Exploring Regional Innovation Dynamics Using Directed Acyclic Graphs

F. Molinari, D. Cisic and B. Kovacic
University of Rijeka, Rijeka, Croatia
mail@francescomolinari.es

***Abstract* – Biannually, the European Commission publishes the Regional Innovation Scoreboard (RIS) as a tool to benchmark the performance of 240 regions, 22 EU Member States (MS) and 4 Associated Countries (AC): Norway, Serbia, Switzerland, and the UK. Only Cyprus, Estonia, Latvia, Luxembourg, and Malta are included at the country level. The RIS consists of a synthetic index that starts from the arithmetic mean of a set of 21 indicators. This allows ranking the various territories of Europe for their innovative performance. EU Regions, MS and AC are expected to take their position on the ranking into account when designing innovation policies. Due to its importance for policymaking, the RIS is subject to recurrent assessments of the extent to which it constitutes a meaningful measure of a territory's innovation performance. This paper contributes to the debate by proposing an approach based on algorithms to discover the causal relations among the RIS variables and visualize them as edges of Directed Acyclic Graphs (DAGs). In so doing, we shed some light on the limitations of the RIS tool to assess the dynamics and performance drivers of the EU national/regional innovation systems.**

Keywords – causal discovery; algorithms; DAG; Bayesian networks; territorial innovation

I. INTRODUCTION

Biannually, the Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs of the European Commission (EC) publishes the Regional Innovation Scoreboard (RIS). This is a composite index constituted by the arithmetic average of 21 indicators, measured at regional level for the largest-sized countries and national level for the others. The higher the average of those (normalized) indicators, the better performing – whatever this means – is a regional or national territory said to be.

Currently, the RIS – or the global ranking generated from its data, not to forget the specific rankings by indicator – is used as a policy support tool to benchmark the innovative performance of 240 European Union's (EU) regions, 22 Member States (MS) and 4 Associated Countries (AC), namely: Norway, Serbia, Switzerland, and the UK. Of the 22 EU MS, only Cyprus, Estonia, Latvia, Luxembourg, and Malta are considered at the country, not at the regional, level. Table 1 lists the indicators in use in the two most recent editions of the RIS (2019 and 2021).

TABLE 1 . LIST OF RIS INDICATORS

Short name	*Indicator*
air	Air emissions by fine particulate matter (PM2.5) in the manufacturing sector
bp_inn	SMEs introducing business process innovations as percentage of SMEs
cited_pub	International scientific co-publications per million population
design	Design applications per billion regional GDP
dig_skills	Individuals who have above basic overall digital skills
emp_SMEs	Employment in innovative SMEs
empl_KI	Employment in knowledge-intensive activities (percentage of total employment)
exp_business	R&D expenditures in the business sector as percentage of GDP
exp_public	R&D expenditures in the public sector as percentage of GDP
inn_exp	Non R&D Innovation expenditures
inn_exp_pp	Innovation expenditures per person employed in innovative SMEs
nn_SMEs	Innovative SMEs collaborating with others as percentage of SMEs
int_sci_pub	International scientific co-publications per million population
IT	ICT specialists (as a percentage of total employment)
LLL	Percentage population aged 25-64 participating in lifelong learning
Marketing innovators	SMEs introducing marketing or organisational innovations as percentage of SMEs
most_cited_pub	Scientific publications among the top-10% most cited publications worldwide
patent	PCT patent applications per billion regional GDP
performance	RIS performance indicator
PP_pub	Public-private co-publications per million population
prod_inn	SMEs introducing product innovations as percentage of SMEs
Sales	Sales of new-to-market and new-to-firm product innovations in SMEs as percentage of turnover
SME_collab	Innovative SMEs collaborating with others as percentage of SMEs
SME_in_house	SMEs innovating in-house as percentage of SMEs
tert_edu	Percentage population aged 25-34 having completed tertiary education
trademark	Trademark applications per billion regional GDP
Tert_edu	Percentage population aged 25-34 having completed tertiary education

Part of the research supporting this paper was made possible by an Interreg MED 2014-2020 grant to EFRI, the University of Rijeka's Faculty of Economics and Business, under the project named "Social and Creative" (S&C). However, the opinions expressed here are solely of the authors and do not engage in any way the Interreg MED Joint Secretariat or other partners of the S&C project.

