
# NOTA DI LAVORO

74.2009

## Mapping the Evolution of "Clusters": A Meta-analysis

By **Mario A. Maggioni** and **T. Erika Uberti,** DISEIS and Faculty of Political Science, Catholic University, Milano

**Francesca Gambarotto,** University of Padova

GLOBAL CHALLENGES Series
Editor: Gianmarco I.P. Ottaviano

# Mapping the Evolution of "Clusters": A Meta-analysis

By Mario A. Maggioni and T. Erika Uberti, DISEIS and Faculty of Political Science, Catholic University, Milano
Francesca Gambarotto, University of Padova

## Summary

This paper presents a meta-analysis of the "cluster literature" contained in scientific journals from 1969 to 2007. Thanks to an original database we study the evolution of a stream of literature which focuses on a research object which is both a theoretical puzzle and an empirical widespread evidence. We identify different growth stages, from take-off to development and maturity. We test the existence of a life-cycle within the authorships and we discover the existence of a substitutability relation between different collaborative behaviours. We study the relationships between a "spatial" and an "industrial" approach within the textual corpus of cluster literature and we show the existence of a "predatory" interaction. We detect the relevance of clustering behaviours in the location of authors working on clusters and in measuring the influence of geographical distance in co-authorship. We measure the extent of a convergence process of the vocabulary of scientists working on clusters.

**Keywords:** Cluster, Life-Cycle, Cluster Literature, Textual Analysis, Agglomeration, Co-Authorship

**JEL Classification:** O18 , R12 , Z13, B41

*Address for correspondence:*

Mario A. Maggioni
Catholic University
Largo Gemelli 1
20123 Milano
Italy
Phone: 0272343951
Fax: 0272342475
E-mail: mario.maggioni@unicatt.it

# Mapping the evolution of "clusters": a meta-analysis

Mario A. Maggioni^, Francesca Gambarotto*, T. Erika Uberti^

*^DISEIS and Faculty of Political Science, Catholic University, Milano*
** University of Padova*

**SUMMARY**

This paper presents a meta-analysis of the "cluster literature" contained in scientific journals from 1969 to 2007. Thanks to an original database we study the evolution of a stream of literature which focuses on a research object which is both a theoretical puzzle and an empirical widespread evidence.

We identify different growth stages, from take-off to development and maturity. We test the existence of a life-cycle within the authorships and we discover the existence of a substitutability relation between different collaborative behaviours.

We study the relationships between a "spatial" and an "industrial" approach within the textual corpus of cluster literature and we show the existence of a "predatory" interaction.

We detect the relevance of clustering behaviours in the location of authors working on clusters and at measuring the influence of geographical distance in co-authorship.

We measure the extent of a convergence process of the vocabulary of scientist working on clusters.

**Key words:** Cluster, life-cycle, cluster literature, textual analysis, agglomeration, co-authorship

**JEL CODES**: O18 , R12 , Z13, B41

## 1. Introduction

In recent years many theoretical and empirical papers have analysed the genesis, development, functioning and decline of "clusters"[1]. Different academic disciplines, streams of literature and schools of thought have been involved in the analysis of *spatial* and *industrial* agglomeration of firms. At the same time the "cluster model" has been seized on by public authorities and policy makers as a tool for promoting competitiveness, innovation and growth at local, regional and, sometimes national level.

However, to quote Martin and Sunley (2003), "the mere popularity of a construct is by no means a guarantee of its profundity. Seductive though the cluster concept is, there is much about it that is problematic, and the rush to employ 'cluster ideas' has run ahead of many fundamental conceptual, theoretical and empirical questions" (p. 7).

This is not to deny that the cluster approach, as stressed by Malmberg and Power (2006), "has undoubtedly persuasive and has contributed to substantial progress in the analysis of the classical issues dealt by economic geographers. A the same time it is an elusive, and at time confusing, concept open to multiple interpretation and understanding (…) (which) has equally caused 'recurring headaches' for many of us active in the field of cluster research and cluster-based industrial, regional or innovation policy formulation" (p. 50).

