Jointly published by Akadémiai Kiadó, Budapest
and Kluwer Academic Publishers, Dordrecht

Scientometrics,
Vol. 56, No. 1 (2003) 000–000

Monitoring elasticity between science and technology domains and its visualization

FILIP DELEUS, MARC M. VAN HULLE

Laboratory of Neuro- and Psychophysiology, Katholieke Universiteit Leuven, Leuven (Belgium)

We introduce a new technique for quantifying and monitoring the effect a given set of time series has on the evolution of a single time series. The technique relies on the causal nature of this effect, and expresses the result in terms of partial and cross elasticities. As an application, we consider the case where the single time series consists of the number of patents filed over time, in a given category, and where the set of time series consists of the numbers of scientific articles published over time, for each one of a number of science domains. Finally, we use a quiver map for visualizing the elasticities and as a case study we illustrate our methodology on patents in the field of Biotechnology.

Introduction

As suggested in previous studies, the interaction between industrial and academic research can be examined by analyzing the joint evolution of the time series of the filed patents and the scientific articles published in a given time span.[1,4,6-8] However, the time series were carefully selected by experts prior to the analysis, and the analysis of the joint evolution relies on a visual assessment of the changes in the time series.

The purpose of this article is to extend this analysis of the joint evolution in two ways. First, we perform our analysis without having to rely on a prior selection of the time series. Second, we go beyond a mere visual assessment by introducing a technique that models the joint evolution in terms of *partial* elasticities. Our ambition is to build models that are causally plausible. For example, if there is an increase in filed patents, then this should be due to an increase in published articles in the past, when we assume that scientific developments are a precursor to technological developments. Furthermore, in addition to partial elasticity, we also introduce the concept of *cross* elasticity. Consider two science domains and one technology domain. Cross elasticity is now modeling the effect of the number of articles published in the first science domain

Received July 18, 2002.

Address for correspondence:
FILIP DELEUS
Laboratory of Neuro- and Psychophysiology, Katholieke Universiteit Leuven
Campus Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium
E-mail: filip@neuro.kuleuven.ac.be

0138–9130/2002/US $ 15.00

on the number of patents which refer to articles from the second science domain and which are filed in the given technology domain. Finally, we introduce a quiver map for visualizing both types of elasticities over time.

Patents- and scientific publications databases

We have used the US Patent and Trademark Office (USPTO) patent database and the *Science Citation Index* (SCI) scientific publication database, produced by the Institute for Scientific Information (ISI). The USPTO database lists the number of filed patents in a given category per year, and the references to articles published in scientific journals. Each patent filed in a given year is classified into one of several technology domains. Each technology domain receives a separate IPC code, an 8 digit code. It is standard practice to consider the first 4 digits only (for more information, see Ref. 8). Furthermore, we ignore the references to scientific articles published after the patent has been filed, since they were added after the filing date: we are interested in the articles that led to the actual filing of the patent.

The SCI database contains the titles of the scientific articles, the authors, the journals where the articles appeared, including volume numbers and page numbers. Each scientific journal is classified into a science domain to which corresponds an SCI code, a 3-digit number. Hence, we can take a single time series from the USPTO database and a set of time series from the SCI database, and model the effect the set has on the evolution of the USPTO time series. This should then provide us with an indication of the influence science developments have on technological developments.

For example, assume that we count per year all scientific articles published with the same SCI code. Similarly, we can count per year all patents filed with same IPC code. The resulting time series can then be plotted. For example, in the left panel in Figure 1, the time series of technology domain A01B (‘Soil working or forestry; parts, details, or accessories of agricultural machines or implements, in general’) and science domain 270 (‘Agriculture’) are plotted. Based on visual inspection, one can remark a joint evolution of the two time series and conclude that there is an effect of the science domain on the technology domain. In the right panel of Figure 1, the time series of technology domain A43B (‘Characteristic features of footwear; parts of footwear’) and science domain 297 (‘Dentistry and odontology’) are plotted and based on visual inspection, one can remark that the changes in the science domain are followed by analogue changes in the technology domain after one period. However this apparent joint evolution is an artifact since the patents in technology domain A43B do not refer to the articles in science domain 297. In order to avoid such artifacts, we introduce the concepts of partial time series and of partial- and cross elasticities in the next two sections, respectively.