Based on the RIS evidence, EU regional and national governments are expected to (re)design their innovation (or sectorial) policies, to improve their positioning on the ranking across time. The intuition is that progress in one or more of the indicators constituting the RIS (see Table 1) should be positively associated with the region's (or country's) performance. For instance, the most innovative region in 2021 was Stockholm, followed by Etelä-Suomi (Finland) and Oberbayern (Germany).

Due to its importance for policymaking in Europe, the RIS is subject to recurrent additions and removals of its components, as well as frequent assessments of the extent to which it constitutes a meaningful measure of a territory's innovation performance. On the one hand, if we compare the 2021 with the 2019 edition of the RIS, 4 new indicators have been included: Individuals who have above basic overall digital skills (*dig_skills*), Innovation expenditures per person employed in innovative SMEs (*inn_exp_pp*), Employed ICT specialists as a percentage of total employment (*IT*), and Air emissions of fine particulates (PM2.5) in Industry (*air*) [1]. On the other hand, authors like [2] and [3] have raised significant objections to the methodological approach used to create the RIS dataset and rankings, as summarised here below.

According to [2], the RIS (like its country-level-only companion, the EIS – European Innovation Scoreboard) does not constitute a meaningful measure of a territory's innovation performance. It misses the fundamental relation between inputs (notably financial resources) and outputs of innovation activities. In other words, it embeds a systematic bias towards those regions and countries that are currently "overspending" in innovation support, compared with the EU average, while it should reward the efficiency of the underlying processes, or a territory's capacity of being quite successful even with a limited amount of resources.

In turn, [3] criticize the fact that the RIS comes up as a linear (that is, an unweighted) average of indicators, which implicitly puts them all on the same level of importance. This may not be realistic for diverse territories, where the dynamics of innovation may show slight or even profound variations in their respective evolutionary patterns.

More generally, the underlying (to both RIS and EIS) model of innovation dynamics has been criticized for being too simplistic or dictated by what data is available for a majority of EU regions and countries rather than its effective degree of realism or explanatory power.

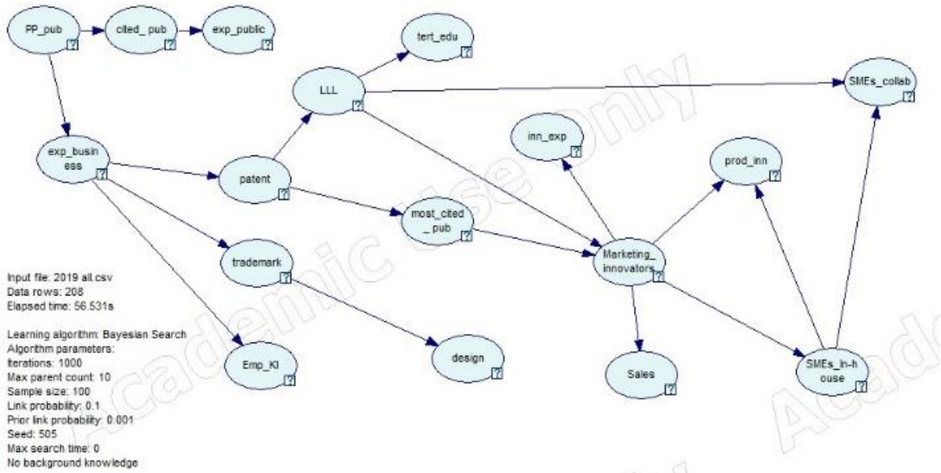
For instance, in earlier editions of the RIS such as in 2009 [4], the triad was presented of Innovation Enablers, Firm Activities and Outputs. The Enablers captured the drivers of innovation that are external to a firm (including human resources). Activities included any innovation effort that a firm can make (and for which data exists, such as for patents and R&D spending). Finally, Outputs encompassed the benefits of innovation for society, such as in terms of employment, sales and (in the reverse, reduction of) air emissions.

This triad is no longer present in the latest RIS edition for 2021 [5] which uses a slightly more complex taxonomy, based on four elements: Framework conditions (such as levels of education, science and digital skills), Investments (including both spend and ICT employment), Innovation activities and Impacts. However, a comparison between EIS and RIS in terms of data sources – such as that done in [5] – highlights all the limitations of the taxonomy's explanatory power as a result of the huge diminution of available time series when moving from the national to the regional level. It sounds therefore rather questionable from a purely statistical point of view the statement made in the RIS report, that "The most innovative regions are typically in the most innovative countries" [5].

