's digital potential and current contributions (European Commission, 2018). A study from EIGE (2017) shows that closing gender gaps in STEM education would have a positive impact on employment, with total EU employment foreseen to rise from 850 thousand to 1.2 million jobs by 2050. Consequently, this would imply an increase in EU GDP per capita from 0.7% to 0.9% by 2030 and from 2.2% to 3% by 2050. The productive capacity and the competitiveness of the EU would clearly increase (Norlén, Papadimitriou & Dijkstra, 2019).

These findings confirm that gender inequalities continue to prevent women from reaching their full potential and hinder EU societies from taking full advantage of women's digital potential and current contributions (European Commission, 2018). Gender equality needs to be introduced as a primary objective in the EU strategy for digital transition, incorporating the measurement of advances in digitalization for women and men as an essential component of this strategy. The monitoring of the effectiveness of public policies governing digital transition (Bánhidi, Dobos & Nemeslaki, 2020) is even more important in the post-COVID-19 economic recovery, in which digital services are becoming a key driver of our economic growth, making the Digital Europe program an essential part of the recovery plan. At least 20% of Next Generation EU will fund investments in digital, which means, roughly, €150 billion.

However, statistical data on digital inclusion are scarce and usually not disaggregated by gender. The Women in Digital (WiD) Scoreboard, formulated in 2019, is one of the few and most recent mechanisms put in place by the European Commission to assess women's inclusion in digital jobs, careers and entrepreneurship. The WiD index, which is part of the Digital Economy and Society Index, brings together twelve relevant indicators to assess the performance of Member States in the areas of Internet use, Internet user skills as well as Specialist skills and employment. The index was constructed to obtain a general characterisation of the performance of individual Member States by observing their overall index score and the scores of the main index dimensions, to pinpoint the areas where performance could be improved and to assess progress over time (European Commission, 2020b). The WiD Scoreboard presents a ranking of countries using a simple arithmetic mean of the twelve normalised indicators.

Using the poset-based approach (poset, for short), in this study we construct a new ranking which fully exploits all information present in the dataset and reduces the need for subjective choices (Badinger & Reuter, 2015). Poset allows to obtain a ranking avoiding the use of aggregation methods (Fattore, 2016; Fattore & Arcagni, 2018; Ivaldi, Ciacci & Soliani., 2020) and without pre-treatment of data: the performance can be evaluated considering all indicators simultaneously (Carlsen & Brüggemann, 2017). Therefore, the poset methodology is useful to overcome the curse of dimensionality without using a parametric model or introducing some subjective criteria. We compare our ranking and the ranking proposed by the Women in Digital Scoreboard for 2020, identifying similarities and differences. By applying the poset-based approach, we can also identify the most significant indicators for each of the three dimensions that compose the WiD, considering both the whole EU-28 and four different macroregions. Our findings about the different significance of indicators depending on the region contribute to identifying areas where policy intervention continues to be needed and to the design of more targeted and effective policies for integrating gender equality in the EU digital transition strategy. Additionally, our analysis provides a clustering of EU countries into four performance groups depending on their level of women's digital inclusion. Although the poset methodology has been already applied to socio-economic issues (Annoni & Brüggemann, 2009; Carlsen & Brüggemann, 2016; Carlsen, 2017; Iglesias et al., 2017; Arcagni et al., 2019; Fattore & Arcagni, 2019), also related to gender discrimination (Di Bella et al., 2018; Di Brisco & Farina, 2018), it has not been applied previously to the analysis of the WiD.

This article is structured in five sections. Section 3.2 presents a literature review of women's digital inclusion within the European Union framework. Section 3.3 defines the data and methods, describing the Women in Digital (WiD) Scoreboard, its dimensions and the poset methodology. Section 3.4 presents the results of the application of the poset methodology to study the different dimensions of WiD in the whole EU-28 and in 4 macroregions in which we divide the EU. The last section presents the discussion of findings as well as the limitations of the study.

3.2 From the digital gender divide to women's digital inclusion

Research on the digital gender divide and women's digital inclusion can be segmented in three main phases. Early feminists and gender studies on the digital revolution were largely optimistic about the potential of digital technologies to empower people. Women were considered as a 'disadvantaged' group that just needed support to reach a level of ICT access like the average of the population. This first-order digital gender divide referred only to the lack of adoption or access to ICT.

However, the second wave of digital divide studies from a gender perspective detected that access to technology alone does not lead directly to more social opportunities and highlighted how digital skills acquisition and uses of the internet are also gender stratified (Castaño, Martín & Martínez, 2011; Helsper, 2010; van Deursen and van Dijk, 2019). The second-order digital divide represents the ICT usage and the proficiency of ICT usage. Technology is gendered, and digital technologies form part of the structure and performance of gender inequalities (Wajcman, 2010; Wyatt, 2008). Digitalisation holds the potential to reorganise gendered work relations since the patterns of the gender division of labour are shaped, negotiated, or affected by digitalisation (Kohlrausch & Weber, 2020). In fact, despite the measures implemented to enhance women's digital skills and to increase the participation of women in the ICT workforce, studies show that disparities on digital skills gaps by gender are still more marked at the highest levels of skills. Gender gaps in the EU are still larger in the higher and more specialized levels of skills, which are broadly considered as key factors for future digital inclusion and employment (OECD, 2018). Women are less engaged in digital technologies, information-seeking activities, content sharing or contributions to free/open collaborative platforms (Hargittai, 2010; Hargittai and Shaw, 2015; Helsper and Eynon, 2013).

Therefore, a third level of digital divide studies focuses on quantifying the impact of the unequal distribution of benefits of internet use (Quan-Haase, Martin & Schreurs, 2016; Meri-Tuulia, Antero & Suvi-Sadetta, 2017; Sáinz, Arroyo & Castaño, 2020; Scheerder, van Deursen & van Dijk, 2017; van Deursen & Helsper, 2015). The third digital gender gap refers to this differentiated use of the most advanced ICT technologies and applications. The recently created EU Women in Digital Scoreboard confirms that there is still a substantial gender gap in specialist digital skills. According to the WiD Scoreboard, even in those Member States where gender mainstreaming is more advanced, 'stereotypes and preconceptions' continue to create obstacles for women and girls (European Commission, 2019) and gender differences have persisted fairly stable along these years (Martínez-Cantos, 2017). These findings are in line with other longitudinal studies from particular contexts, such as the Netherlands, where the gender differences regarding digital skills have remained consistent in recent years (van Deursen & van Dijk, 2015; van Deursen, van Dijk & ten Klooster, 2015).