This paper looks at clusters and, in particular, at cluster life-cycle, from a different perspective. Many papers have been written on the development pattern of industrial clusters and this special issue will add original theoretical insights and new empirical evidence on this issue. However since clusters are a research topic which has been investigated for, at least, 40 years, we think that it is now the right time to look at the evolution (and check whether there is a life-cycle) of clusters not as economic phenomena, but as research objects. A seminal contribution on this issue is Maskell and Kebir (2006).

For this reason, we present a meta-analysis of the "cluster literature"[2] contained in scientific journals from 1969 to 2007. Thanks to an original database, built for this purpose, we are able to study the evolution of clusters through a plurality of techniques (from characteristics textual analysis to population ecology, from social network and geographical analysis to correspondence analysis), by studying the development of a stream of literature which is at the cross-roads of different scientific disciplines (economics, geography, environmental studies, regional science, urban studies, business and management, transport studies) and focuses on a research object which is both a theoretical puzzle and an empirical widespread evidence.

The rest of the paper is organised as follows: section 2 describes the selection and construction of the bibliographic database; section 3 describes the life-cycle of clusters with respect to the comparison of "new" vs. "old" authors' behaviours; studies the evolution of the vocabulary used in this literature and compares the diffusion of "industrial" vs. "spatial" dimensions of the concept of cluster; section 4 analyses the existence of "clustering" phenomena and dynamics in the cluster literature with reference to two distinct geographies: "physical geography" (i.e. cities where scientists have their academic affiliations) and "textual geography" (i.e. based on the specific vocabulary used by scientists in different countries); section 5 concludes the paper.

## 2. The database

To analyse the evolution of the concept and use of "clusters" in the economic literature we selected articles published in international scientific journals collected in two databases, ISI-Thomson "Web of Science" and "EconLit", following a three-steps procedure.

[1] See, among others, Bresnahan and Gambardella (2004); Breschi and Malerba (2005); Asheim et al. (2006); Braunerhjelm and Feldman M. (2006); Karlsson (2008).

[2] For our definition of "cluster literature", see section 2.

Firstly, in December 2007, we downloaded all titles and abstracts (where available) of all articles containing the "word" <cluster*>.

In particular, as far as the ISI-Thomson "Web of Science" (henceforth ISI) database is concerned, we limited the analysis to a subset (i.e. economics, planning and development, geography, management, environmental studies, business and urban studies) of all subject categories contained in the social science citation index. In addition we excluded those articles referred to <cluster analysis> as a statistical procedure.

Similarly, we conducted the same textual search in the EconLit database. To avoid any unbalance between Econlit and ISI - and to include information needed to conduct this analysis (i.e. texts of abstracts and affiliations to geo-code authors) - we excluded volumes, collective volume articles, working papers and Ph.D thesis which are contained in the EconLit database, and we limited our search to journal articles.