This paper contributes to the debate by proposing an alternative approach to analysing the innovation dynamics of a region. The approach is based on the use of Artificial Intelligence algorithms to "discover" the causal relations among the RIS variables for which indicators exist, in a fundamentally atheoretical way. By atheoretical, we mean that there is no prior commitment to any specific model of the economy that could predict the direction of influence between two or more RIS indicators.

Instead, the proposed approach situates itself in the domain of the so-called Causal Discovery, an advanced data analytics discipline that has been successfully applied to medicine, genetics, and ecology, but is still in its early stages in terms of economic applications [6].

The essence of Causal Discovery is to start from data collections (such as the RIS database, attributing a value to each normalized indicator in each EU country and region for a particular year) and to implement Machine Learning techniques across a sufficiently high number of iterations, leading to visualize the most likely causal relations between variables as edges of Directed Acyclic Graphs (DAGs), connecting some nodes with one-way (causal) relations.



```
Input file: 2019 all.csv
Data rows: 208
Elapsed time: 56.531s

Learning algorithm: Bayesian Search
Algorithm parameters:
Iterations: 1000
Max parent count: 10
Sample size: 100
Link probability: 0.1
Prior link probability: 0.001
Seed: 505
Max search time: 0
No background knowledge

Best score in iteration 380: -5099.16
EM Log Likelihood: -4284.35
```

Figure 1. Illustrative example of the RIS 2019 innovation dynamics

An illustrative example of using such technique is visualised in Figure 1, which starts from data of an earlier edition of the RIS, referred to the year 2019.

As one can notice, the DAG starts on the top left with a node without parents and two children. Public-private co-publications per million population (*PP_pub*) as a measure of connectedness between business and academia is shown to directly influence R&D expenditures in the private sector (*exp_business*) and R&D expenditures in the public sector (*exp_public*) only indirectly, i.e. through the intermediate node International scientific citations (*cited_pub*).

This influence looks rather counterintuitive according to both theory and evidence, as publications are normally seen as outputs, instead of enhancers of R&D expenditure. In the case of public administration, scientific publications seem to matter to extent they are cited internationally. A possible explanation can be the importance of reputational factors, deriving from past history of collaborations with the private sector, for academia (as well as enterprises) to receive additional funding for new R&D projects.

Moving one step forward, one can notice that R&D expenditures in the business sector (*exp_business*) have an influence on employment in knowledge-intensive activities (*empl_KI*), on patent applications (*patent*) and on design applications (*design*), although not directly but through trademark applications (*trademark*). Again, this influence is not immediately evident from the experience, unless one accepts that most design innovations in search of protection are in fact those related to a mere aesthetic improvement of some trademark aspects.

Further inspection of the DAG shows an influence of patent applications to lifelong learning (*LLL*), which in turn has influence on the percentage of population with tertiary education (*tert_edu*), the percentage of SMEs collaborating with others (*SME_collab*) and the percentage of SMEs introducing some marketing or organisational innovations (*Marketing innovators*). However, the two latter relations are not exclusive: *SME_collab* are also influenced by the SMEs innovating in-house (*SME_in_house*), which may sound quite reasonable, while *Marketing innovators* are also influenced by the top-10% cited scientific publications (*most_cited_pub*) which has no easy-to-interpret meaning.

Finally, *Marketing innovators* are shown in the DAG to have influence on non-R&D based innovation expenditures (*inn_exp*), on SMEs introducing product innovations (*prod_inn*), on Sales of new-to-market and new-to-firm product innovations in SMEs (*Sales*), as well as on the percentage of SMEs innovating in-house (*SME_in_house*). The latter also has some influence on *prod_inn* together with *Marketing innovators*.

Again, at least part of the above evidence is puzzling. As an example, the direct connection between *LLL* and the propensity of SMEs to collaborate with others or introduce marketing or organisational innovations may sound quite reasonable, but the same cannot be said for the influence of lifelong learning – which pertains to the sphere of vocational training – on an increased percentage of people with tertiary education in a region or country. Same goes for the direct influence of *most_cited_pub* on SMEs acting as *Marketing innovators*, which as we have outlined above, cannot be explained by any promptly available theory.

Without anticipating the conclusions of the paper, we suspect that either a few important indicators are missing from the RIS evidence base, which could be found to act as co-influencers of the dynamics of some variables, or all the heterogeneities that exist within the EU innovation system are too broad to be captured in a single, bird's eye view cutting across the wide diversities of 240 regions and 26 countries of Europe.