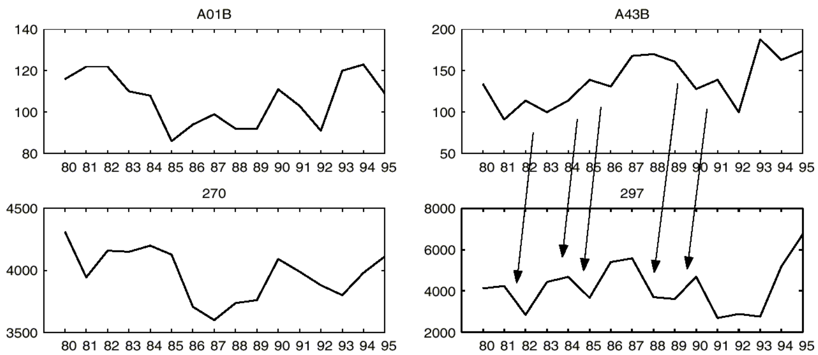

Figure 1. Panels in left column: number of patents filed in technology domain A01B (top) and articles published in science domain 270 (bottom) as a function of time. Panels in right column: idem, but for patents filed in technology domain A43B (top) and articles published in science domain 297 (bottom). Arrows indicate time instances where the simultaneous increase and decrease in both time series is artifactual. See text

Partial time series

Before we can introduce our technique, we first need a number of definitions. Let $NPAT(td_i,t)$ be the number of patents of technology domain td_i, filed in year t. Similarly, let $NPUB(sd_i,t)$ be the number of publications of science domain sd_i, published in year t. Assume now that we take a patent from technology domain td_i, filed in year t. If that patent refers N times to an article from science domain sd_k, and also M times to all other articles of other science domains, then we count the reference to sd_k as $N/(N+M)$. When we repeat this procedure over all patents in technology domain td_i, and summate all the resulting ratios, then we obtain the quantity $NPAT_{fract1}(td_i,t,sd_k)$.

In order to model causal effects between science and technology, we further introduce two extensions of the previous definition. First, assume we take a patent from technology domain td_i, filed in yeart. When that patent refers N times to an article of science domain sd_k, published *no later* than t_{sd_k}, and M times to all other articles from science domain sd_k, published *later* than t_{sd_k}, or to articles from other science domains (irrespective of their publication date), then we obtain the ratio *N*/(*N*+*M*). When we repeat this for each patent in technology domain td_i, filed in year t, and summate the ratios, we obtain the quantity $NPAT_{fract2}(td_i, t, sd_k, t_{sd_k})$.

Second, assume again that we take a patent from technology domain td_i, filed in year t. If that patent refers $N1$ times to an article of science domain sd_k, published *no later* than t_{sd_k}, $N2$ times to an article from science domain sd_j, irrespective of its publication date, and M times to all other articles, then we count the reference to sd_j, with respect to sd_k and t_{sd_k}, as $N1.N2/\{(N1+N2+M)^2\}$. When we repeat this for each patent in technology domain td_i, filed in year t, and summate the ratios, we obtain the quantity $NPAT_{fract3}(td_i, t, sd_k, t_{sd_k}, sd_j)$.

Finally, we note that the following relationships hold between the definitions of $NPAT$ and $NPAT_{fract}$:

$$NPAT(td_i, t) = \sum_k NPAT_{fract1}(td_i, t, sd_k) \tag{1}$$

$$NPAT_{fract1}(td_i, t, sd_k) = NPAT_{fract2}(td_i, t, sd_k, t) \tag{2}$$

$$NPAT_{fract2}(td_i, t, sd_k, t_{sd_k}) = \sum_j NPAT_{fract3}(td_i, t, sd_k, t_{sd_k}, sd_j) \tag{3}$$

Partial and cross elasticities

One way to model the effect of the publication of scientific articles on the filing of patents is to consider it in terms of an elasticity: the relative change in the number of filed patents as a result of a relative change in the number of articles published, after a certain delay. We can further specify this elasticity and express it as a function of time, and of science and technology domains. This leads to the following tentative definition:

$$\varepsilon_{td_i, sd_j, t, delay} = \frac{\dfrac{NPAT(td_i, t+delay) - NPAT(td_i, t-1)}{NPAT(td_i, t-1)}}{\dfrac{NPUB(sd_j, t) - NPUB(sd_j, t-1)}{NPUB(sd_j, t-1)}} \tag{4}$$

However, since the patents in technology domain td_i not necessarily refer to articles in science domain sd_j, this definition of elasticity could lead to erroneous results.