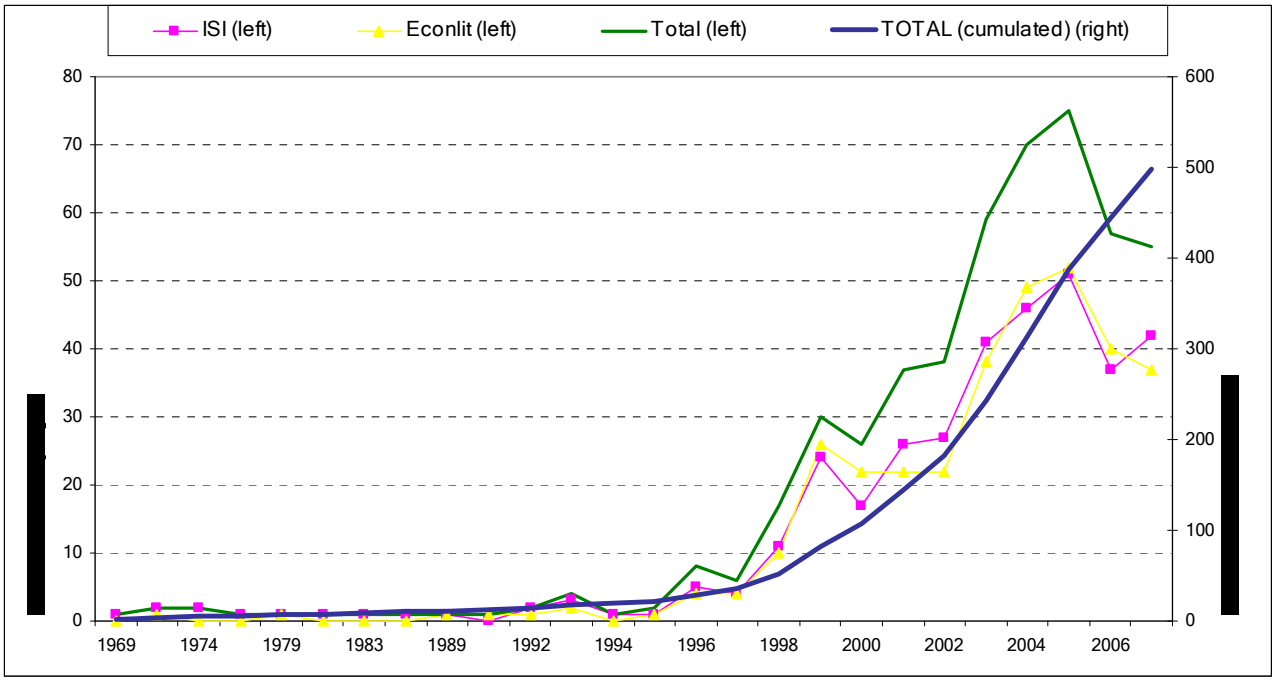
Secondly, we excluded from our database those articles (such as comments, introductions of special issues, editorials and book reviews) that do not contain original research output and which lack information needed for both textual and geographical location analyses.

Thirdly, we merged both databases in order to avoid duplications so to obtain a final set of 499 scientific articles ranging from 1969, when the first article[3] on clusters was published, until 2007.

Due to the different purposes of ISI and EconLit databases, we were able to include in our analysis papers published in academic journals with a national and international coverage, hence to capture possible national differences. The degree of overlap between the two databases varies between 13% (in 1996) and 67% (in 1999)[4].

As it can be seen in figure 1 (left axis), the cluster topic followed similar patterns (as measured by the number of articles published per year) in both databases with a slow start (with 13 blank years before 1992), a booming period and a slow down in the last 2 years (which may be due also to the updating delays of the bibliographic databases). On the right axis, one can read the cumulated number of articles which follows a typical "S-Shaped" (possibly logistic) form.

*Figure 1: Number of Articles on clusters*




[3] The first article in the database is "Some properties of a cluster point process", by M.F. Dacey, and published in the *Canadian Geographer*.

[4] While in 1979 and 1989 both databases include the same articles.

Source: our calculations on ISI and EconLit databases.

Since we wanted to investigate the existence of clustering behaviour displayed by people writing on clusters and, more generally, we wanted to analyse the spatial distribution of this literature, we extracted information from each record included in our final database regarding each author's affiliation[5]. In this way we were able to associate each paper to one (or more) city and country[6], in order to conduct both a geographical analysis of the "cluster" scientific community and a textual analysis focused on national differences and similarities of the vocabulary[7].

In our database papers were written by authors working in 282 cities located in 45 nations; however, in the "textual geography" analysis, described in section 4.2, we focused exclusively on the most relevant countries (Canada, France, Germany, Italy, Netherlands, Spain, Sweden, United Kingdom, United States) which accounted for over 85% of the papers.

As far as the time span of the database is concerned, for different analyses we partitioned the covered period (1969-2007) into a number of periods according to three different procedures: "statistical", "historical" and a "textual analysis" procedure.

