



Tracing technological shifts: time-series analysis of correlations between patent classes

M. Maragakis^{1,2,3,a} , M. A. Rouni^{1,b} , E. Mouza^{1,c}, M. Kanetidis^{1,d} , P. Argyrakis^{1,2,e} 

¹ School of Physics, Aristotle University of Thessaloniki, Thessaloniki, Greece

² Complexity Center, Aristotle University of Thessaloniki, Thessaloniki, Greece

³ Department of Physics, International Hellenic University, Kavala, Greece

Received: 15 May 2023 / Accepted: 2 August 2023

© The Author(s) 2023

Abstract Patents are used as a reliable indicator for the study of technological evolution in specific fields. Patent citation networks can further enlighten the relation between individual classes of patents that are used to categorize innovation. The tightening or loosening of bonds between a pair of them can point to a changing landscape in either of the two, or in both. It does, however, clearly signal one or more changes. Thus, it is important to point out pairs of classes that undergo processes of this kind, and try to provide plausible explanations for them. We use patent citation data from the European Patent Office to create the time series of all IPC classes. We then examine all pairs of patent classes for correlations, and discuss those which show the greatest increase, or decrease, over time. We identify classes which show both a significant decrease in their correlation with one class and simultaneously an increase with another. We further proceed to check the cross correlations of all pairs in order to identify pairs which show a time lag in following one another. By implementing specific criteria for the selection of the most promising pairs we distinguish some cases which exhibit strong correlation values with time lags of several months (3–10), and for which we can provide a plausible explanation.

1 Introduction

Patents are widely considered to be a major indicator of current technological developments. They are often used as a metric of the research and innovation output of a given society, depicting current technological trends, and indicating its economic potential [1, 2]. It has long been established that a significant amount of time is often needed to advance from the innovation stage to the production one [3]. This is partly due to the fact that advances in production require similarly advanced levels in many different and distinct innovation fields [4]. This translates to a need to understand and track the co-evolution of innovation fields and their cross-correlations.

Such questions, however, have been sufficiently analyzed on a spatial [5, 6] level mainly, and not as much a temporal one. Indeed, the focus of studies on patents and patent citations has mainly centered around the dynamics of specific domains, such as the Internet of Things [7], synthetic biology [8], the water sector [9] etc. There are also papers [10–12] that have tried to identify the dynamics of the patent citation networks as a whole, and understand the hidden processes behind the growth and evolution of the patent citation networks' structure. The effect of patent citations added by patent examiners on such studies should not be overlooked. However, it has been proven to mostly have a spatial and self-citing angle, and not a temporal, cross-sectorial, aspect [13].

In econophysics, biophysics, and many other fields where data in most cases carry a timestamp (which is an extremely useful temporal mark), time-series analysis is very commonly used [14]. It serves as a method to extract information on the evolution of specific quantities, and can be used to understand whether some hidden dependencies exist [15]. For example, it can be used to extract hidden information that may provide hints on the future values of stock market indices [16], or it may be used to describe the metabolism of free-moving mice under controlled ambient temperature [17]. Another issue in time-series analysis is that of representation and dimensionality reduction, for which there have been several different approaches such as [18, 19]. Regarding

^a e-mail: mmara@auth.gr

^b e-mail: arouni@physics.auth.gr

^c e-mail: elmouza@physics.auth.gr

^d e-mail: mkanetid@auth.gr

^e e-mail: panos@auth.gr (corresponding author)

patents and innovation, work has also been done in the field of forecasting future trends that are found when combining time series [20].

To the best of our knowledge, what seems to be lacking in this field is a study on the correlations and cross-correlations between individual time series derived from data coming from technological fields. The time series of citations of a technological field in patents describes the potential this area has on applied innovation. The influence innovations on one field may have on other fields is usually due to new trends that may arise. To cover this, we will try to identify specific innovation fields (as defined in the patent network) which are seemingly unrelated, and which through our research show a high probability of having a well-defined correlation decrease, or increase over time, or exhibit cross-correlation.

The remainder of this study is organized as follows. Section 2 lists the data used for our study. Section 3 describes the method followed to treat our data in order to get results and reach our basic goals, which are listed in Sect. 4. Our study concludes with Sect. 5 which lists our basic observations and our ideas on the future prospects of this work.

2 Data

The International Patent Classification (IPC) system is a hierarchical system used worldwide to categorize patents based on the scientific or technical areas they cover. It comprises eight sections (labeled A–H), each representing a broad field of technology, which are further subdivided into classes, subclasses, groups, and subgroups. An IPC code consists of a combination of letters and numbers. The first letter indicates the section, while the first letter along with the subsequent two digits represent the class. The remaining characters and numbers provide further details on the classification, specifying the subclasses and groups. Each patent is assigned one or more IPC codes based on the specific technical areas it pertains to, making it easier for patent examiners, researchers, and inventors to search for relevant prior art and analyze the patent landscape in specific technology fields.

Our data has been obtained from the Organization for Economic Co-operation and Development (OECD), and covers European Patent Office (EPO) data which span 37 full years, for the period of 1979 to 2015. The network created from the patents' citations, has cited and citing patents as nodes and the citation itself is depicted by a directed link. Each link has a timestamp, allowing for a detailed record of the evolution of link creation within the network.

Our primary focus is on the IPC classes of the cited patents. We tally each patent citation and attribute it to the respective IPC classes the patent is associated with, taking note of the citing date. Thus, even when the cited patent belongs to more than a single class, every citing patent's class adds a citation event (all of equal weight) in each of these classes. Through this process, we generate a time series of the total citations for each of the 124 IPC classes. Since the EPO citation data is not uniformly recorded at consistent time intervals, we aggregated the data on a monthly basis, which yielded approximately 450 time points in each time series.