What we need is an elasticity that captures the relative change in technological output in domain td_i between year t-1 and year t+*delay*, which can be due to a change in scientific output in domain sd_j between year t-1 and year t. More precisely, we define the change in technological output as $NPAT_{fract2}(td_i,t+delay,sd_j,t)-NPAT_{fract2}(td_i,t-1,sd_j,t-1)$. This difference is, given a certain technology domain td_i and a certain science domain sd_j, the change in the fractional number of patents from year t-1 to year t+*delay*, where the patents refer to articles published no later than in year t-1 and year t, respectively. We relate this quantity to the total number of patents in td_i filed in year t-1, i.e., $NPAT(td_i,t-1)$. This leads to the concept of *partial* elasticity of which a plausible definition is:

$$\varepsilon_{td_i,sd_j,t,delay} = \frac{\dfrac{NPAT_{fract2}(td_i,t+delay,sd_j,t)-NPAT_{fract2}(td_i,t-1,sd_j,t-1)}{NPAT(td_i,t-1)}}{\dfrac{NPUB(sd_j,t)-NPUB(sd_j,t-1)}{NPUB(sd_j,t-1)}} \tag{5}$$

The partial elasticity measures the effect of a change in the number of articles published in science domain sd_j between year t-1 and year t, on the number of patents filed in technology domain td_i in year t + *delay*. When calculating $\varepsilon_{td_i,sd_i,t,delay}$ for different td_i and sd_j values, we obtain the following partial elasticity matrix:

$$\varepsilon_{T,S,t,delay} = \begin{pmatrix} \varepsilon_{td_1,sd_1,t,delay} & \varepsilon_{td_2,sd_1,t,delay} & \cdots \\ \varepsilon_{td_2,sd_1,t,delay} & \varepsilon_{td_2,sd_2,t,delay} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{6}$$

Evidently, this matrix can be determined for different combinations of times t and delays *delay*, however, this will quickly result in an explosion of elasticity values. In order to obtain an overview, we introduce the following graphical technique. We calculate $\varepsilon_{td_i,sd_j,t,delay}$ for different delays *delay*, but keep time t constant, and select the maximal ε-value. We then enter this maximum in the (i,j)th entry in a new matrix. This new matrix will then reflect the maximal elasticities reached over different delays. This matrix can then be visualized as follows. We represent each entry in this new matrix as a vector ("quiver") with length equal to the logarithm of the absolute value of the entry, and with angle proportional to the magnitude of the delay. A horizontal quiver corresponds to a zero delay; a non-zero delay results in a counter-clockwise rotation of the quiver such that a vertical quiver corresponds to a delay of 10 years. When the maximal ε-value is negative, we flip the quiver by rotating it over 180 degrees. When the quiver has zero length, we put a dot instead. The result is called a quiver map, an example of which will later be shown in Figure 6 for the case of patents filed in the field of Biotechnology. The rows correspond to science domains and the columns to technology domains. The modeled effect is thus from a row element onto a column element.

Consider now two science domains and one technology domain. We can now model the effect of a change in the number of articles in the first science domain on the number of patents which refer to articles from the second science domain. This effect is captured by what we call *cross elasticity*.