The "statistical" procedure, based on quartile distribution, was adopted to obtain a roughly equal number of articles per each period. Due to the unequal distribution of papers (see figure 1 and table 1), the first period spans from 1969 to 2000 (and includes 108 articles), while the remaining three periods divide the last 7 years as follows: the second period (including 134 articles) ranges between 2001 and 2003; the third period (containing 145 articles) ranges between 2004 and 2005; and the fourth period includes 112 articles published between 2006 and 2007.

The "historical" procedure was adopted to discuss some specificities of different phases of the cluster literature. For this reason we identified three periods: the "take-off" period, spanning from 1969 to 1989, which includes 11 articles; the "development" period, spanning from 1990 to 1999, which includes 71 articles; and the "maturity" period, spanning from 2000 to 2007, which includes 417 papers.

Finally, to describe the evolution of the vocabulary used in the titles and abstracts of the articles included in our dataset, we used a "textual analysis" procedure limiting the analysis to the last 10 years (1998-2007), since over 90% of the total scientific production (and of the vocabulary[8]) has been published in these years. As it will be described in greater detail in section 3.2, the selection of a given initial year is crucial to calculate the IT index and to associate the trends, hence any choice involving previous years as the reference period would have distorted the analysis upward.

These different procedures are functional to the different analyses conducted in this paper. According to the statistical procedure, we are able to define periods in order to distribute papers homogenously over time avoiding an extreme unbalancing of the distribution; according to the historical procedure, we exogenously identify three periods, defined as by relevant phases of the scientific literature on clusters; finally the selection of the last 10 years period, according to the "textual analysis" procedure, was required in order to obtained unbiased textual analysis indexes.

## 3. The life-cycle of clusters

Since the main focus of the paper is about the existence of life-cycle of clusters in the economic literature, it is interesting to move forward from the mere description of the publishing trends of

[5] Since our interest was mainly "geographic" in its scope, we limited the analysis to cities and we avoided any investigation on specific universities, centres of research and/or departments.

[6] We extracted city information from each affiliation as defined by authors. In case of multiple affiliations, since in both databases there is no possibility to distinguish between primary and secondary affiliations, we maintained both information.

[7] When authors, based in different countries, write jointly a paper, for the purpose of this textual analysis we attributed t the paper to both countries.

[8] If we plot the whole vocabulary, we obtain a similar distribution resulting in figure 1.

papers on clusters, as plotted in figure 1, and to evaluate the development process of this stream of literature from a twofold perspective.

The first refers to the "fitness" of the cluster concept, where fitness is measured by the attraction of new scientists to this field of analysis (section 3.1); the second refers to the evolution of the vocabulary contained in this literature and how it evolved over time (section 3.2).

A stream of literature is in fact vital when it is able to attract new scientists, while keeping the "established base" stable and when its vocabulary evolves in order to take into account both new theoretical approaches and new empirical evidences.

### 3.1 The life-cycle of authors: "New" vs. "Old"

When we analysed the temporal evolution of papers and authors writing on clusters, we noted some interesting features. As described in the previous section, the number of papers is concentrated in the last ten years; however, during the whole period, it is possible to build some behavioural indexes that may help to identify the existence of different waves of the diffusion of this topic in the scientific literature.

As previously highlighted, the temporal distribution of papers is very skewed. Indeed from 1969 till 1995, less than 2 papers were published, on average, each year: in particular during this period, lasting nearly 30 years, only 4,2% of total papers was published and only 3,5% of total number of authors appeared (table 1). From 1995's to 1999, the number of articles increased, reaching 12,2% of total number of papers, while the vast majority of papers (83,6%) has been published in the last 7 years.