3 Methodology

We calculate the Pearson correlation coefficient, which is a statistic that measures linear correlation between two variables x and y , where x and y are time series of individual IPC classes [21]. The coefficient value r_{xy} , of the 124×124 pairs of IPC classes for all 37 years of citation measurements are then plotted in a matrix. The correlation values are calculated using the following equation:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

where \bar{x} and \bar{y} are the mean values of x and y , and n is the number of data points in the time series. We can then plot the values in a correlation matrix. The diagonal of such matrices (r_{xy}) has fixed value 1, which is the correlation value of any IPC class time series to itself. Given the previous equation, it is straightforward that correlation coefficient values range from -1 to 1 . However, since pairs are in the order of their IPC classification system, which can be considered as random order in that it is not related to correlations, r_{xy} values are also in the same order. To rearrange them in a more meaningful and visually clear order, and identify some sort of clusters, we use the Hierarchical Binary Cluster Tree (HBCT) method [22].

For this reason, we estimate the correlation distance of the rearranged matrix, δ_{xy} :

$$\delta_{xy} = \sqrt{2(1 - r_{xy})} \quad (2)$$

where r_{xy} is the Pearson correlation coefficient of variables x , y and δ_{xy} is the corresponding distance that takes values $0 < \delta < 2$.

Distance values are preferable over Pearson correlation coefficients, since IPC classes can cluster together when their distance is small. By rearranging the matrix, it is possible to distinguish with greater ease differences between pairs of classes.

A basic correlation analysis would not enable us to confirm the presence or absence of temporal variations in the correlation between different IPC classes. To address this, we employ the "sliding windows method" [23, 24], which allows us to divide the aggregated network into smaller time periods with various starting dates. In particular, we establish a 4-year window and apply a 1-year time step, resulting in a total of 34 consecutive windows. In each window, we calculate the correlation coefficients (r_{xy}) and

keep track of the changes in their values, thereby constructing a time series that represents the evolution of r_{xy} for every pair of IPC classes.

Subsequently, we evaluate the temporal dependence of all these pairs using cross-correlation analysis. This process involves determining whether a displaced similarity exists between the pairs by temporally shifting one of the two time-series forwards or backwards, and by using the same equation (Eq. 1) as above. To ensure more reliable statistical analysis, we set an abstract limit for time-series with low citation counts, which corresponds to an average of approximately 3 citations per month or 1500 citations over the 37-year period.

Additionally, we implement measures to account for any generic trends, such as the increase in citations over time. We apply a ‘pre-whitening’ process to mitigate such concerns, flattening the spectrum [25] and rendering the time series stationary. We examine variance over time, auto-correlation, and seasonality, ensuring that any potential impact on our results is neutralized.

To eliminate the possibility that our findings are due to random statistical occurrences or artifacts of our methodology, we generate a large set of randomized time-series samples by shuffling the existing values randomly. This allows us to compare the outcomes derived from the original real-world data with those from the randomized instances.

4 Results and discussion

The Pearson correlation coefficient methodology yields a unique correlation value per pair, r_{xy} , enabling us to generate a correlation matrix for all IPC class pairs. The resulting values are displayed in Fig. 1a, where each row corresponds to the correlation of each IPC class with all others, including itself on the diagonal. Since IPC class codes are arranged in the IPC classification ordering, there is no evidence of clustering in the correlation matrix.

To shed more light on this, we employ the HBCT method to generate a dendrogram (Fig. 1b), which rearranges all IPC classes according to their existing clustering patterns, using the distance values δ (derived from Eq. 2) as the basis for this reorganization. We then proceed to produce the rearranged correlation matrix (Fig. 1c), where the diagonal consists of coefficients with exact value 1 as it refers to the pairs of each IPC class with itself. With this rearrangement, it becomes more evident that some clustering takes place, as there are some same-colored regions mostly located in the edges, although the majority of IPC classes do not distinctly belong to a specific cluster. This is an indication of a yet hidden relation between individual IPC classes.

It should be noted that correlation values are dynamic from year to year or even from month to month due to seasonal phenomena or significant social, technological and economic events, and, thus, vary greatly. For this reason, we will make time series analysis using the ‘sliding window method’ mentioned in Sect. 3. The width of the window is chosen to be 4 years and the step of the slide forward time is that of a year, thus creating 34 time windows (Fig. 2). However, and given that the first few years have limited recorded patent data on most IPC classes, we omit the first 4 time windows and keep the remaining 30 for all future analysis.

In order to ensure that only those IPC classes with a significant amount of citations are being used, and that correlations between pairs of them are meaningful, we apply a minimum of 1500 overall citations per IPC class. This reduces the amount of classes to 78. Thus, the pairs whose correlations are investigated are $78 \cdot 77 / 2$, which means that we have about 3000 such pairs.

Furthermore, given that we have 30 sliding windows per pair, we obtain 30 correlation coefficient values, one for each window, for each of these pairs. This helps us construct a new time series of the correlations of each pair. As expected, some pairs show strong and continuous correlation throughout their entire time series, while others seem to lose their correlation depending on the period under study (Fig. 3).

We focus on those cases exhibiting a trend with either positive or negative continuous slope. We calculated the least squares slope estimator in simple linear regression. All presented slopes were found statistically significant at the level of 99.9%, with p -value $< 10^{-3}$. As seen in the results (Table 1 and Fig. 4), some pairs have a cumulative change of up to ± 0.6 in the correlation coefficient of those 30 windows.

It is worth noting that there are many more pairs that show a strong systematic decrease in their correlation values than increase. This is mainly due to some specific IPC classes that seem to have lost their correlation to many other classes over time. Looking specifically at one, G21, we see that it regards patents in the field of: “Physics/Nuclear physics; Nuclear engineering”. A simple explanation that may be offered is that nuclear physics materials are used less often in innovation in recent times than a few decades ago. Thus, any association with this class has potentially gradually declined, and loss of correlation to other fields may have ensued.