The digital gender divide becomes even more pronounced when it comes to women as creators of technology. Women are still under-represented in information and communication technology (ICT) jobs, top management and academic careers. This pattern applies to almost all developed countries and is largely independent of the country's level of economic development (Sorgner et al., 2017). Though 57% of tertiary graduates in the EU are women, only 20% of tertiary graduates in ICT-related fields are women and the share of women in ICT jobs is 19% (EIGE, 2020; European Commission, 2021b). There is no progress, as these figures have been stable over the last few years, but the 2030 Digital Compass has set the target that the EU should have 20 million employed ICT specialists, with convergence between women and men, by 2030 (European Commission, 2021a).

Beyond ICT, a striking gender gap exists among scientists and engineers in the high-technology sectors likely to be mobilised in the design and development of new digital technologies. In 2019, across the EU, there were close to 32 million scientists and engineers employed in high-technology sectors, of whom only one fifth were women. And even when women do study STEM, they face a glass ceiling preventing them from holding senior positions. Software development is also a male-dominated club. The majority of software packages are still authored by men. Start-ups and venture capital investment point to socio-cultural gender bias in equity financing: 93% of innovative start-ups seeking venture capital investments have been founded by men, women-owned start-ups receive 23% less funding and are 30% less likely to have a positive exit (European Commission, 2019; OECD, 2018). In summary, STEM sectors do not seem to be able to incorporate, retain and promote women properly. Gender inequalities remain and generate equity and efficiency problems that hamper economic growth and welfare for all, but especially for women (Vergés et al., 2021).

Summing up, while a number of positive policy developments can be noted, major challenges remain if gender equality in the digital world of work is to be achieved. One of the main challenges is the development of gender-specific and gender-sensitive indicators and indices that provide insights into the depth and breadth of women's digital inclusion, since ICT-focused indices which include gender dimensions have a relatively short history (Brimacombe & Skuse, 2013) The EU digital strategy 'Shaping Europe's digital future' (European Commission, 2020a) and the EU gender equality strategy 2020–2025 are the last steps taken by the EU for the integration of a gender perspective in this area. These initiatives are placing an emerging emphasis on the collection of sex-disaggregated data and development of indicators. However, the Women in Digital Scoreboard is still the only measurement framework to monitor the progress of European countries towards women's digital inclusion. Therefore, in this study we critically analyse it using the poset methodology, construct a more refined ranking and examine the differences by macroregions and clusters of countries.

3.3 Material and methods

The Women in Digital (WiD) Scoreboard is part of the Digital Economy and Society Index (DESI) and assesses in detail women's participation in the digital economy in the EU-28 countries. It is based on 12 indicators divided in three dimensions (European Commission, 2020c), namely internet use, internet user skills, and specialist skills and employment.

The first dimension (internet use) is composed of six indicators, listed as follows: 1.1 *% of women who use the internet at least once a week*; 1.2 *% of women who never used the internet*; 1.3 *% of women who used the internet in the previous three months to use online banking*; 1.4 *% of women who used internet in the previous three months for doing an online course*; 1.5 *% of women who used internet in the previous three months for taking part in on-line consultations of voting to define civic or political issues*; 1.6 *% of women internet users who, during the previous year, needed to send filled forms to the public administration*. The breakdown for the indicators of this dimension is all females aged 16-74, and the source of the data is the Community survey on ICT usage in households and any individuals, provided by Eurostat.

The second dimension (internet user skills) consists of three indicators, which are: 2.1 *% of women with basic or above basic digital skills in information, communication, problem solving and software for content creation*; 2.2 *% of women with above basic digital skills in information, communication, problem solving and software for content creation*; 2.3 *% of women who have used advanced spreadsheet functions, created presentation or document* *integrating text, pictures and tables or charts, or written code in a programming language*. The breakdown and the source of the data are the same of the ones of the first dimension.

The third dimension (specialist skills and employment) contains the last three indicators of the index: 3.1 *Women graduates in STEM per 1000 individuals ages 20-29*; 3.2 *% of women aged 15-74 employed ICT specialist based on the ISCO-08 classification*; 3.3 *Gender pay gap in unadjusted form, considering all employees working in firms with ten or more employees*. The source of the data of this dimension is Eurostat questionnaire on education statistics, the labour force survey, and the structure of earnings survey. Indicator 3.3 measures the difference between male's average gross hourly earnings and female's one as a percentage of male's average gross hourly earnings.

In the WiD index 2020, all the indicators are considered of equal importance, and the aggregation of the indicators into the three dimensions and into the overall index is constructed as the simple unweighted arithmetic average of the normalised scores. In this paper, we use the normalised scores available from the Women in Digital website. No missing data are detected. The data matrix considered in this study is composed of 28 countries and 12 indicators; hence, the total number of observations is 336.

In this study we apply the partial order theory – or poset-based approach –, a discipline associated with discrete mathematics, in which the objects of a data set, composed of multiple indicators, are compared and ordered to obtain a ranking (Brüggemann & Patil, 2011). According to the poset theory, one object could be considered better than another if and only if it has better performance in all indicators of a data set, or, alternatively, if it is better performing in just one indicator and it ties in all the others. Furthermore, all the ordered pairs of objects could be graphically represented in the so-called Hasse diagram.

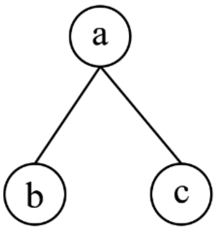
In the analysis presented in this work, the first step consists in the identification of the Hasse diagram, which represents the relations between the 28 countries according to their scores considering all 12 indicators together. To better understand poset's theory, consider the example in Table 3.1, identifying three countries (A, B, and C) and three indicators (q_1, q_2, and q_3).

Table 3. 1 – Example: three countries (a, b, and c) and three indicators (q1, q2, and q3)

Country	q_1	q_2	q_3
A	4	3	2
B	3	2	2
C	4	2	0

In the poset analysis it is crucial to compare all countries based on all indicators. Therefore, we could say that country A is better than both country B and country C since, even if it ties in q_3 with country B and in q1 with country C, it shows a higher score on all other indicators. What is not possible to compare is country B with country C: country B shows a higher score in q_3 (2 > 0), but a lower score in q_1 (3 < 4); hence, country B is incomparable with country C. The relations among the comparable countries are country A > country B as well as country A > county C. At the same time, country B || country C (where || is the sign to represent incomparability). Figure 3.1 shows the Hasse diagram of our example.