It is defined as follows. Define first the change in the number of filed patents as the difference between $NPAT_{fract3}(td_i,t+delay,sd_j,t,sd_k)$ and $NPAT_{fract3}(td_i,t\text{-}1,sd_j,t\text{-}1,sd_k)$. This is the change in the fractional number of patents from year t-1 to year $t+delay$ where the patents refer to articles from science domain sd_j published no later than in year t-1 and year t respectively and where the patents refer to articles from science domain sd_k, irrespective of their publication date. In order to obtain the relative change, this quantity is divided by $NPAT_{fract2}(td_i,t\text{-}1,sd_j,t\text{-}1)$, by virtue of Eq. 3. The cross elasticity then becomes:

$$\varepsilon_{td_i,sd_j,sd_k,t,delay} = \frac{\dfrac{NPAT_{fract3}(td_i,t+delay,sd_j,t,sd_k) - NPAT_{fract3}(td_i,t-1,sd_j,t-1,sd_k)}{NPAT_{fract2}(td_i,t-1,sd_j,t-1)}}{\dfrac{NPUB(sd_j,t) - NPUB(sd_j,t-1)}{NPUB(sd_j,t-1)}} \tag{7}$$

Thus, the cross elasticity measures the effect of a change in the number of articles published in science domain sd_j between year t-1 and year t, on the number of patents filed in technology domain td_i in year $t+delay$, which use articles taken from science domain sd_k. When calculating this for different sd_j and sd_k, we obtain the following cross elasticity matrix:

$$\varepsilon_{td_i,t,delay} = \begin{pmatrix} \varepsilon_{td_i,sd_1,t,delay} & \varepsilon_{td_i,sd_1,sd_2,t,delay} & \cdots \\ \varepsilon_{td_i,sd_2,sd_1,t,delay} & \varepsilon_{td_2,sd_2,sd_2,t,delay} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{8}$$

As was the case with the partial elasticity matrix, the cross elasticity matrix can also be determined for different (*t,delay*)-combinations, and the result summarized and visualized with the quiver map. The rows correspond to the science domains of which the causal effects onto the science domains listed in the columns are modeled. The modeled effect is thus again from a row element onto a column element.

Results

As an example, we take patents from the USPTO database which all belong to technology area Biotechnology. Following the classification of *Grupp* and *Schmoch*[3] and its updated version, which is also used by the Observatoire des Sciences et des Technologies (OST) and the Institut national de la propriete industrielle (INPI), the technology area Biotechnology has been defined as the collection of 7 IPC 4-digit classes: C07G, C12M, C12N, C12P, C12Q, C12R and C12S. We illustrate our analyses on these technology domains.

The most important IPC class in terms of number of patents is technology domain C12N. In Figure 2, the number of patents filed in domain C12N is plotted as a function of time. We rank the science domains by the number of references made to them by patents in the Biotechnology area. The most important science domains in biotechnology patents are the science domains with SCI codes 279, 350, 347, 329, 320, 394, 282, 284, 296, 344. The time series of the two most important science domains are plotted in Figure 3. The evolution in the fractional number of patents in technology domain C12N with respect to science domains 279 and 350, i.e., $NPAT_{fract1}$(C12N,t,279) and $NPAT_{fract1}$(C12N,t,350) is plotted in Figure 4. Next as an illustration of the delay between publications in a certain science domain and the number of patents refering to these publications, we plot $NPAT_{fract2}$(C12N,t,279,1985) and $NPAT_{fract2}$(C12N,t,350,1985) in Figure 5, being the evolution in the fractional number of patents which refer to articles published no later than in year 1985 in science domains 279 and 350 respectively.

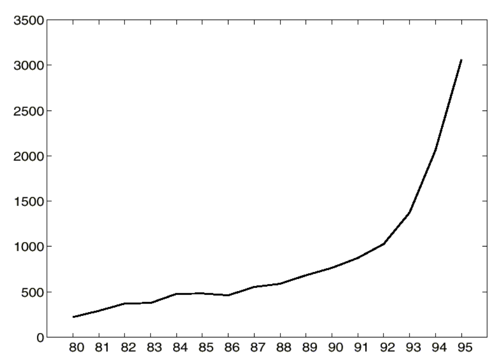

Figure 2. Temporal evolution of number of patents filed in technology domain C12N

The partial elasticities for year 1985 between the technology domains belonging to the Biotechnology area and their most important linked science domains are calculated as defined by Eq. 5 and their summaries are plotted as a quiver map in Figure 6. We observe that the most cited science domains also have the largest relative effects, but that there are differences in the delays of these effects. For example, the more horizontal quiver between science domain 350 and technology domain C12N illustrates that the effect from this science domain on that technology domain is quicker than the effect of the science domain 279 on the same technology domain.