Furthermore, it is also interesting that IPC classes B60, D04 and F16 exhibit a peculiar behaviour, in that they are seen both in pairs that show a significant increase in the correlation with one class, and a significant decrease with another (or two in the case of F16). The exact description of these classes are:

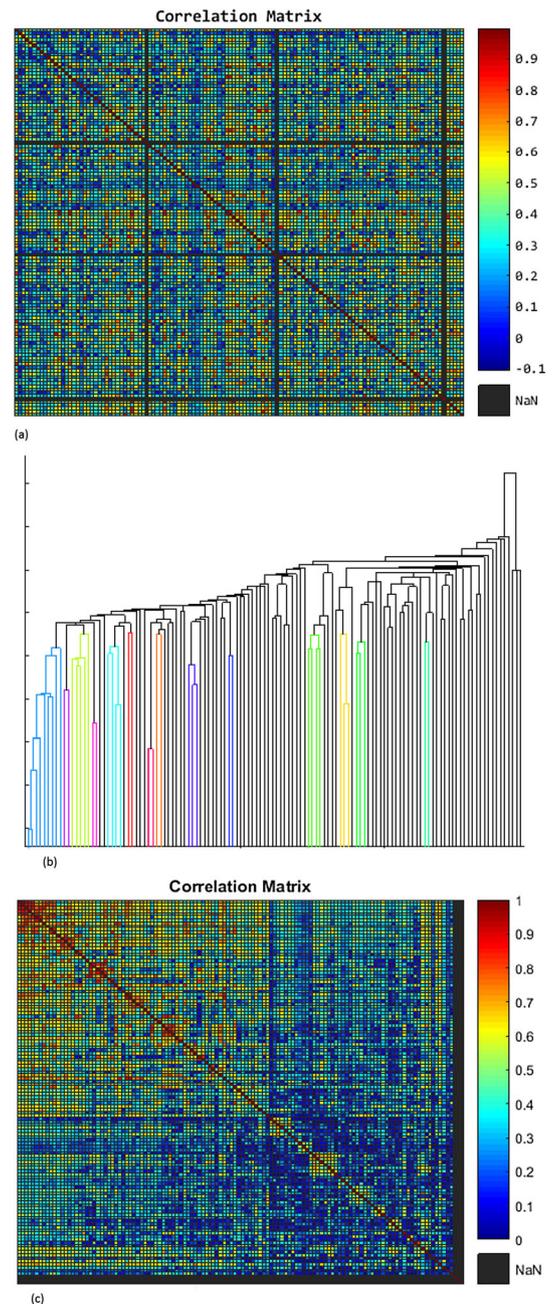
B60: Performing Operations; Transporting/Vehicles in General.

D04: Textiles; Paper/Braiding; Lace-making; Knitting; Trimmings; Non-woven fabrics.

F16: Mechanical engineering; Lighting; Heating; Weapons; Blasting/Engineering elements or units; General measures for producing and maintaining effective functioning of machines or installations; Thermal insulation in general.

Their behaviour may point to a change in the use of innovations on these specific fields over time. For example, D04 undergoes a decrease in its correlation value to A01 and an increase to that of B24. Those classes are agriculture and animal husbandry for the first, and grinding and polishing for the second. This means that it was more common in the past to have innovations in textiles and

Fig. 1 **a** The correlation matrix of all IPC classes, as produced when no ordering is imposed. **b** The dendrogram as produced by the HBCT method, which is used to order the IPC Classes. **c** The resulting ordered correlation matrix of all IPC classes



paper that were related to innovations in agricultural practices, while it may now be more common for innovations in textiles and paper to be registered along with grinding and polishing. A simple observation is that textiles were more related to raw materials produced through animal husbandry 40 years ago than now, and innovations in paper are now more often related to polishing (e.g., glossy papers) and grinding (e.g., paper recycling) than decades ago. Furthermore, agricultural practices and textiles/paper production have been driven further apart due to the advent of smart textiles that are not connected to agriculture. In comparison, properties needed in paper/textiles such as water repellency, antimicrobial activity, or improved strength, can be achieved through innovations in grinding and polishing of modern times.

Similar explanations hold for the behaviour of the correlation of F16 to pairs B62 (increasing correlation), and C07 and C12 (decreasing correlation), where innovation in mechanical engineering is more bound to land vehicles nowadays and is less correlated to innovation in organic chemistry and in brewing.

Next, we focus on understanding whether any hidden relations exist between pairs of IPC classes that are not direct, and can, instead, be found only when there is some time delay. Uncovering this would help identify innovation driving fields (IPC classes) and innovation following ones. Therefore, cross correlation was performed to all possible IPC pairs in order to detect whether

Fig. 2 Sliding time windows method on the patent data. Red and green windows on the left are two consecutive windows. Yellow, blue and red time series are 3 indicative time series of IPC classes showing some of the most extreme behaviour. Yellow being one with many citations per year, red one with very few, blue one with an extreme peak