Figure 3. 1 – Example: Hasse diagram



The second step of the analysis includes the identification of the downset of any country as well as the incomparabilities in order to construct the ranking of the countries. The downset of country x consists of those countries y such that $y \leq x$; its cardinality is denoted as D(x). If $y < x$ for one or more indicators and $y > x$, then x and y are incomparable; the number of countries that are incomparable with a country x is denoted as I(x). In our example, we obtain the results as shown in Table 3.2.

Table 3. 2 – Example: downsets and incomparabilities of the objects, in numbers.

Country	D(x)	I(x)
A	3	0
B	1	1
C	1	1

According to Table 3.2, the downset of country A is composed of three elements (country A itself, country B, and country C). To rank the countries, we apply the Local Partial Order Model (LPOM), where the "final score" of the countries is a function of D(x) and I(x). The formula to compute the "final score" of is as follows: (Brüggemann & Patil, 2011)

$$\delta(x) = D(x)\ [(n+1)/(n+1-I(x))] \qquad (1)$$

where x is the country of interest and n indicates the total number of countries, in our example, $n = 3$. For instance, the score of country A, applying the formula, is: 3 * (3 + 1) / (3 + 1 – 0) = 3 * 4 / 4 = 3. By contrast, the score of both countries B and C is: 1 * (3 + 1) / (3 + 1 – 1) = 1 * 4 / 3 = 1,33. Thus, we obtain the following ranking: first position for country A and second position for both countries B and C (tie). If we create a ranking by simply computing the unweighted arithmetic average, we will obtain a different ranking with country B better ranked than country C. In our analysis we will use the LPOM to create the ranking considering first the whole dataset, and then we will repeat the same process for each of the three dimensions of indicators.

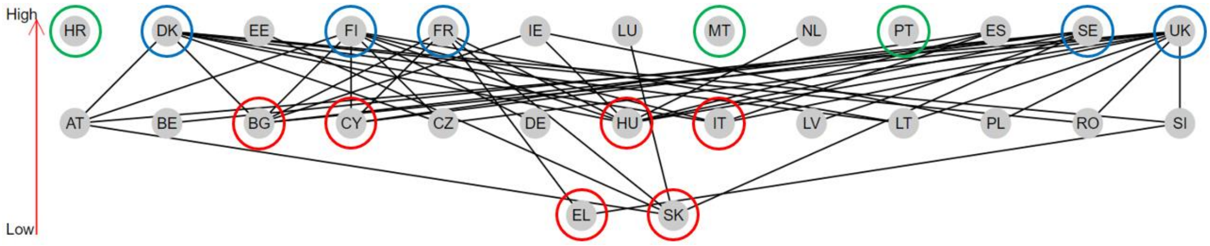
The third step of the analysis consists in the detection of the most significant indicators for each of the three dimensions through the "attribute-related sensitivity" analysis. The aim is to examine how an indicator influences the position of the countries in the Hasse diagram by removing one indicator from the data matrix (Brüggemann & Patil, 2011).

Our goal is to find the four out of six most important indicators of dimension 1, and the two out of three most relevant indicators of both dimension 2 and 3. We will conduct this analysis considering first the whole 28 countries and then just the countries of each of the four European macroregions (northern, western, eastern, and southern Europe) to examine in-depth regional variations. The last step of the analysis is the comparison between our results and the results of the Women in Digital Scoreboard for 2020, identifying similarities and differences among both rankings. The poset-based approach is applied using the online software called "PyHasse", available at https://posets.pyhasse.org/.

3.4 Results

In this section, we present the main results of the analysis, starting from the first step of the analysis, namely the Hasse Diagram of the 28 countries considering all the 12 indicators of the three dimensions (Figure 3.2).

Figure 3. 2 – Hasse Diagram, 28 countries and 12 indicators



The Hasse Diagram shows the connections between the countries analysed according to their data. The lines connecting two countries reveal that the country in the higher level is better than the country in the lower level, since it has higher scores in all the 12 indicators. On the one hand, the countries circled in blue (Denmark, Finland, France, Sweden, and United Kingdom) are in a higher-level respect to, at least, five countries; on the other hand, the countries circled in red (Bulgaria, Cyprus, Greece, Hungary, Italy, and Slovakia) are low-performing states in all indicators respect to, at least, five countries. The three countries circled in green (Croatia, Malta, and Portugal) are incomparable with all other countries; this means that they are very good performing in at least one indicator, as well as very low performing in other indicator(s).

From the Hasse Diagram it is now possible to move on to the second step of the analysis: to compute the downsets and the number of incomparabilities of each country, for calculating the final scores using the Local Partial Order Model (LPOM), and then constructing our own ranking, according to Figure 3.3.

Figure 3. 3 – Ranking of the countries according to their scores obtained as a function of the downsets and the incomparabilities

Rank	Country	downset	incomp	Score
1	United-Kingdom	12	16	26.77
2	Denmark	11	17	26.58
3	Finland	10	18	26.36
4	Sweden	7	21	25.38
5	France	6	22	24.86
6	Ireland	4	24	23.20
6	Spain	4	24	23.20
8	Estonia	2	26	19.33
8	Luxembourg	2	26	19.33
8	Netherlands	2	26	19.33
11	Croatia	1	27	14.50
11	Malta	1	27	14.50
11	Portugal	1	27	14.50
14	Slovenia	2	24	11.60
15	Austria	2	23	9.67
15	Belgium	1	26	9.67
15	Germany	1	26	9.67
15	Latvia	1	26	9.67
19	Poland	1	25	7.25
19	Romania	1	25	7.25
21	Czechia	1	24	5.80
21	Lithuania	1	24	5.80
23	Cyprus	1	23	4.83
23	Greece	1	23	4.83
23	Italy	1	23	4.83
26	Bulgaria	1	21	3.63
27	Hungary	1	19	2.90
27	Slovakia	1	19	2.90

The countries are grouped in four categories depending on their final score. Blue represents the “women digital participation leaders” (leaders, for short) with a final score higher than 20; green is the “medium-high women digital participation” group (medium-high, for short) with a final score between 10 and 20; in yellow the "medium-low women digital participation" group (medium-low, for short) is identified with a final score between 5 and 10; finally, red represents the "emerging women digital participation" group (emerging, for short) a final score lower than 5. All groups are composed of seven countries except the medium-low group (in yellow), which counts eight countries since Czechia and Lithuania have the same score.