*Table 1: Descriptive statistics on articles and authors*

| Year | Statistical period | Historical period | Articles | Authors | Average Authorship | "New" Authors | "Old" Authors | % of "Old" Authors |
|---|---|---|---|---|---|---|---|---|
| 1969 | 1 | 1 | 1 | 1 | 1,00 | 1 | 0 | - |
| 1973 | 1 | 1 | 2 | 2 | 1,00 | 1 | 1 | 50,0 |
| 1974 | 1 | 1 | 2 | 4 | 2,00 | 4 | 0 | - |
| 1978 | 1 | 1 | 1 | 1 | 1,00 | 1 | 0 | - |
| 1979 | 1 | 1 | 1 | 2 | 2,00 | 2 | 0 | - |
| 1982 | 1 | 1 | 1 | 1 | 1,00 | 1 | 0 | - |
| 1983 | 1 | 1 | 1 | 1 | 1,00 | 1 | 0 | - |
| 1987 | 1 | 1 | 1 | 1 | 1,00 | 1 | 0 | - |
| 1989 | 1 | 1 | 1 | 1 | 1,00 | 1 | 0 | - |
| 1990 | 1 | 2 | 1 | 1 | 1,00 | 1 | 0 | - |
| 1992 | 1 | 2 | 2 | 3 | 1,50 | 3 | 0 | - |
| 1993 | 1 | 2 | 4 | 7 | 1,75 | 7 | 0 | - |
| 1994 | 1 | 2 | 1 | 1 | 1,00 | 1 | 0 | - |
| 1995 | 1 | 2 | 2 | 3 | 1,50 | 3 | 0 | - |
| 1996 | 1 | 2 | 8 | 11 | 1,38 | 10 | 1 | 9,1 |
| 1997 | 1 | 2 | 6 | 8 | 1,33 | 7 | 1 | 12,5 |
| 1998 | 1 | 2 | 17 | 27 | 1,59 | 26 | 1 | 3,7 |
| 1999 | 1 | 2 | 30 | 42 | 1,40 | 37 | 5 | 11,9 |
| 2000 | 1 | 3 | 26 | 40 | 1,54 | 33 | 7 | 17,5 |
| 2001 | 2 | 3 | 37 | 57 | 1,54 | 50 | 7 | 12,3 |
| 2002 | 2 | 3 | 38 | 62 | 1,63 | 53 | 9 | 14,5 |
| 2003 | 2 | 3 | 59 | 102 | 1,73 | 87 | 15 | 14,7 |
| 2004 | 3 | 3 | 70 | 127 | 1,81 | 97 | 30 | 23,6 |
| 2005 | 3 | 3 | 75 | 132 | 1,76 | 106 | 26 | 19,7 |
| 2006 | 4 | 3 | 57 | 105 | 1,84 | 77 | 28 | 26,7 |
| 2007 | 4 | 3 | 55 | 97 | 1,76 | 72 | 25 | 25,8 |
| **TOTAL** | | | **499** | **839** | **1,43** | **683** | **156** | |

Source: our calculations on ISI and EconLit databases.

Furthermore, the authors' cooperative behaviour (i.e. the average number of people co-authoring a paper) increases steadily through the period. From 1969 to 1995 (with the exception of years 1974, 1979, 1992, 1993 and 1995) all papers were written by a single author (see table 1). From 1996 co-authorship begun to diffuse and continuously increased over time, although the average number of authors per paper remains always less than two (i.e. 1,43 person per paper)[9]. Limiting the analysis to different periods, identified according to the "statistical" and "historical" procedures, this upward trend is confirmed. In particular, the average number of authors per paper in the statistical periods increases by 37% from the first period (when the average authorship is equal to 1,32) to nearly 2 persons per paper in the last period (when the average authorship is equal to 1,80)[10].

It is also interesting to analyse the entry behaviour of scholars, by comparing the behaviour of "New" and "Old" scientists.

We consider as "New", any author that appears for the first time in the database, irrespectively to the kind of authoring (single or co-authorship); successively any scholar that, according to the records contained in this database, has already written at least once previously, is considered as an "Old" author.

As far as the presence of "New" and "Old" authors is concerned, table 1 (last column) clearly shows a structural break year (1996) that splits the database into two parts: indeed from 1996 on[11] an increasing percentage of "Old" authors stabilises its presence in cluster literature continuing to publish papers on this topic.