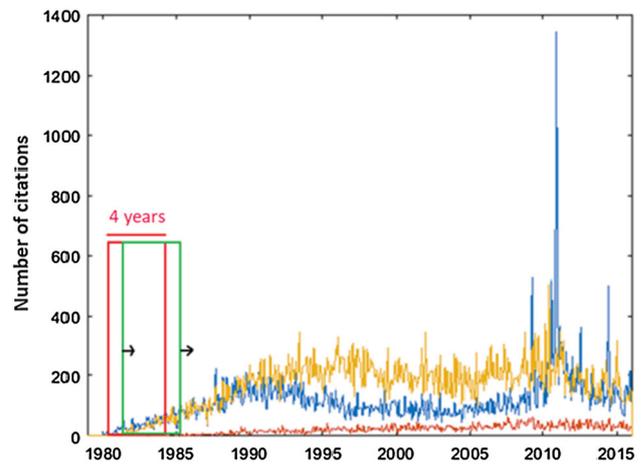
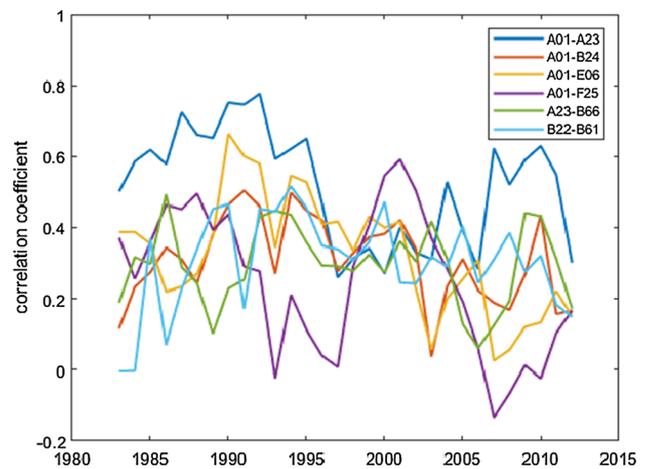


Fig. 3 Indicative correlation values of various IPC class pairs over the 30 time windows. Some pairs exhibit strong and continuous correlation, others exhibit periods or significant correlation, and other periods of no correlation. Year in the x-axis corresponds to the year where each 4-year sliding window starts. For a full definition of all related IPC Codes found in the figure above see Appendix A



such intriguing cases exist. Given that the number of pairs is 3003, and that the cross correlation of a pair for several months was investigated, additional criteria were required to identify the most promising results in a reliable way.

Thus, we applied 3 additional criteria to select the strongest and meaningful cross correlations. Firstly, a threshold at $r_{\max} > 0.75$ was imposed for the maximum cross correlation value. Secondly, a cross correlation coefficient peak which is much higher than the correlation value of the pair was set at $\Delta r = r_{\max} - r_0 > 0.45$. Thirdly, only those cases where the peak was reached at a time delay larger than 1 month were examined, thus, $\Delta t = |t_{\max} - t_0| > 1$ is required. In the above, t_{\max} indicates the time of the peak cross correlation values, t_0 the zero-time delay, r_{\max} indicates the peak cross correlation values, r_0 the value at t_0 , which are essentially the correlation values, as derived before. If these criteria (visually presented at Fig. 5) are fulfilled for a pair of timeseries, then this pair indicates that a variation in the first time series may be the cause for a change in the second one, following some time Δt .

Using this process, it was found that only 68 pairs satisfied the requirements discussed above, the first 20 of which in terms of higher correlation values are shown in Table 2. Furthermore, for the purpose of providing an indicative image of some cases, we also show four such pairs (Fig. 6a–d).

As before, to ensure that such results would not be produced randomly, we shuffle each time series 10,000 times and calculate the cross-correlation values. All the cases presented demonstrate statistical significance at a 99.99% confidence level, with a p -value of 1×10^{-4} .

It is evident that there are some pairs where strong cross-correlation exists. Also, given that these pairs often cover cases between IPC Classes for which it can be surmised by their description that a hidden relation may exist, our results seem fitting. For example, in Fig. 6b there is a pair of IPC classes which have the following exact descriptions:

B65: Performing Operations; Transporting/Conveying; Packing; Storing; Handling thin or filamentary material—and.

C23: Chemistry; Metallurgy/Coating metallic material; Coating material with metallic material; Chemical surface treatment; Diffusion treatment of metallic material; Coating by vacuum evaporation, by sputtering, by ion implantation or by chemical vapour deposition, in general; Inhibiting corrosion of metallic material or incrustation in general.

This pair can easily be thought of as connected, since a new innovation in Coating metallic material, Chemical surface treatment or Coating by vacuum evaporation can have an effect on innovations on packing or handling thin or filamentary material. Our results

Table 1 Selected IPC class pairs with highest and lowest slopes and their respective 95% confidence intervals. The definitions of all presented IPC codes are given in Appendix A

IPC pair	Slope value with error bar	IPC pair	Slope value with error bar
C22–C30	0.0217 ± 0.0044	C07–F02	− 0.0229 ± 0.0085
B61–F01	0.0168 ± 0.0045	C07–F16	− 0.0227 ± 0.0085
B62–H04	0.0156 ± 0.0054	C12–G21	− 0.0219 ± 0.0073
B60–B62	0.0154 ± 0.0055	C07–G21	− 0.0209 ± 0.0068
F21–G08	0.0139 ± 0.0059	C12–F16	− 0.0208 ± 0.0082
B61–H05	0.0137 ± 0.0068	A23–G21	− 0.0198 ± 0.0058
B41–B62	0.0136 ± 0.0042	A01–D04	− 0.0197 ± 0.0076
B24–F01	0.0133 ± 0.0045	G03–G21	− 0.0197 ± 0.0043
B26–B61	0.0131 ± 0.0070	C11–G21	− 0.0194 ± 0.0070
B61–B67	0.0126 ± 0.0053	A01–G21	− 0.0194 ± 0.0063
B24–D04	0.0124 ± 0.0061	C10–G11	− 0.0191 ± 0.0076
B42–F21	0.0121 ± 0.0049	A01–C02	− 0.0190 ± 0.0057
B62–F16	0.0119 ± 0.0050	C12–F02	− 0.0189 ± 0.0095
B61–G01	0.0118 ± 0.0063	A01–F02	− 0.0186 ± 0.0076
B62–G06	0.0117 ± 0.0054	B22–C07	− 0.0184 ± 0.0060
F21–H05	0.0116 ± 0.0051	C04–C07	− 0.0182 ± 0.0045
B62–B65	0.0114 ± 0.0060	A61–G21	− 0.0181 ± 0.0059
A63–E05	0.0112 ± 0.0043	C04–C12	− 0.0178 ± 0.0057
E21–F15	0.0110 ± 0.0061	C07–H03	− 0.0176 ± 0.0082
F04–G07	0.0110 ± 0.0056	B60–C12	− 0.0175 ± 0.0081

point to the fact that there is indeed such a relation, and we even estimate that a period of about 3–4 months is the time needed for the maximum cross correlation value to be achieved.