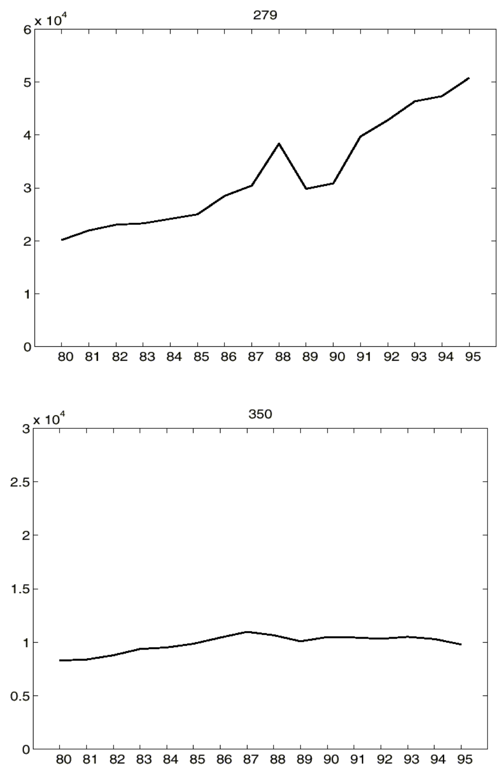

Figure 3. Temporal evolution of number of articles published in science domains 279 (above) and 350 (below)

Finally, we analyze technology domain C12N, and show the time series of $NPAT_{fract3}$(C12N,t,279,1985,279) and $NPAT_{fract3}$(C12N,t,279,1985,350) in Figure 7. The quantities $NPAT_{fract3}$ are used to calculate the cross elasticities as defined by Eq. 7. These elasticities, which measure how one science domain exerts an effect on the fractional number of patents with respect to a second science domain, are done for year 1985 and are summarized in Figure 8. We observe that the quivers in the top left part of the map are the longest ones, which is due to the ordering of the science domains.

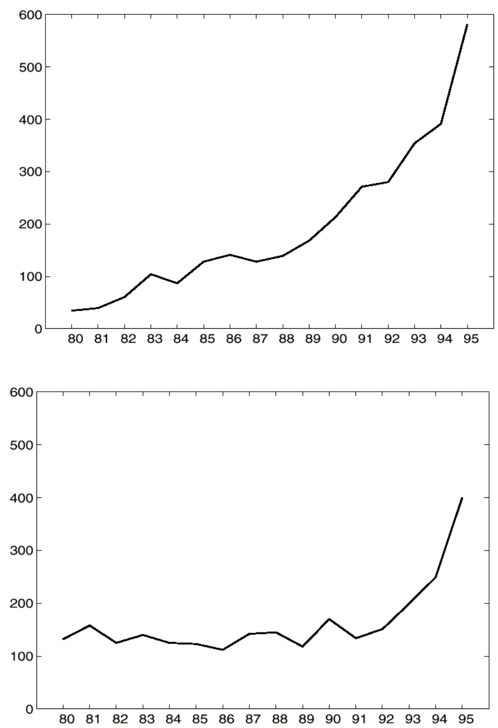

Figure 4. Temporal evolution of $NPAT_{fract1}$(C12N,t,279) (above) and $NPAT_{fract1}$(C12N,t,350) (below)

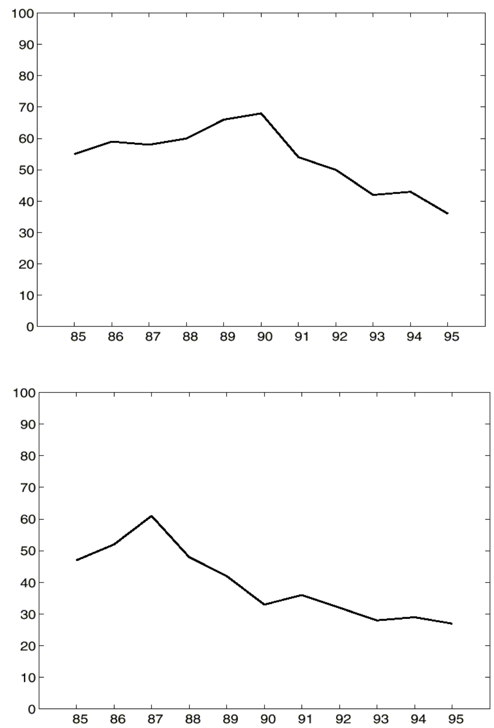

Figure 5. Temporal evolution of $NPAT_{fract2}$(C12N,*t*,279,1985) (above) and $NPAT_{fract2}$(C12N,*t*,350,1985) (below)