If we focus the attention on "New" authors we can interpret their presence in the database as a proxy variable for the attractiveness of the cluster topic. In particular we may measure the attractiveness of clusters topic as the share of "New" authors writing on clusters – without any collaboration with older scholars – respect to the total number of "New" authors: this share is always very high and ranges between 80% and 90%, confirming that there is room for newcomers to write on this topic in the journals included in this database. It is however interesting to note that this share is slightly decreasing in last years, since more and more "New" authors collaborate with "Old" scholars, whose presence in the clusters literature is already consolidated. Respect to this issue, it seems therefore that "Old" authors play a mentoring "pull-effect" in the scientific publishing activity by introducing "New" authors in this stream of literature.

If we focus the attention on the authorship and coauthorship behaviour of "Old" scholars, we may identify four different strategies, ranging from very autarkic to very relational ones. The *autarkic* behaviour identifies scholars that maintain an attitude to write on their own, taking advantage from their previous works on this topic. A different behaviour, labelled *invisible colleges*[12], identifies scholars that exploit their belonging to an established scientific community, and when writing on the same issue (i.e. clusters) they rather prefer to collaborate with other "Old" scholars. Finally we identified two other collaborating behaviours: the *mentoring-effect* of older scholars that introduce new people to the issue and possibly exploit their new competencies, and the *mixed- effect*, when "Old" scholars write jointly with "New" and "Old" authors, combining new and established competencies in order to produce an original piece of research.

By plotting "Old" scholars' behaviour from 1996 onward we may have a hint on how the sociological structure of the "cluster literature" modifies over time (figure 2).

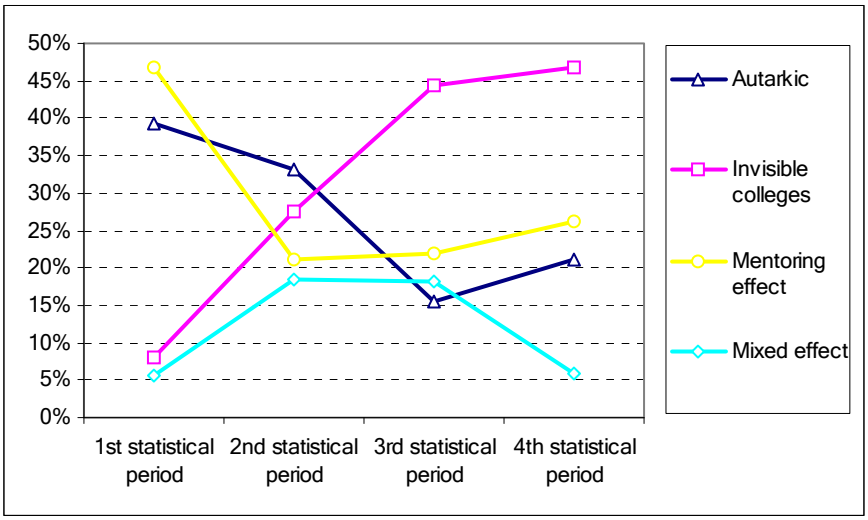
[9] Remarkably lower than in other scientific disciplines (studied by: Kretschmer, 1994; Barabasi et al. 2001; Newman, 2004; Glanzel and Schubert, 2004).

[10] Similarly analysing the average number of authors according to the historical procedure, there exists an analogous upward trend: the average number of authors per paper increases by 42% from the first to the last period.

[11] In 1973, only one article was written by an "Old" author, M.F. Dacey. After 1996, the presence of "Old" authors slow down in 1998, but after this year there is a continue increase of this quota.

[12] The label name is a tribute to the seminal contribution of Diana Crane (1972).

*Figure 2: Average "Old" authors' behaviours according to the statistical procedure*




Source: our calculations on ISI and EconLit databases.

The *autarkic behaviour* is constantly decreasing over time, while the *invisible colleges* strategy appears to be steadily increasing through the periods. Both the m*entoring-effect* and the *mixed-effect* show non-monotonic time patterns. However while the *mentoring-effect* shows a relevant decrease from the first to the second period and a slow recovery in the last period, the *mixed-effect* displays an inverted-U pattern with a higher level recorded in the two central periods.