Similarly, pairs displayed in Fig. 6a, although seemingly unrelated per their definitions (A47 is Furniture; Domestic Articles or Appliances; Coffee Mills; Spice Mills; Suction Cleaners in General, and B62 is Land Vehicles for Traveling Otherwise than on Rails) which display a lag of 8 months could be due to the fact that technological advances in furniture design and related materials or devices, may precede their incorporation into land vehicles, as it may require some time for them to be adopted or adapted for use in the automotive industry. As for Fig. 6c, where the pair again seems initially unrelated (C03 is for Glass; Mineral or Slag Wool, H05 is for Electric Techniques not Otherwise Provided for), and which shows a time lag of 3 months, some assumptions can be made. It is for example reasonable to assume that specialized glass compositions, coatings, or manufacturing processes covered by C03, may have implications for electrical applications of H05, and the time lag may be due to just that. Finally, for Fig. 6d, a pair where it is again not easy to find direct relations (B62 is for Land Vehicles for Traveling Otherwise than on Rails, and G01 is for Measuring; Testing), there is a time lag of 9 months. This can perhaps be related to innovations in measuring and testing technologies, such as advanced sensors, diagnostic tools, or testing methodologies covered by G01, which may be developed and patented before their incorporation into land vehicles of B62.

5 Conclusion

Patents are one of the most common established indicators of innovation, and can often amount to a hidden market value of several billions for large tech companies. For example, in a recent patent auction Yahoo had sold around 3000 of its patents for an estimated price of 1 billion US dollars. Companies such as Apple or Samsung have a considerably higher patent portfolio value. The value alone, however, cannot determine the way a specific innovation impacts others in related, or seemingly unrelated, fields. Thus, we have tried to find methods to understand the effects that patent classes can have on each other, so as to predict with high likelihood the changes in one class when observed in another.

Our methodology which is based on time series analysis, and the results obtained, verify that in some cases there is significant, and statistically improbable, change in the correlation of some pairs of IPC classes. In fact, we have clearly identified several cases where there is a gradual (or in some cases more abrupt) change in the correlation values between the time series of a pair of classes. We have provided an explanation for this change, and we have identified specific classes whose patent content has probably changed over time, in relation to other classes. As an example, we mentioned the case of class G21 which may nowadays have a smaller technological impact in innovation overall. We further found classes such as B60, D04 and F16 that have both significantly increased their correlations to some classes and decreased it with others. This can be attributed to a change in the use of innovations of each IPC class, especially when compared to its correlated class.