To discuss this ranking, we should look at the downsets: 12 countries have greater results in all indicators with respect to at least one country (downset greater than 1). The largest downset is the United Kingdom's one: the UK has higher scores than 11 countries in all indicators. Hence, the UK leads the ranking thanks to its good performance in all indicators. By contrast, the countries with the downset equal to 1, are underperforming in at least one indicator respect to all other countries; 16 countries are in this situation, and their final score drops as the number of incomparabilities decreases, which means that the number of countries with better results in all the indicators of the dataset increase. According to our results, the last positions of the ranking are occupied by Hungary and Slovakia: they both present incomparabilities with 19 countries, which means that 8 different countries have higher scores in all indicators with respect to them. We show these results also in Figure 3.4 through political maps, one for each of the four macroregions.

Figure 3. 4 – Results of the poset analysis considering all three WiD 2020 dimensions. Clockwise from top left: northern Europe, western Europe, eastern Europe, southern Europe.

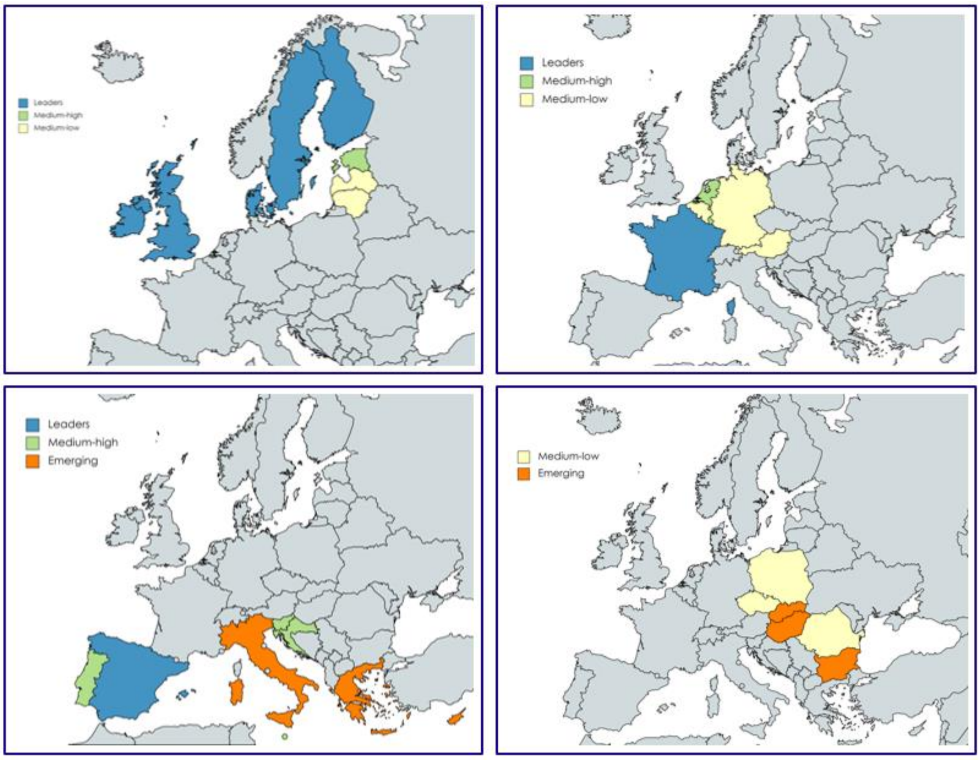


Figure 3.4 shows that the results significantly differ among the macroregions. Northern European countries are all in the leaders' group, except the Baltic countries (Estonia is in the middle-high group, Latvia and Lithuania are in the middle-low group). Western European countries range from the leaders' group (France) to the medium-low group (Austria, Belgium, and Germany). Southern European countries present Spain in the leaders' group, but at the same time Cyprus, Greece, and Italy are in the emerging group. Finally, eastern European countries belong only to the last two groups of the ranking: Czech Republic, Poland, and Romania are in the middle-low group, whereas Bulgaria, Hungary, and Slovakia are in the emerging group. Therefore, on the one side, northern and western European countries are at the forefront regarding women's participation in the digital economy (especially UK and Scandinavian countries). On the other side, some southern and eastern European countries present great shortcomings in this regard.

The third step of the analysis consists in the attribute-related sensitivity analysis. We identify the two out of three most significant indicators for each dimension. Since the first dimension is composed of six indicators, for this dimension we identify the four most important indicators. The analysis is repeated five times: first considering all the countries together, and then considering one of the four macroregions at a time. The results are presented in Table 3.3. The indicators are listed following the enumeration presented in Section 3.3.

Table 3. 3 – Most impacting indicators according to the attribute-related sensitivity analysis, both at EU-28 level, and at macro-regional level

Indicator	EU-28	Northern EU	Western EU	Southern EU	Eastern EU
1.1	X	X		X	X
1.2	X				X
1.3			X		X
1.4					
1.5	X	X	X	X	
1.6	X				X
2.1					
2.2	X	X	X	X	X

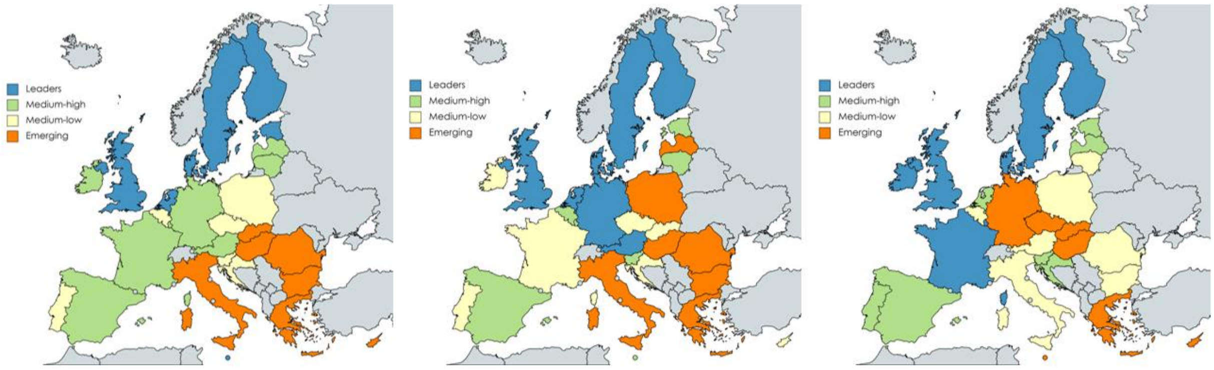
2.3	X			X	X
3.1	X	X	X	X	X
3.2			X	X	
3.3	X	X	X	X	X

The indicators in bold in the first column are the most relevant ones for all the EU-28 countries. Regarding the analysis at a macro-regional level, it is important to underline the following considerations: first, notice that it was not possible to identify the two out of three indicators with the highest impact for all macroregions in all dimensions (for instance, the most important indicators considering northern European countries are just two out of six in dimension 1, and just one out of three in dimension 2); second, in some cases it was not possible to find more important indicators than others in a dimension (for instance, all three indicators of dimension 3 have the same impact in western and southern European countries) and for this reason all indicators of that dimension are considered as equivalent.