The remainder of this paper is structured as follows: in section II we describe in more detail the proposed approach to explore the innovation dynamics of a territory in an atheoretical fashion. In the following three sections we first introduce Bayesian networks as a special category of DAGs and then use them to build an interpretative model of the RIS 2021 database. The model is used for visualisation but also simulation purposes, albeit with some limitations due to unavailable data and lack of computing power. In the last section of the paper, we draw some preliminary conclusions and outline directions for future work.

II. PROPOSED APPROACH

We adopt the Bayesian networks as tools for analysing the dynamics of EU regional innovation sytems. Bayesian networks are a type of probabilistic graphical model used to represent visually and help reflect about uncertain events [6, 7, 8]. They consist of a Directed Acyclic Graph (DAG) that encodes the conditional dependencies between a set of random variables, each associated with a probability table specifying the probability for that variable to have a certain value given the values of the parent nodes in the graph.

The key feature of Bayesian networks is their ability to handle probabilistic inference under uncertainty conditions in a principled and transparent way. They particularly allow for representation and simulation of complex relationships between variables and provide a framework for integrating prior knowledge and data-driven evidence [9]. Bayesian networks are widely used for causal inference and decision-making, as they lead to the identification of the most influential variables in a system and to the calculation and optimisation of alternative decision outcomes despite the uncertainty of available information [10].

Bayesian networks are also used in various thematic fields for modelling and reasoning about complex systems that involve uncertainty and incomplete information. They can be helpful in predicting the likelihood of future events, diagnosing diseases, and making decisions supported by stochastic models. These models are appealing because of their capability to explain intricate processes and provide a well-structured approach for acquiring (solid) knowledge from (noisy) observations [11].

When examining a particular dataset, the objective of Bayesian model selection is to identify the most probable set of connections (causal relations) among the variables that compose it. Unfortunately, if the amount of available data is limited, there could be numerous different models that possess substantial posterior probabilities. Hence, we must use machine learning algorithms that approximate the final model by a repeated number of iterations. While these approaches are computationally compliant, they lack a firm guarantee of the resulting graph's quality. In fact, the graph space is highly "non-convex", and algorithms may become stuck at suboptimal regions.

Additionally, it is known that the exact computation of conditional probabilities in Bayesian belief networks is NP-hard [11], which means that it belongs to a complexity class that is at least as hard as the hardest decision problems that a non-deterministic Turing machine can solve. Thus, an all-too-conservative approach to the visual representation of the 21 indicators now used in the RIS would be very time and resource consuming. However, a variety of algorithms and techniques can be used to simplify the task. These include maximum likelihood estimation, Bayesian search and estimation, and various other methods of machine learning.