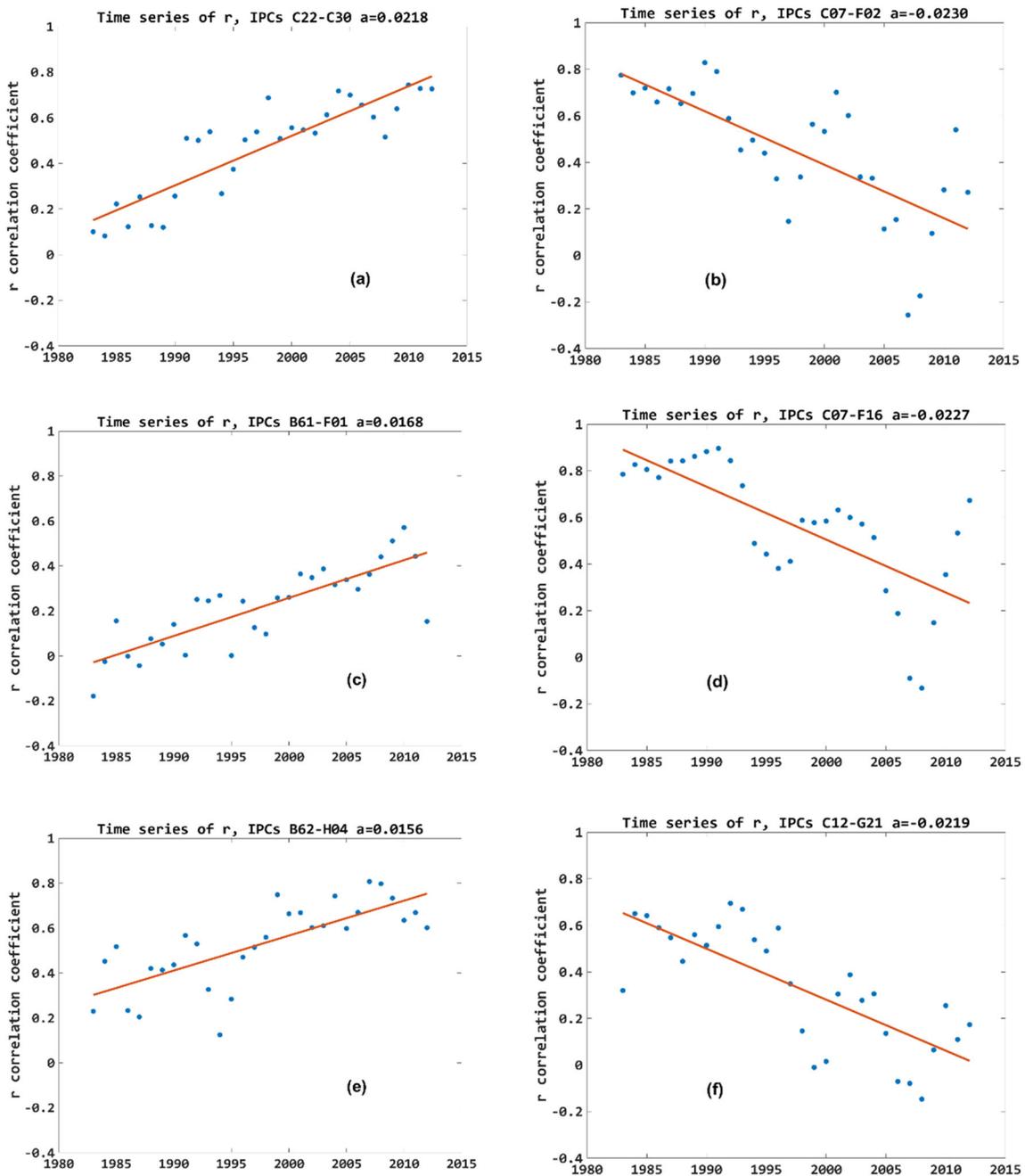


Fig. 4 Correlation coefficient values of various pairs over each sliding window. Each subfigure corresponds to the correlation of a specific pair. The pairs selected exhibit the highest and lowest slope values. Pairs on the left correspond to the highest slope, on the right to the lowest. **a** C22–C30, **b** C07–F02, **c** B61–F01, **d** C07–F16, **e** B62–H04, **f** C12–G21. Year in the x-axis is the same as in Fig. 3. Full definition of IPC codes can be found in Appendix A

The methodology also identifies whether there is a temporal lag between the correlation of two IPC classes, and estimates the strength of this cross correlation. We have chosen through the application of additional criteria, some example cases where there appears to be significant cross correlation. One such case is that of the pairs mentioned in the results section between B65 and C23. We have also provided a plausible explanation for all four such cases shown in the related figure, based on some of the specific subfields that exist in these classes.

Fig. 5 Visual representation of the criteria for the selection of the most promising pairs with significant cross correlation peaks in their time delays

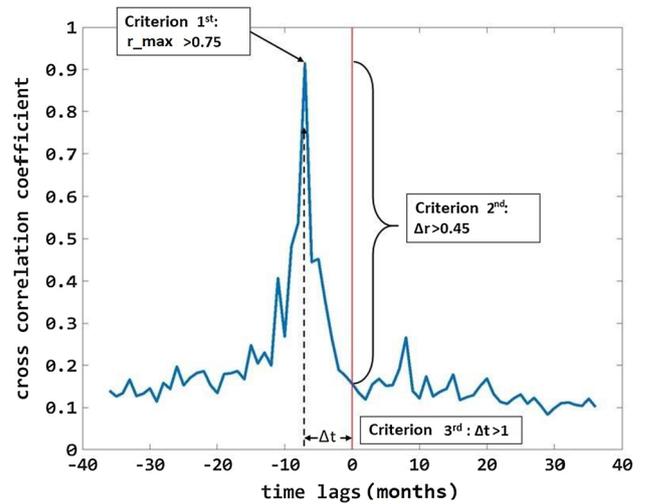


Table 2 Cross-correlation coefficients of 20 pairs, ordered by the highest cross-correlation values, and time lag of the max (peak) value. Positive time lag values signify that IPC Code 2 is ahead of IPC Code 1, and negative values the opposite. Full definition of IPC codes can be found in Appendix A

IPC class 1	IPC class 2	Cross-correlation value	Time lag (in months)
B32	H01	0.93	10
A61	H01	0.92	-2
B28	B62	0.91	-7
A47	B62	0.88	-8
B29	B65	0.88	-3
B62	H02	0.88	10
D06	F25	0.88	-6
F16	G08	0.88	-4
B62	G06	0.87	9
D06	H05	0.87	-3
B42	F23	0.86	4
B62	G01	0.86	9
B62	G07	0.86	9
C03	H05	0.86	-3
F23	F28	0.86	-5
A63	B32	0.85	9
B29	B62	0.85	-9
B32	F16	0.85	3
B62	C23	0.85	9
E06	G10	0.85	4