The attribute-related sensitivity analysis has then revealed 8 out of 12 most significant indicators for the EU-28. Specifically, three indicators about internet user skills, specialist skills and employment have the highest impact in all four macroregions, namely: 2.2 *% of women with above basic digital skills in information, communication, problem solving and software for content creation*; 3.1 *Women graduates in STEM per 1000 individuals ages 20-29*; 3.3 *Gender pay gap in unadjusted form, considering all employees working in firms with ten or more employees*. Furthermore, two more indicators about internet use are the most relevant in three macroregions: 1.1 *% of women who use the internet at least once a week*; 1.5 *% of women who used internet in the previous three months for taking part in on-line consultations of voting to define civic or political issues*. Moreover, two indicators are significant in some macroregions even if they are not in the EU-28 analysis: indicator *1.3* in western and eastern Europe; and indicator *3.2* in western and southern Europe. Only two indicators are left out in all the analyses: 1.4 *% of women who used internet in the previous three months for doing an online course*; 2.1 *% of women with basic or above basic digital skills in information, communication, problem solving and software for content creation*.

Another interesting aspect of the analysis that deserves attention is the ranking obtained considering the three dimensions individually. The ranking is expressed in the form of the four performance categories discussed above. The results are showed in the maps of Figure 3.5.

Figure 3. 5 – Results of the poset analysis considering the three WiD 2020 dimensions singularly. From left to right: dimension 1, dimension 2, and dimension 3.



Looking at Figure 3.5 we can first consider that only a few countries are in the same performance category in all the dimensions. On the one hand, Denmark, Finland, Sweden, and the UK are the only four countries that are leaders in each of the three dimensions; on the other hand, Greece and Hungary are the only two countries in the bottom of the ranking in all the three dimensions. Our findings identify in which dimension some countries could improve the most. For instance, Austria, Germany, Lithuania and Luxembourg have good results in the first two dimensions, but they have strong weaknesses (especially Germany and Luxembourg) in the dimension related to specialist skills and employment. Another example is represented by France, Ireland, and Latvia, which are in the middle-high category in both dimensions one and three, but they could improve significantly in the dimension regarding internet user skills. Slovenia is the only country that has its strongest lacks in the first dimension (internet use).

The last step of the analysis concerns the comparison between the results and ranking proposed in the Women in Digital Scoreboard 2020 and our results obtained using the poset-based approach, by computing the Spearman correlation coefficient ρ, and the τ Kendall correlation, as in Alaimo et al. (2021a & 2021b). First of all, to test the validity of our ranking we calculate the Spearman correlation coefficient ρ, using the following formula:

$$\rho = 1 - \frac{6 * \sum d_i^2}{n * (n^2 - 1)} \qquad (2)$$

where d is the pairwise distance of the ranks of the different countries and n is the number of countries. The result (in a range between 0 and 1, where 0 is total discordance and 1 is total concordance) is 0.881 with a p value < 0.001. We also calculate the τ Kendall rank correlation, applying the following formula:

$$\tau = \frac{c - d}{c + d} \quad (3)$$

where c is the number of concordant pairs and d the number of discordant pairs. The result (in a range between -1 and 1, where -1 is total discordance and 1 is total concordance) is 0.687 with a p value < 0.001. The high values of the coefficients mean that the results obtained in the two ranking are similar, even if there are some differences, which we try to explain starting from Figure 6, which shows the scores of the 28 countries comparing our analysis and the score reported in the WiD Scoreboard 2020, and in Figure 3.6, which present the countries that change at least three positions in the ranking.

Figure 3. 6 – Comparison between the scores of the 28 countries in the Women in Digital Scoreboard 2020 and in the poset-based approach analysis

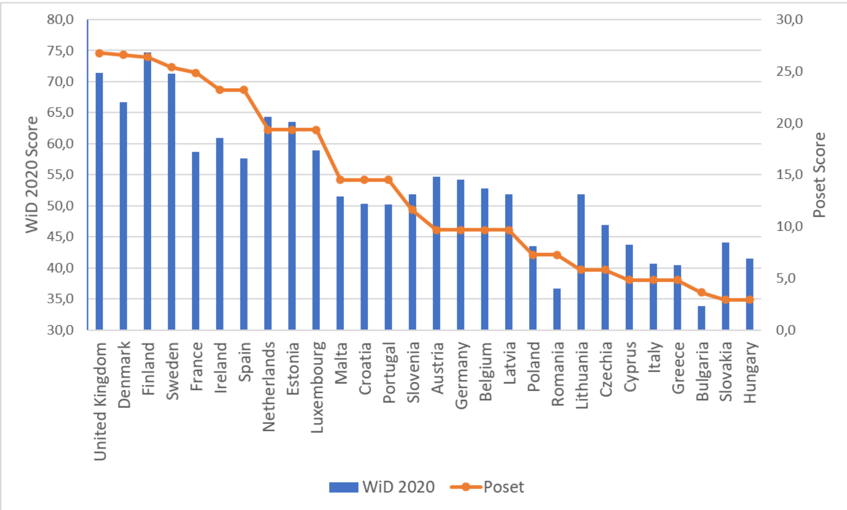
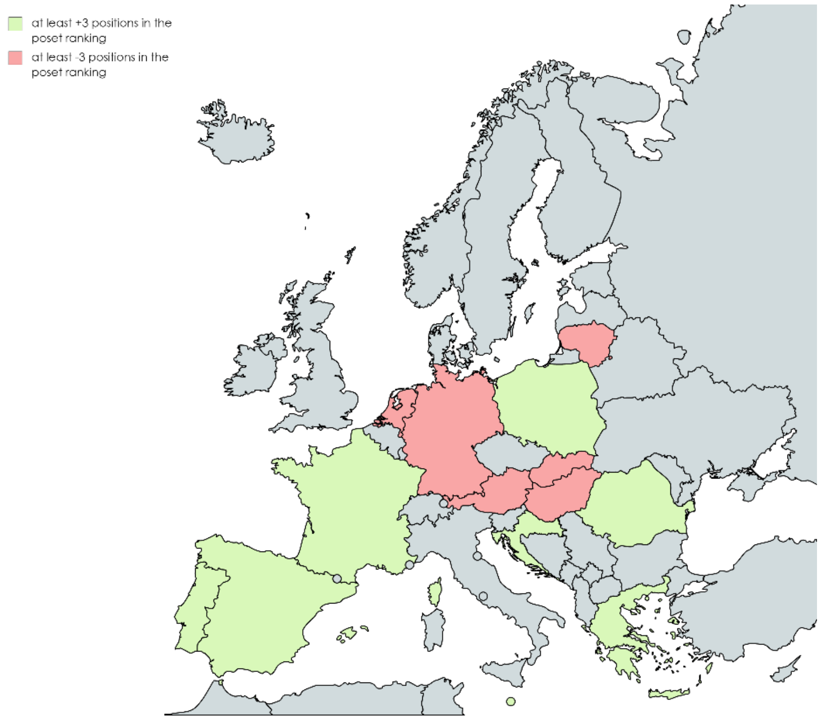