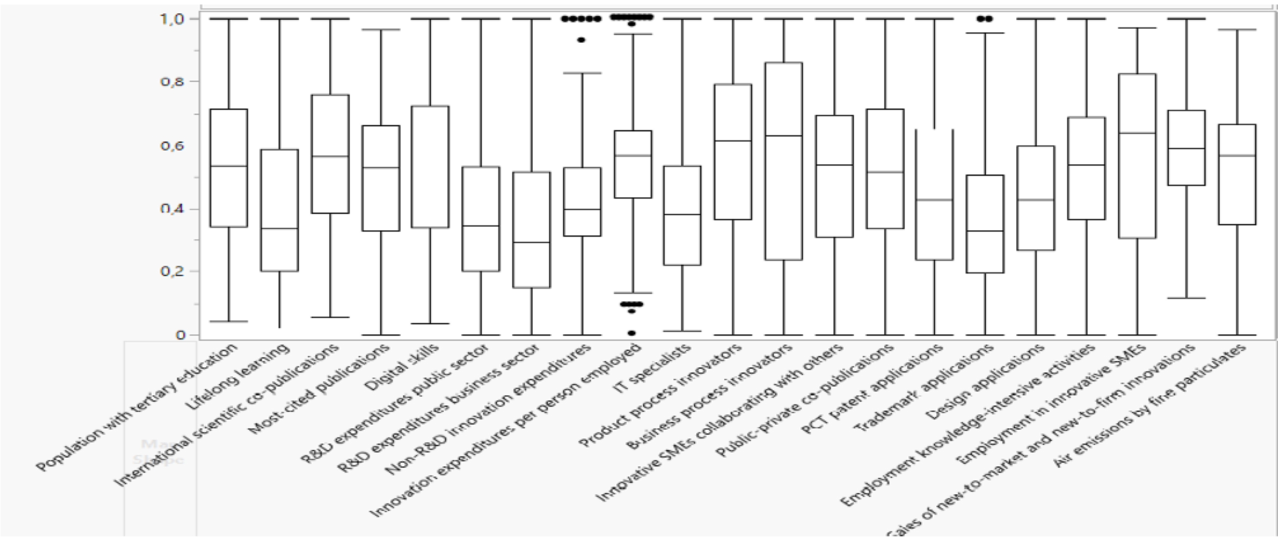


Figure 1. RIS 2021 dataset

The authors have used a Bayesian search algorithm and found out that the results are dependent on the number of iterations and the scoring metric used. The Bayesian search algorithm used for structure learning is Hill Climb with random restarts [12]. The algorithm works by starting with an initial network structure and iteratively making small changes to the structure to improve its score.

One of the key parameters for success is the number of iterations, which determines how many times the algorithm will be run. This number can have a significant impact on the resulting Bayesian model. If the number of iterations is too low, the algorithm may not have enough time to explore all possible network structures and may get stuck in a suboptimal structure. On the other hand, if the number of iterations is too high, the algorithm may overfit the data and create a complex network structure that is not generalizable to new data.

It is also worth noting that the number of iterations is just one of many factors that can affect the quality of the resulting Bayesian model. Other factors include the initial network structure, the quality of the data, and the scoring metric used to evaluate the network.

III. Model building

As mentioned in the Introduction, the two latest datasets RIS 2019 and RIS 2021 do not match each other, because there has been a significant change in indicators and even in the number of EU regions. For instance, compared to the RIS 2019, regional coverage has changed for Croatia from two to four regions, following a revision from the 2016 to the 2021 edition of the NUTS classification. Additionally, the 2019 indicators named “SMEs innovating in-house” and “Marketing or organizational innovators” are no longer in RIS 2021, which introduced four additional ones: Digital skills, IT specialists, Innovation expenditures per person employed, and Air emissions by fine particulates. As the RIS 2019 and RIS 2021 are not fully compatible, we cannot integrate them to get a larger dataset, leaving us with only 240 data rows for 21 indicators. The dataset from RIS 2021 is presented in Figure 2.

To allow Bayesian network modelling, every indicator has been split into 5 equal groups, or quintiles [13], each representing 20% of the given indicator’s range. Spreading indicator rankings to quintiles prevents the data from being too thin to be used. This method is normally practiced in economic policy related calculations [6].

For the RIS 2021, we have used the Bayesian Search algorithm with 1000 iterations, max parent count of 10 and 0,001 prior link probability. The resulting network is shown in Figure 3. It was built with the help of the GeNIe modeler [14].

Interestingly, starting point of the DAG is now R&D expenditure in the business sector (*exp_business*), which has influence on patent applications (*patent*), and through them on design applications (*design*), public-private publications (*PP_pub*) and digital skills (*dig_skills*). The latter in turn influence lifelong learning (*LLL*), non-R&D innovations (*inn_exp*), Employment in innovative SMEs (*emp_SMEs*), SMEs introducing business process innovations (*bp_inn*), Air emissions by fine particulate matter (*air*), and top-10% most cited publications (*most_cited_pub*). The latter variable is also influenced by International scientific co-publications (*cited_pub*) and in turn influences Air pollution, which is shown to have some influence on Employment in innovative SMEs. And the description of causal relations between parent and children nodes could continue further, moving along the DAG from left to right.

Globally, the results of this exercise of model building as represented in Figure 3 are rather different from those in Figure 1. This difference cannot be easily explained with