By looking at figure 2 one may suggest the existence of a substitutability relation between two couples: autarkic vs. invisible colleges and mentoring vs. mixed effects.

In order to expand the insight on the sociological structure of the "cluster literature", as it can be measured by the co-authorship behaviours, we decided to add the spatial aspect into the picture. These results are thoroughly discussed in section 4.1.1.

### 3.2 The life-cycle of words and concepts

The main aim of the paper is to detect the evolution pattern of the cluster literature. In order to do so we used different statistical tools applied to textual analysis, as analysis of characteristic textual terms and correspondence analysis, to identify the main research topics of this literature; to stress the role played by different countries; and to test whether a research topic convergence has emerged.

The analysis of characteristic textual terms (presented in this section) allows the study of research topics and their enrichment; correspondence analysis (described in section 4.2), which is based on textual longitudinal data, is a graphical description of textual elements by time periods and it is useful to detect the relationships existing between specific research topics and specific geographical areas (i.e. countries in this paper).

In order to perform these two statistical exercises, we processed all records contained in the database and we performed a content analysis to produce a set of categorical variables (research topics) and to detect the characteristic textual units emerging during time. Content analysis is a research tool that uses words as variables and that explores linguistic and textual properties of texts[13].

[13] Content analysis has been recently used as analytical tool for economic and business analysis and, especially, within the field of organizational studies (Duriau et al., 2007). For an application to labour economics and regional science see Gambarotto and Walter (2005), Gambarotto and Bramanti (2009).

Before building the matrix of graphical forms (*word-types*) we analyzed the quality of the textual *corpus*[14]. The *corpus* dimension, *N*, is given by the number of *word tokens* (statistical textual units) while the vocabulary includes *V* distinct words, or *word-types*. Each word-type is associated with its number of *word tokens*, i.e. with its frequency. Observing our textual *corpus*, it results that it is a medium-large *corpus*[15] including 76.262 word tokens (*N*) while the extracted vocabulary is made up of 6.468 word-types (*V*). Our analysis will be based on the vocabulary (*V*) to identify the main features of the cluster literature and to interpret its evolution (Lebart et al., 1998).

We evaluated the lexicographic richness of our textual corpus. Content analysis uses two fundamental lexicographic indexes: the first one is the ratio between the size of the vocabulary and the corpus extension (*V*/*N*) and measures the richness of vocabulary. Content analysis can be applied to a corpus only if this ratio is lower than 20%. The second index, the hapax percentage, measures the words appearing only once in the text (*hapax*). When the number of hapax ($V_1$) is too large, (i.e. $V_1/V > 50\%$) it means that the language used in the corpus is over-refined (on average, each word is used only twice) and the statistical analysis is therefore not possible. Our textual corpus satisfies both conditions: the lexical richness of vocabulary (*V*/*N*) is 8,48% and the "share" of hapax in the corpus ($V_1/V$) is 43,26%.

By using the abovementioned software application then we applied the following procedures to extract textual information form the corpus:

1. Parsing: this is a routine phase in which data files are cleaned of errors and misprints, English and American spellings are harmonised and the use of uppercase, abbreviations, and delimiters is taken into account.
2. Analysis of vocabulary: we extracted the vocabulary from the normalized corpus. We performed a statistical analysis of frequency classes of word-types and, finally, we calculated the vocabulary coverage, which evaluates the weight of a word-type with respect to vocabulary.
3. Segmentation: we divided the text into distinct units (with different length) and we selected those units embodying a linguistic sense. This is a very important step because we identify the core of the distinctive language used in papers. We identify the principal polyforms and polyrematics using the Index of Significance (IS index[16]) and transforming single distinct words composing a segment into one single statistical unit (complex expressions).
4. Lexical analysis: the original corpus is restored adding those segments of meaningful words for our investigation. At this stage we obtain the complete vocabulary with text segments selected as significant for our analysis. We have extracted 11.931 text segments composed by 2, 3 or 4 word-types and selected 351 meaningful segments: among others, those containing <cluster*>, <innovation>, <region*>, <technol*>, <local*>, <sector*>, <geograph*>, <industr*> . The lemmatized vocabulary includes 6.807 word-types.
5. Characteristic textual analysis: we looked for the peculiarity of the language. In our case the characteristic language of titles and abstracts of papers published each year since 1969.
6. Correspondence analysis, presented in section 4.2, is applied to vocabulary in order to synthesize graphically the association between selected categorical variables (in our case nationality of researchers) and items (word-types). The analysis was performed over the corpus partitioned in four statistical periods to identify the evolution of cluster concept within this stream of literature.