Figure 3.6 shows that the ranking obtained with the poset-based approach is not the same as the one in the Women in Digital Scoreboard 2020. Countries like France, Spain or Ireland improve their ranking a lot with the poset methodology, while Lithuania, Slovakia or Hungary move down in the ranking. Therefore, our results show a quite distinct order of the ranking of countries.

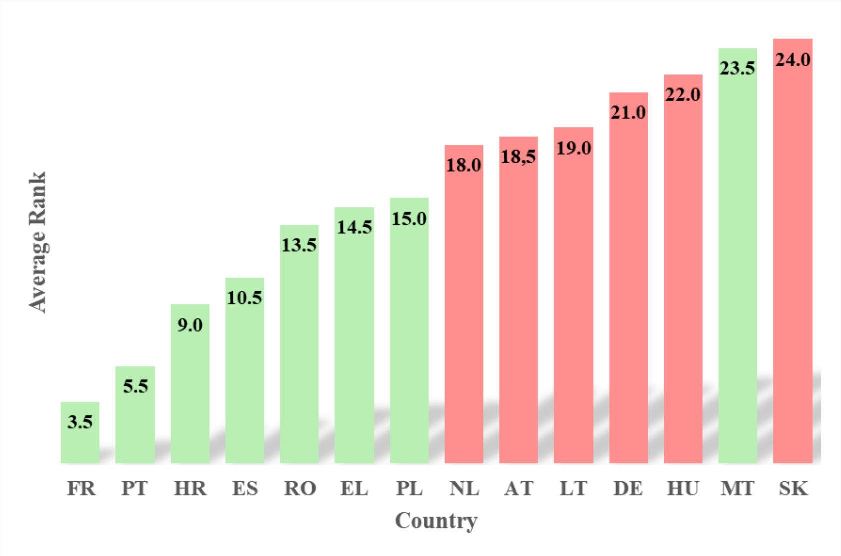
Figure 3.7 represents in green the countries that improve their results of at least three positions in the poset-based analysis compared to their ranking in the Women in Digital Scoreboard: Portugal and Romania (8), Croatia (7), Malta (6), France, Poland, and Spain (4), and Greece (3). In red, the countries that fall off at least three positions in the ranking: Lithuania (-7), Slovakia (-6), Austria (-4), Germany, Hungary, and Netherlands (-3). The explanation of the differences between the two rankings lies in the performance of these countries in the most important indicators, particularly indicators *1.5*, *1.6*, *3.1*, and *3.3* (the last two even more substantially). Good performances in these indicators led to best ranking in the poset-based analysis as well as deficiencies in these indicators led to worse results in the ranking. To better understand this phenomenon, Figure 3.8 shows the average ranking of the aforementioned countries considering only indicators *3.1* and *3.3*.

Figure 3. 7 – Comparison between the poset-based approach analysis and the Women in Digital Scoreboard 2020 results



According to Figure 3.8, we can identify two groups of countries considering the results in indicators *3.1* and *3.3*, except Malta, which climbs up the ranks even though its results in the relevant indicators of the third dimension are not good (25th in *3.1*, and 21st in *3.3*). However, as explained in the previous pages, Malta belongs to the small group of countries that cannot be compared with the rest, according to the poset-based approach theory, thanks overall to its very good results in indicator *1.5*, where Malta ranks 2nd. We can conclude that, in general, countries that have good results in the majority of indicators should improve their results in the most relevant indicators: *Women graduates in STEM per 1000 individuals ages 20-29* and *Gender pay gap in unadjusted form, considering all employees working in firms with ten or more employees*.

Figure 3. 8 – Average ranking of the 13 countries underlined in Figure 3.6, according to indicators 3.1 and 3.3



3.5 Discussion and conclusions

In this article we applied the poset theory to analyse women's digital inclusion in the EU-28 countries using the data from the 12 indicators of the Women in Digital Scoreboard 2020. The poset methodology allowed to construct a new ranking that avoids the shortcomings of the aggregative approaches. The analysis resulted also in a classification of countries, according to our new ranking, in four groups depending on their performance level. The leaders group is composed of the United Kingdom, Denmark, Finland, Sweden, France, Ireland, and Spain (leaders group); by contrast, the countries where women are most underrepresented are Slovakia, Hungary, Bulgaria, Italy, Greece, and Cyprus (emerging group).

According to the poset-based approach, the leaders group is composed of those countries who present better results in all indicators compared to at least other three different countries. United Kingdom is the country leading the ranking, since it has better scores in all indicators with respect to other eleven different countries; Ireland and Spain, the last two countries of the leaders group show better results in all indicators respect to three different countries. Similarly, the emerging group is composed of those countries underperforming in all indicators compared to at least four different countries (as in the case of Cyprus, Greece, and Italy) –eight in the case of Hungary and Slovakia, who are in the last positions of the ranking. Moreover, three countries (Croatia, Malta, and Portugal) are incomparable with all the other countries; this means that they present very good results in some indicators and very low scores in other indicators.