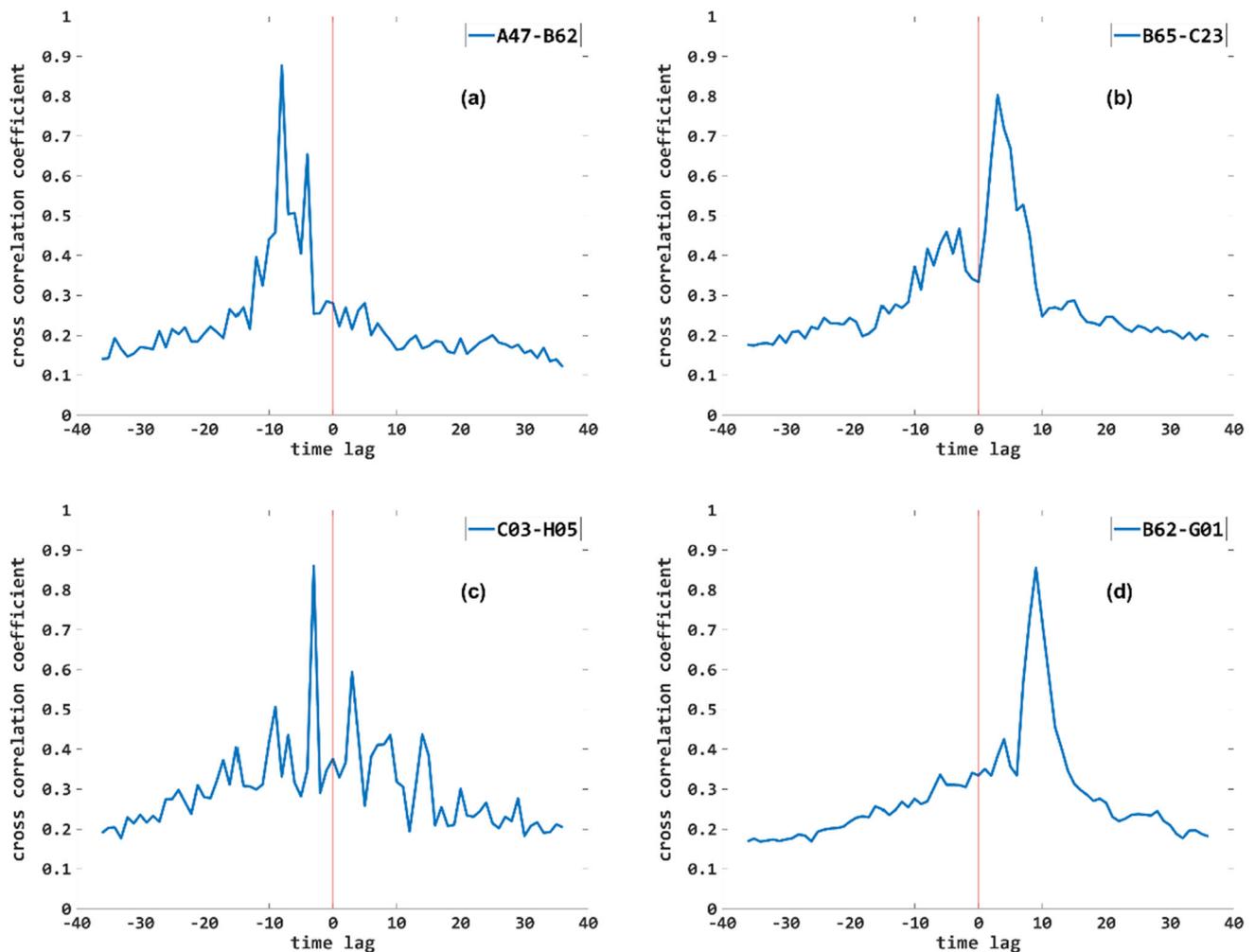


Fig. 6 Cross correlation of specific pairs over their time lags (in months). Pairs with some of the highest cross correlation values that cover all previously set criteria. Full definition of IPC codes can be found in Appendix A

Our approach lacks the ability to identify cases where there are peaks for only short periods of time in the correlation values over the span of the 34 years of data. It also cannot identify pairs where cross correlation may be more complex, as would be the case if both classes were to influence one another after some time lag. Future work could focus on finding more robust ways to identify such changes, as well as set indisputable criteria to identify and specify exact cross correlation temporal lags between pairs.

Funding Open access funding provided by HEAL-Link Greece. No funding was received for conducting this study.

Data Availability Statement This manuscript has associated data in a data repository. [Authors' comment: The data that support the findings of this study are available from OECD but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of OECD.]

Declarations

Conflict of interest The authors have no competing interest to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A

See Table 3.

Table 3 IPC Codes

IPC code	IPC class title
A	Human Necessities
A01	Agriculture; Forestry; Animal Husbandry; Hunting; Trapping; Fishing
A23	Foods or Foodstuffs; Their Treatment, not Covered by Other Classes
A47	Furniture; Domestic Articles or Appliances; Coffee Mills; Spice Mills; Suction Cleaners in General
A61	Medical or Veterinary Science; Hygiene
A63	Sports; Games; Amusements
B	Performing Operations; Transporting
B22	Casting; Powder Metallurgy
B24	Grinding; Polishing
B26	Hand Cutting Tools; Cutting; Severing
B28	Working cement, clay, or stone
B29	Working of plastics; Working of substances in a plastic state in general
B32	Layered Products
B41	Printing; Linotype; Typewriters; Stamps
B42	Bookbinding; Albums; Files; Special Printed Matter
B60	Vehicles in General
B61	Railways
B62	Land Vehicles for Traveling Otherwise than on Rails
B65	Conveying; Packing; Storing; Handling Thin or Filamentary Material
B67	Opening or Closing Bottles, Jars or Similar Containers; Liquid Handling
C	Chemistry; Metallurgy
C02	Treatment of Water, Waste Water, Sewage, or Sludge
C03	Glass; Mineral or Slag Wool
C04	Cement, Concrete, Artificial Stone, Ceramics; Refractories
C07	Organic Chemistry
C10	Petroleum, Gas or Coke Industries; Technical Gases Containing Carbon Monoxide; Fuels; Lubricants; Peat
C11	Animal or Vegetable Oils, Fats, Fatty Substances or Waxes; Fatty Acids Therefrom; Detergents; Candles
C12	Biochemistry; Beer; Spirits; Wine; Vinegar; Microbiology; Enzymology; Mutation or Genetic Engineering
C22	Metallurgy; Ferrous or Non-Ferrous Alloys; Treatment of Alloys or Non-Ferrous Metals
C23	Coating Metallic Material; Coating Material with Metallic Material; Surface Colloidal Coating; Vacuum Evaporation; Electrolysis; Ion-Implantation
C30	Crystal Growth
D	Textiles; Paper
D04	Braiding; Lace-Making; Knitting; Trimmings; Non-Woven Fabrics
D06	Treatment of textiles or the like; Laundering; Flexible materials not otherwise provided for
E	Fixed Constructions
E05	Locks; Keys; Window or Door Fittings; Safes
E06	Doors, windows, shutters, or roller blinds, in general; Ladders
E21	Earth or Rock Drilling; Mining
F	Mechanical Engineering; Lighting; Heating; Weapons; Blasting
F01	Machines or Engines in General; Engine Plants in General; Steam Engines