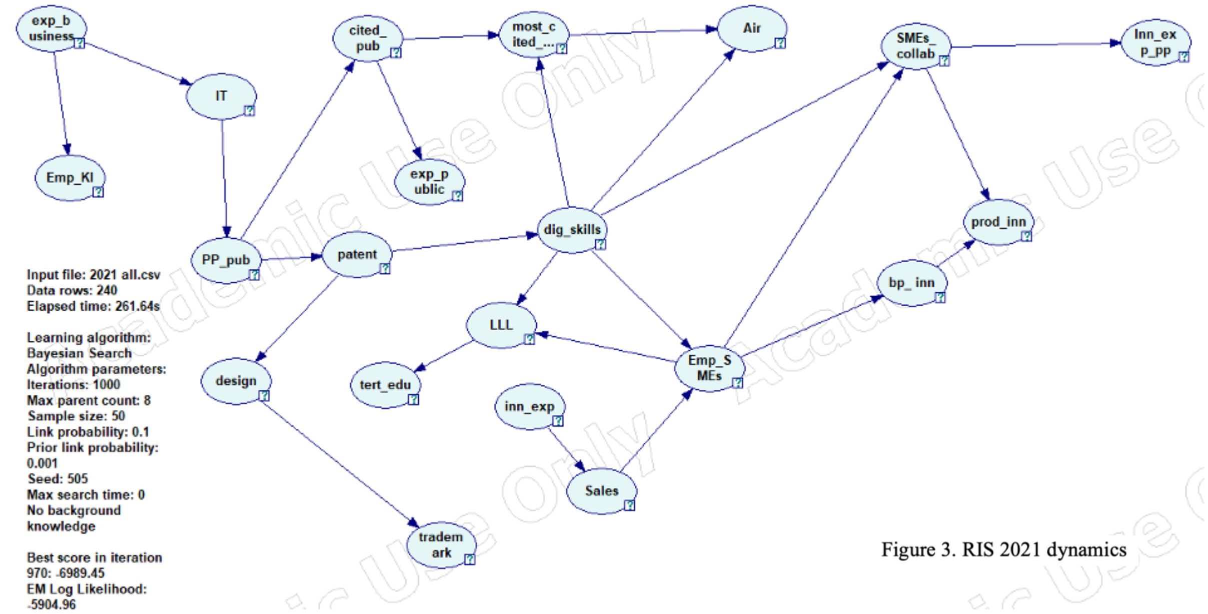


Figure 3. RIS 2021 dynamics

progress of time or with the inclusion/exclusion of some variables in/from the database.

IV. Model simulation

When using Bayesian networks for policy definition, the impacts of changes in a node value on another node in the same network are of great interest.

In fact, one of the key reasons for building a Bayesian network is to fit the conditional probabilities of its nodes. This involves estimating the probability distribution of each node given the values of its parents. The process in question involves collecting data and using it to estimate the probability distributions for each node. Estimating the conditional probabilities can be a complex task, especially for networks with a large number of nodes.

Once the conditional probabilities have been estimated, a Bayesian network can be used to make predictions about the relationships between variables. This can be done by propagating the probabilities throughout the network. As it is shown in Figure 4, the RIS2021 Bayesian network model has defined conditional probabilities and predictions that can be calculated as affecting one indicator and then having influence on others.

To exemplify the approach, we have chosen four RIS variables that can be influenced by policy measures. These are Public R&D expenditure, Expenditure on R&D by