We also analysed the data by macroregions, and the results seem to confirm the socio-economic pattern among European countries: northern countries are mostly in the leaders group, western countries are between the leaders and the middle group, southern countries are mostly between the middle and the emerging group, and eastern countries are mostly in the emerging group. Thus, countries in a macroregion usually belong to the same group, with very few exceptions: the former soviet Baltic states are the only countries in the North of Europe that are not in the leaders group, and Spain is the only southern European country represented in the leaders group.

Comparing our results with the ones proposed in the Women in Digital Scoreboard 2020 report, we found that half of the countries have equal or similar positions in the two rankings while the other half move up or down the ranking at least three positions. These differences depend mainly on the performance of countries in the most significant indicators revealed by the attribute-related sensitivity analysis that we have conducted, considering all countries first, and then the countries in each macroregion. Among the most relevant indicators two of them belong to the first dimension, internet use *(% of women who used internet in the previous three months for taking part in on-line consultations of voting to define civic or political issues*, and *% of internet users who, during the previous year, needed to send filled forms to the public administration*), and other two belong to the third dimension, specialist skills and employment (*Women graduates in STEM per 1,000 individuals ages 20-29*, and *Gender pay gap in unadjusted form, considering all employees working in firms with ten or more employees*). The last two indicators are even more important in the determination of the ranking. In fact, the countries who improved their ranking are those with good results in indicators *3.1* and *3.3* (except for Malta). The macroregional analysis performed results in different relevant indicators for each European region. For instance, our analysis suggests that only for western and southern European countries an important indicator is *% of women aged 15-74 employed ICT specialist based on the ISCO-08 classification*. Therefore, our research contributes to better understand the dynamics and the underlying causes of women's digital inclusion in the EU-28 countries. And the different significance of indicators in the EU and in the four macroregions helps to design more targeted and effective policies, showing the specific areas in which each country should focus to reduce the gender digital divide.

This study presents some limitations, both theoretical and methodological. Under the theoretical point of view, the set of indicators is not so exhaustive, and the data are collected just at country level. The methodological limitations are mainly related to the fact that, even in quite large dataset, the application of the poset-based approach can lead to a high number of incomparabilities, generated in some cases by small differences in the performance of some indicators.

For future research, it would be very interesting to collect regional data in order to replicate the analysis at a regional level and explore in detail the regional variances in the gender digital divide across Europe, as regional socioeconomic differences in some countries (such as Italy and Spain) are usually very high. Moreover, if more gendered data were available the set of indicators could be further enlarged, including for example many of the other indicators of the Digital Economic and Society Index (DESI), which is currently composed of 25 indicators. Finally, as data for more years is available, we will be able to expand the study to include a longitudinal analysis.

References

Alaimo, L.S., Arcagni, A., Fattore, M., & Maggino, F. (2021a). Synthesis of Multi-indicator System Over Time: A Poset-based Approach. *Social Indicators Research 157*, 77-99, https://doi.org/10.1007/s11205-020-02398-5.

Alaimo, L.S., Ciacci, A., & Ivaldi, E. (2021b). Measuring Sustainable Development by Non-aggregative Approach. *Social Indicators Research 157*, 101-122, https://doi.org/10.1007/s11205-020-02357-0.

Arroyo, L. (2020). Implications of Digital Inclusion: Digitalization in Terms of Time Use from a Gender Perspective. *Social Inclusion, 8(2)*, 180-189. https://doi.org/10.17645/si.v8i2.2546.

Arroyo, L., & Valenduc, G. (2016). Digital Skills and Labour Opportunities for Low-Skilled Woman. *Dynamics of Virtual Work, 6*, 16.

Bánhidi, Z., Dobos, I., & Nemeslaki, A. (2020). What the overall Digital Economy and Society Index reveals: A statistical analysis of the DESI EU28 dimensions. *Regional Statistics, 10(2)*, 46-62. https://doi.org/10.15196/RS100209.

Brimacombe, T., & Skuse, A. (2013). Gender, ICTs, and Indicators: Measuring Inequality and Change. *Gender, Technology and Development, 17(2)*, 131-157. https://doi.org/10.1177/0971852413488713.

Brüggemann, R. & Patil, G.P. (2011). Ranking and Prioritization for Multi-Indicator Systems. Introduction to Partial Order Applications. Springer-Verlag New York. ISBN: 978-1-4419-8476-0. doi:10.1007/978-1-4419-8477-7.

Castaño, C., Martín, J., & Martínez, J. L. (2011). La brecha digital de género en España y Europa: Medición con indicadores compuestos. *Revista Española de Investigaciones Sociológicas, 136*. https://doi.org/10.5477/cis/reis.136.127.

Di Bella, E., Leporatti, L., Maggino, F. & Gandullia, L. (2018). A poset based indicator of gender equality at sub-national level. ASMOD 2018. *Proceedings of the International Conference on Advances in Statistical Modelling of Ordinal Data*. https://doi.org/10.6093/978-88-6887-042-3.

Di Brisco, A.M. & Farina, P. (2018). Measuring Gender Gap from a Poset Perspective. *Social indicators research 136*, 1109-1124. https://doi.org/10.1007/s11205-017-1582-8.

EIGE (2017). *Economic Benefits of Gender Equality in the EU: EU and EU Member States overviews*, Publications Office of the European Union, Luxembourg.

EIGE (2020). *Gender Equality Index 2020—Digitalisation and the future of work*. https://eige.europa.eu/publications/gender-equality-index-2020-digitalisation-and-future-work.

Elena-Bucea, A., Cruz-Jesus, F., Oliveira, T. et al. (2021). Assessing the Role of Age, Education, Gender and Income on the Digital Divide: Evidence for the European Union. *Information Systems Frontiers 23*, 1007–1021 https://doi.org/10.1007/s10796-020-10012-9.

European Commission (2018). *Women in the digital age: Final report*. Directorate General for Communications Networks, Content and Technology. Publications Office. https://data.europa.eu/doi/10.2759/526938.

European Commission (2019). *Women in Digital - Shaping Europe's digital future*. https://digital-strategy.ec.europa.eu/en/library/women-digital.

European Commission (2020a). *Shaping Europe's Digital Future*. Retrieved from https://ec.europa.eu/info/sites/info/files/communication-shaping-europes-digital-future-feb2020_en_4.pdf.

European Commission (2020b). *Women in Digital Index 2020. Methodology*. https://comunic.ro/wp-content/uploads/2021/03/WomeninDigitalIndex2020-methodology.pdf.