Table 3 continued

IPC code	IPC class title
F02	Combustion Engines; Hot-Gas or Combustion-Product Engine Plants
F04	Positive Displacement Machines for Liquids; Pumps for Liquids or Elastic Fluids
F15	Fluid-Pressure Actuators; Hydraulics or Pneumatics in General
F16	Engineering Elements or Units; General Measures for Producing and Maintaining Effective Functioning of Machines or Installations; Thermal Insulation in General
F21	Lighting
F23	Combustion Apparatus; Combustion Processes
F25	Refrigeration or cooling; Combined heating and refrigeration systems; Heat pump systems; Manufacture or storage of ice; Liquefaction or solidification of gasses
F28	Heat Exchange in General
G	Physics
G01	Measuring; Testing
G03	Photography; Cinematography; Analogous Techniques Using Waves Other than Optical Waves; Electrography; Holography
G06	Computing; Calculating; Counting
G07	Checking Devices
G08	Signaling
G10	Musical Instruments; Acoustics
G11	Information Storage
G21	Nuclear Physics; Nuclear Engineering
H	Electricity
H01	Basic Electric Elements
H02	Generation, conversion, or distribution of electric power
H03	Basic Electronic Circuitry
H04	Electric Communication Technique
H05	Electric Techniques not Otherwise Provided for

References

1. L. Soete, The impact of technological innovation on international trade patterns: the evidence reconsidered. *Res. Policy* **16**(2–4), 101–130 (1987)
2. S. Nagaoka, K. Motohashi, A. Goto, Patent statistics as an innovation indicator, in *Handbook of the Economics of Innovation, Volume 2*. (Elsevier, 2010), pp.1083–1127. [https://doi.org/10.1016/S0169-7218\(10\)02009-5](https://doi.org/10.1016/S0169-7218(10)02009-5)
3. H.-D. Hausteiner, E. Neuwirth, Long waves in world industrial production, energy consumption, innovations, inventions, and patents and their identification by spectral analysis. *Technol. Forecast. Soc. Change* **22**(1), 53–89 (1982)
4. C.L. Benson, C.L. Magee, Quantitative determination of technological improvement from patent data. *PLoS ONE* **10**(4), e0121635 (2015)
5. M.M. Fischer, D.A. Griffith, Modeling spatial autocorrelation in spatial interaction data: an application to patent citation data in the European union. *J. Region. Sci.* **48**(5), 969–989 (2008)
6. C. Yang, S. Liu, Spatial correlation analysis of low-carbon innovation: a case study of manufacturing patents in China. *J. Clean. Prod.* **273**, 122893 (2020)
7. L. Ardito, D. D’Adda, A.M. Petruzzelli, Mapping innovation dynamics in the internet of things domain: evidence from patent analysis. *Technol. Forecast. Soc. Chang.* **136**, 317–330 (2018)
8. D. van Doren, S. Koenigstein, T. Reiss, The development of synthetic biology: a patent analysis. *Syst. Synth. Biol.* **7**(4), 209–220 (2013). <https://doi.org/10.1007/s11693-013-9121-7>
9. L. Compagnucci, F. Spigarelli, Fostering cross-sector collaboration to promote innovation in the water sector. *Sustainability* **10**, 4154 (2018)
10. P. Érdi, K. Makovi, Z. Somogyvári, K. Strandburg, J. Tobochnik, P. Volf, L. Zálányi, Prediction of emerging technologies based on analysis of the US patent citation network. *Scientometrics* **95**(1), 225–242 (2013)
11. K. Angelou, M. Maragakis, K. Kosmidis, P. Argyrakis, A hybrid model for the patent citation network structure. *Phys. A* **541**, 123363 (2020)
12. K. Angelou, M. Maragakis, P. Argyrakis, A structural analysis of the patent citation network by the k-shell decomposition method. *Phys. A* **521**, 476–483 (2019)
13. J. Alcácer, M. Gittelman, Patent citations as a measure of knowledge flows: the influence of examiner citations. *Rev. Econ. Stat.* **88**(4), 774–779 (2006)
14. W.A. Fuller, In *Introduction to Statistical Time Series* (2009)
15. T.-C. Fu, A review on time series data mining. *Eng. Appl. Artif. Intell.* **24**(1), 164–181 (2011)
16. J.K. Shareef, *Prediction of International Stock Market Movements Using a Statistical Time Series Analysis Method* (United Scholars Publication, 2015)
17. G.A. Sunagawa, M. Takahashi, Hypometabolism during daily torpor in mice is dominated by reduction in the sensitivity of the thermoregulatory system. *Sci. Rep.* **6**(1), 37011 (2016)
18. M. Ali, M.W. Jones, X. Xie, M. Williams, TimeCluster: dimension reduction applied to temporal data for visual analytics. *Vis. Comput.* **35**(6–8), 1013–1026 (2019)
19. E. Eldele, M. Ragab, Z. Chen, M. Wu, C. K. Kwok, X. Li, C. Guan, Time-series representation learning via temporal and contextual contrasting. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence* (2021)

20. M. Chakraborty, S.A. Bahrainian, F. Crestani, Forecasting patent growth by combining time-series signals using covariance patterns. In *Proceedings of the First Joint Conference of the Information Retrieval* (2020)
21. K. Pearson, VII. Note on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.* **58**(347–352), 240–242 (1895)
22. R.R. Sokal, C.D. Michener, *A Statistical Method for Evaluating Systematic Relationships* (University of Kansas Science Bulletin, 1958)
23. M. Datar, A. Gionis, P. Indyk, R. Motwani, Maintaining stream statistics over sliding windows. *SIAM J. Comput.* **31**(6), 1794–1813 (2002)
24. K. Angelou, M. Maragakis, K. Kosmidis, P. Argyrakis, Dynamics of regional multilinks in research innovation temporal networks. *EPL (Europhys. Lett.)* **130**(2), 28001 (2020)
25. C. Chatfield, H. Xing, *The Analysis of Time Series: An Introduction with R*, 7th edn. (Chapman & Hall, 2019)