This is also the case for the quivers along the diagonal of the quiver map. Hence, the corresponding science domains, listed row-wise, are the ones that exert the strongest effects on the science domains listed column-wise. Furthermore, we also notice that parts of the map have no quivers, hence, there are not any causal effects exerted by the science domains listed row-wise. Finally, these results clearly show that we can indeed analyze the interactions between science and technology, without having to rely on a prior selection of the science- and technology domains done by an expert.

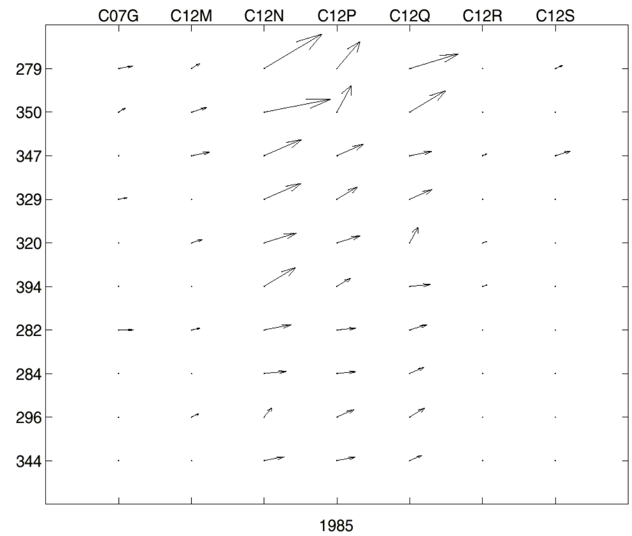

Figure 6. Quiver map for partial elasticities in the Biotechnology area for 1985, the length of the quiver corresponds to the strength of the elasticity, the angle corresponds to the delay (a horizontal quiver means a zero delay, a vertical quiver means a delay of 10 years). Four-digit codes refer to IPC classes, three-digit number refer to SCI codes

Conclusion

We have introduced a new technique for quantifying the causal effect the publication of scientific articles has on the filing of patents. The effect was examined in terms of partial and cross elasticities, and the results summarized and visualized as quiver maps. With these maps one can quickly spot the most important science domains by the extent of their causal effects.

There are at least two interesting applications of our technique. First, by performing the analysis on specific subsets of articles from given science domain or on subsets of patents from given technology domains, one can analyse the effects between these subsets. These subsets can for example correspond to articles or patents from different economic regions, and hence one can model the causal effects between economic regions. For example, the effect of scientific articles published in Europe on the filing of patents in the USA. Second, instead of using predefined science and technology domains based on SCI classifications and IPC classifications, the relevance of which can be debated, one could consider user-defined domains, e.g., the patents and scientific articles obtained by searching the respective databases for specific keywords, or one could perform a cluster analysis, e.g., on the patents by their references to scientific articles.[2]

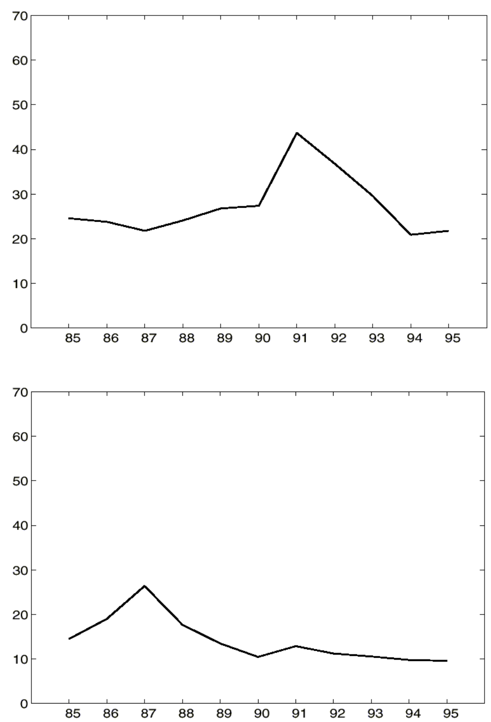

Figure 7. Temporal evolution of $NPAT_{fract3}$(C12N,*t*,279,1985,279) (above) and $NPAT_{fract3}$(C12N,*t*,350,1985,350) (below)