European Commission (2020c). *Women in Digital Scoreboard 2020*. Available at https://bit.ly/3CPgXnH, last consulted: September 27, 2021.

European Commission (2021a). *2030 Digital Compass: the European way for the Digital Decade*. https://eufordigital.eu/wp-content/uploads/2021/03/2030-Digital-Compass-the-European-way-for-the-Digital-Decade.pdf

European Commission (2021b). *Women in Digital Scoreboard 2021*. Available at https://digital-strategy.ec.europa.eu/en/news/women-digital-scoreboard-2021, last consulted: December 1, 2021.

Hargittai, E. (2010). Digital Na(t)ives? Variation in Internet Skills and Uses among Members of the “Net Generation”. *Sociological Inquiry, 80(1),* 92-113. https://doi.org/10.1111/j.1475-682X.2009.00317.x.

Hargittai, E., & Shaw, A. (2015). Mind the skills gap: The role of Internet know-how and gender in differentiated contributions to Wikipedia. *Information, Communication & Society, 18(4)*, 424-442. https://doi.org/10.1080/1369118X.2014.957711.

Helsper, E. J. (2010). Gendered Internet Use Across Generations and Life Stages. *Communication Research, 37(3)*, 352-374. https://doi.org/10.1177/0093650209356439.

Helsper, E. J., & Eynon, R. (2013). Distinct skill pathways to digital engagement. *European Journal of Communication, 28(6)*, 696-713. https://doi.org/10.1177/0267323113499113.

JRC (2019). *The changing nature of work and skills in the digital age*. Publications Office of the European Union. https://doi.org/10.2760/679150.

Kohlrausch, B, & Weber, L. (2020). Gender Relations at the Digitalised Workplace: The Interrelation Between Digitalisation, Gender, and Work. *Gender a výzkum / Gender and Research, 21(2)*, 13–31, http://dx.doi.org/10.13060/gav.2020.010.

Mariscal, J., Mayne, G., Aneja, U., & Sorgner, A. (2019). Bridging the Gender Digital Gap. *Economics, 13(1)*, 20190009. https://doi.org/10.5018/economics-ejournal.ja.2019-9.

Martínez-Cantos, J. L. (2017). Digital skills gaps: A pending subject for gender digital inclusion in the European Union. *European Journal of Communication, 32(5),* 419-438. https://doi.org/10.1177/0267323117718464.

Meri-Tuulia, K., Antero, K., & Suvi-Sadetta, K. (2017). Differences between the genders in ICT skills for Finnish upper comprehensive school students: Does gender matter? *SEMINAR.NET, 13(2)*, 17.

Norlén, H., Papadimitriou, E., Dijkstra, L. (2019). *The Regional Gender Equality Monitor. Measuring female disadvantage and achievement in EU region*, EUR 29679 EN, doi:10.2760/472693, JRC115814.

OECD (2001). *Understanding the digital divide*, https://www.oecd.org/sti/1888451.pdf.

OECD (2018). *Bridging the digital gender divide*. Include, Upskill, Innovate, https://www.oecd.org/digital/bridging-the-digital-gender-divide.pdf.

Quan-Haase, A., Martin, K., & Schreurs, K. (2016). Interviews with digital seniors: ICT use in the context of everyday life. Information, *Communication & Society, 19(5)*, 691-707. https://doi.org/10.1080/1369118X.2016.1140217.

Rodríguez-Modroño, P. (2011). Las mujeres y las tecnologías de la información y las comunicaciones, in O. Marcenaro (Ed.), *La cambiante situación de la mujer en Andalucía* (Colección Realidad Social 07, pp. 159-184). Sevilla: Fundación Pública Andaluza Centro de Estudios Andaluces.

Rogers, E.M. (2001). The Digital Divide. *Convergence 7(4)*, 96-111.

Scheele, A. (2007). Gender and the quality of work: An overview of European and national approaches. *Transfer: European Review of Labour and Research, 13(4)*, 595-610. https://doi.org/10.1177/102425890701300406.

Scheerder, A., van Deursen, A., & van Dijk, J. (2017). Determinants of Internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. *Telematics and Informatics, 34(8)*, 1607-1624. https://doi.org/10.1016/j.tele.2017.07.007.

Seybert, H. (2007). Gender differences in the use of computers and the Internet. *Statistics in focus, Population and Social Conditions* 119/2007, Eurostat.

Sorgner, A., E. Bode and C. Krieger-Boden (2017). The Effects of Digitalization on Gender Equality in the G20 economies, Kiel Institute for the World Economy, https://www.ifw-kiel.de/pub/e-books/digital_women-final_report.pdf.

van Deursen, A. J. A. M., & Helsper, E. J. (2015). The Third-Level Digital Divide: Who Benefits Most from Being Online? in L. Robinson, S. R. Cotten, J. Schulz, T. M. Hale, & A. Williams (Eds.), *Studies in Media and Communications* (Vol. 10, pp. 29-52). Emerald Group Publishing Limited. https://doi.org/10.1108/S2050-206020150000010002.

van Deursen, A. J. A. M., & van Dijk, J. A. G. M. (2015). Toward a Multifaceted Model of Internet Access for Understanding Digital Divides: An Empirical Investigation. *The Information Society, 31(5)*, 379-391. https://doi.org/10.1080/01972243.2015.1069770.

van Deursen, A. J. A. M., van Dijk, J. A. G. M., & ten Klooster, P. M. (2015). Increasing inequalities in what we do online: A longitudinal cross sectional analysis of Internet activities among the Dutch population (2010 to 2013) over gender, age, education, and income. *Telematics and Informatics, 32(2)*, 259-272. https://doi.org/10.1016/j.tele.2014.09.003.

van Deursen, A. J., & van Dijk, J. A. (2014). The digital divide shifts to differences in usage. *New Media & Society, 16(3)*, 507-526. https://doi.org/10.1177/1461444813487959.

Vergés Bosch, N., Freude, L., Almeda Samaranch, E., & González Ramos, A. M. (2021). Women working in ICT: Situation and possibilities of progress in Catalonia and Spain. *Gender, Technology and Development*, 1-19. https://doi.org/10.1080/09718524.2021.1969783.

Wajcman, J. (2010). Feminist theories of technology. *Cambridge Journal of Economics, 34(1)*, 143-152. https://doi.org/10.1093/cje/ben057.

Wilhelm, A. (2004). *Digital Nation. Toward an inclusive Information Society*. Cambridge, Massachusetts: MIT Press.