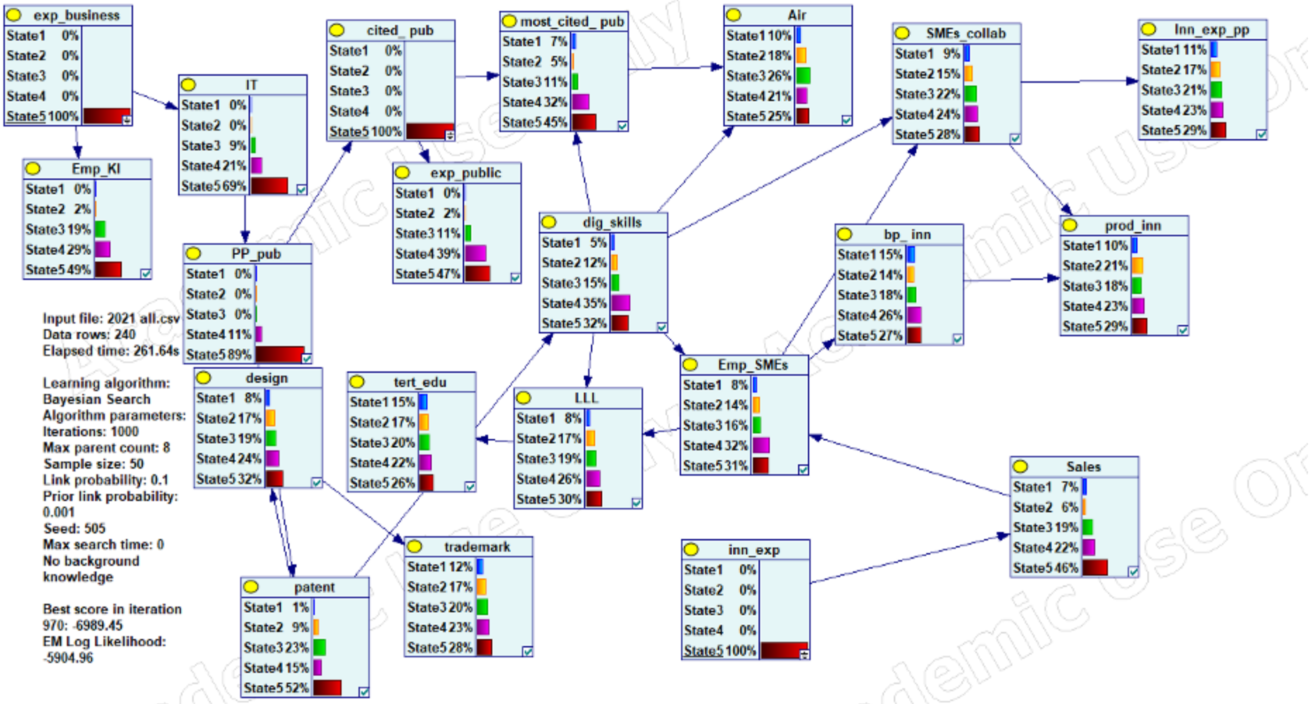


Figure 4 RIS2021 propagating probabilities calculation

business entities, Digital skills and Lifelong learning. Then we simulated the consequences of changing the probability of one indicator and calculated the impact of the change on the other indicators. Some results are displayed in Figure 5.

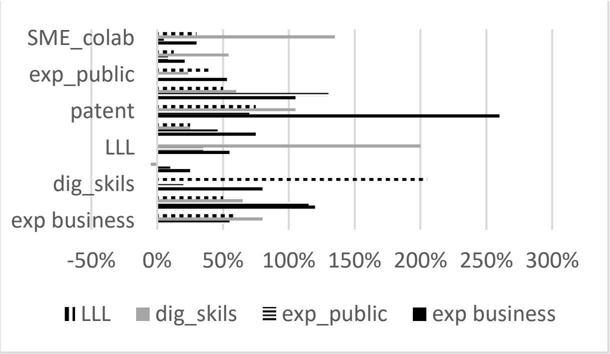


Figure 2. Impacts on indicators when changing another indicator

If one increases public R&D expenditure significantly (e.g. to match the largest expenditure level in the EU), this would increase business R&D expenditure by 55%, patents and publications by 115% and 130%, employment in SMEs by 46%, lifelong learning by 35%, digital skills by 20%, innovation expenditure in SMEs by 8%, employment in knowledge-intensive activities by 10% and collaboration between SMEs by only 5%.

Using policy to boost learning of digital skills would increase public R&D expenditure by 24% and expenditure on R&D in the business sector by 80%. Employment in SMEs would increase by 25%, patent applications by 105%, lifelong learning by 200%, innovation expenditures would grow by 54% and SME collaboration would increase by 135%, while the employment in knowledge-intensive activities would decrease by 5%.

V. Discussion

Bayesian networks seem to be a promising instrument for modelling complex relationships between variables that describe innovation systems. The intuition is to think of Bayesian networks as causal graphs, where every arc (edge) represents a direct causal influence between two connected variables. This view may be too informal and not always mathematically correct, but is widely used by practitioners and can provide a valuable framework for understanding complex systems.

One of the key advantages of using Bayesian networks is that they can help to identify causal factors for certain stochastic variables. A directed arc from X to Y captures the knowledge that X is a causal factor for Y, providing a valuable tool for identifying the most important variables in a system. Additionally, the lack of arcs between pairs of variables can be used to express simple facts about the absence of causal influences between them. This allows for the construction of a data driven model that visualizes all the main causal relationships between variables.

Implementing the proposed approach to the RIS 2019 and RIS 2021 datasets has delivered inconsistent results, both across time and compared with available economic theories and practical experiences. Such inconsistencies are irredeemable, from a purely technical perspective, unless we would consider the possibility of replicating the whole structure learning exercise with the required computing power.

Alternative explanations for the puzzling direction of several causal relations between variables can be manifold: on the one hand, the limited number of observations (only 240 regions compared with 21 indicators) impedes the most obvious corrective action one might consider, that of using RIS data for smaller sized sets of regions – for instance, those belonging to the same geographical areas of Europe or characterised by the same positioning in the ranking (e.g. what the RIS methodology defines “moderate innovators”).