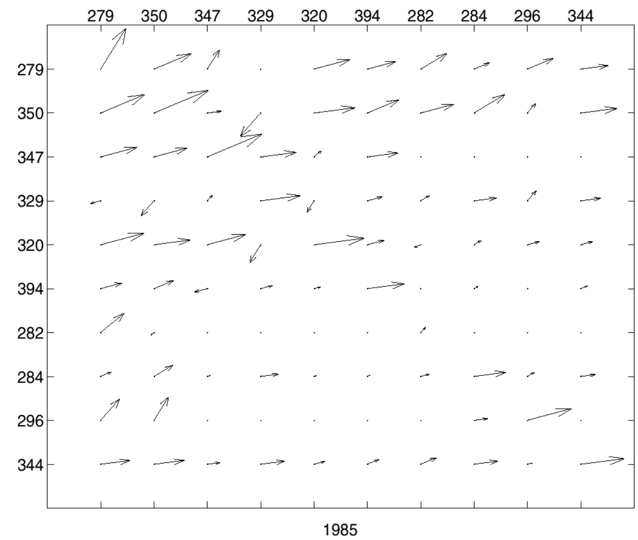

Figure 8. Quiver map for cross elasticities in technology domain C12N for 1985, the length of the quiver corresponds to the strength of the elasticity, the angle corresponds to the delay (a horisontal quiver means a zero delay, a vertical quiver means a delay of 10 years). Three-digit numbers refer to SCI codes

*

F.D. is supported by a scholarship from the Flemish Ministry for Science and Technology (VIS/98/012). M.M.V.H. is supported by research grants received from the Fund for Scientific Research (G.0185.96N), the National Lottery (Belgium) (9.0185.96), the Flemish Regional Ministry of Education (Belgium) (GOA 95/99-06; 2000/11), the Flemish Ministry for Science and Technology (VIS/98/012), and the European Commission, 5th framework programme (QLG3-CT-2000-30161 and IST-2001-32114).

References

1. CARPENTER, M. P., NARIN, F., Validation study: Patent citations as indicators of science and foreign dependence, *World Patent Information*, 5 (1983) 180–185.
2. DELEUS, F., VAN HULLE, M. M., Science and technology interactions discovered with a new topographic map-based visualization tool, *Proceedings of the Visual Data Mining Workshop* held on the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2001), San Francisco, USA, 2001, pp. 42–50.
3. GRUPP, H., SCHMOCH, U., Perception of scientification of innovation as measured by referencing between patents and papers, In: GRUPP, H. (Ed.), *Dynamics of Science-Based Innovations*, Springer Publishers, Berlin/Heidelberg, 1992, pp. 73–128.

4. MEYER, M., *Patents citing scientific literature: is the relationship causal or casual?*, *Institute for Prospective Technological Studies Reports*, 28 (1998) 11–18.
5. NARIN, F., NOMA, E., Is technology becoming science?, *Scientometrics*, 7 (3-6) (1985) 369–381.
6. NARIN, F., HAMILTON, K. S., OLIVASTRO, D., The increasing linkage between U.S. technology and public science, *Research Policy*, 26 (3) (1997) 317–330.
7. SCHMOCH, U., Indicators and relations between science and technology, *Scientometrics*, 38 (1) (1997) 103–116.
8. VERBEEK, A., DEBACKERE, K., LUWEL, M., VAN LOOY, B., ANDRIES, P., VAN HULLE, M., DELEUS, F., Linking science to technology using bibliographic references in patents to build linkage schemes, *Proceedings of the 8th International Conference on Scientometrics & Informetrics*, Sydney, Australia, 2001, pp. 717–732,