On the other hand, the continuous changes in the data series and metadata definitions do not help consolidate a database with consistent indicators across time, which may be considered as an alternative approach to increasing data size to the required extent. Same goes for moving from a biannual to a yearly publication of RIS statistics, especially for the regional level, which would enable considering time series instead of punctual datasets, and introducing also the time lag dimension, which might be reasonable in a number of cases.

Having more observations and more quality data would be essential to clarify if the heterogeneities that exist within the EU innovation system are too broad to be captured by a single, bird’s eye view cutting across the wide diversities of involved regions and countries.

VI. Conclusion

This paper has proposed an alternative approach to the RIS for analysing the innovation dynamics of a region. The approach is based on the experimental use of Bayesian networks and Machine Learning algorithms to “discover” and visualise causal relations between variables for which RIS indicators exist, without a prior definition of the nature of those variables – such as Innovation Enablers, Firm Activities and Outputs – and therefore in an atheoretical fashion.

To the extent that the proposed approach can be deemed reliable, despite some limitations in computing power and the number of observations, to develop a data driven model predicting the direction of influence between two or more variables, the results of its implementation with both the RIS2019 and RIS2021 datasets has led to counterintuitive results according to both economic theory and experience (or common sense). This raises additional concerns to those expressed in previous literature on the limitations of the RIS tool to assess the dynamics and performance drivers of the EU national/regional innovation systems.

Future work by the authors will include testing again the proposed approach on fewer RIS indicators clustered by typology (for instance, making reference to the triad and quadruplet of elements in earlier and later versions of the methodology). The mapping of results obtained from such exercise would be helpful to shed light on whether a few important datasets are missing, which could be found to act as co-influencers of some variables, or if the direction of some causal relations as (unconvincingly) identified in this preliminary study should instead be reversed.

REFERENCES

[1] European Commission, “Regional Innovation Scoreboard 2021 Methodology Report”. Document date: 14/06/2021. Retrieved online at: https://ec.europa.eu/docsroom/documents/45972

[2] C. Edquist, J. M. Zabala-Iturriagagoitia, J. Barbero, and J. L. Zofío, “On the meaning of innovation performance: Is the synthetic indicator of the Innovation Union Scoreboard flawed?”, Research Evaluation, 27(3), 2018, 196–211.

[3] E.G. Carayannis, Y. Goletsis, and E. Grigoroudis, “Composite innovation metrics: MCDA and the Quadruple Innovation Helix framework”, Technological Forecasting and Social Change 131, 2018, 4-17.

[4] European Commission, “Regional Innovation Scoreboard 2009”. Document date: December 2009. Retrieved online at: https://op.europa.eu/en/publication-detail/-/publication/438811f6-bc27-4049-9872-ad76db87f01e/language-en

[5] European Commission, “Regional Innovation Scoreboard 2021”. Document date: 21 June 2021. Retrieved online at: https://op.europa.eu/en/publication-detail/-/publication/b76f4287-0b94-11ec-adb1-01aa75ed71a1/language-en/format-PDF/source-242412276

[6] A. Darwiche, Modeling and reasoning with Bayesian networks, Cambridge: Cambridge University Press, 2009.

[7] J. Cheng, R. Greiner, J. Kelly, D. Bell, and W. Liu, “Learning Bayesian networks from data: An information-theory based approach”, Artificial Intelligence 137, 2002, 1-2, 43-90.

[8] N. Friedman and D. Koller, “Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks”, Machine Learning 50, 2003, 1-2, 95-125.

[9] J.F. Carriger, M.G. Barron and M.C. Newman, “Bayesian networks improve causal environmental assessments for evidence-based policy”, Environmental Science and Technology 2016, 50, 2.

[10] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, “Using Bayesian networks to analyze expression data”, Journal of Computational Biology 7, 2000, 3/4, 601-620.

[11] P. Dagum and M. Luby, “Approximating probabilistic inference in Bayesian belief networks is NP-hard”, Artificial Intelligence, 60, 1993, 1, 141-153.

[12] D. Heckerman and R. Shachter, “Decision-theoretic foundations for causal reasoning”, Journal of Artificial Intelligence Research, 3, 1995, 405-430.

[13] World Bank, “What are quintiles?” Retrieved online at: https://datahelpdesk.worldbank.org/knowledgebase/articles/1986160-what-are-quintiles, accessed 2.2 2023.

[14] GeNIe Modeler by BayesFusion LLC (academic license). Retrieved online at: www.bayesfusion.com/genie