The characteristic textual analysis (or distinctive language) allows us to identify the thematic words that have contributed to define research topics over time. "The characteristic textual element is the

[14] All textual analyses have been performed with a specific software application, Taltac2 (www.taltac.it).

[15] *Corpus* dimensions are classified as following: small dimension, when the length of the *corpus* is less than 15.000 word tokens; medium dimension when the length varies between 15.000-50.000 word tokens, medium-large if it varies between 50.000-100.000 tokens and a large *corpus* when its size exceeds 100.000 word tokens (Tuzzi, 2003).

[16] See Lebart et al. (1998) for the statistical properties of this index.

minimum set of words that maximally represents the vocabulary" (Bolasco, 1999). This means that it is possible to observe the vitality of the cluster concept exploring the evolution of its distinctive language over time (Lebart et al., 1998; Bolasco, 1999). Characteristic elements are calculated for each subtext (composed by the papers published each year) so that for each word or segment we are able to assign an atypical frequency (i.e. overused/underused word-types). To identify these words, we used a hypergeometric model, asymptotically approximated to a Gaussian distribution, by which a probability of frequency of each word-type in each historical period is calculated. When the effective word-type frequency approximates the probabilistic value, the word-type is defined "banal" (i.e. it belongs to the basic vocabulary of the corpus); when it is higher, it is defined "positively specific" for that period; while a negative specificity (rare word-types) emerges when the effective frequency is lower than the probabilistic value.

The probability level used to select characteristic words, or *p-value*, is set equal to 0.02 while we did not introduce an exogenous frequency threshold for word-types since we are interested in detecting new conceptual entries in the vocabulary. Therefore we calculated characteristic textual units for word-tokens appearing only once.

Tables 2, 3, 4 present some results of the positive characteristic textual words by historical periods.

We grouped positive characteristic textual elements according to the historical procedure since, in this way, we are more able to detect the life-cycle of cluster literature. In fact the first period represents the take-off period (1969-1989) with a slow starting scientific production (only 11 papers); the second period, the development period (1990-1999), includes years of enlarging research and the early diffusion of the concept (with 71 papers), and finally the maturity period (2000-2007), recording 417 papers, includes years of concept mining, when the adoption of cluster concept increased and assumed a relevant economic role in the scientific literature.

Moreover, we consider the positive characteristic textual words as key-words of three distinct but complementary and intertwined approaches: the *methodological* approach, where research activity is focused on the theoretical description and the analytical tools designed to interpret the clustering phenomenon and its dynamics, the *regional* approach, in which clustering is analysed mainly as a spatial process and, in addition, the focus of the analysis is centred around the economic advantages of firms' location; and finally the *industrial* approach, in which the agglomeration of firms is mainly due to innovation and organizational processes within a given industry, or set of industries and the clustering process directly derives from marketing and organisational strategies of firms[17].

As synthesised in table 2, during the "take-off" period, the analysis of clusters was grounded on "classical" location theory – with reference to the models of Christaller (1933) and Lösh (1940) – and addressed <principle> or <properties> related to regional topics. Urban and industrial agglomerations were considered spatial phenomena looking for spatial economic description and, at that time, the available theoretical tools were those of classical location analysis, i.e. central place and gravity theories. The industrial approach (exemplified by words such as <activities> and <complementary> appearing in 1974) was subordinated to the regional approach since during this period the main interest was to explain the contribution of spatial agglomeration to economic development.

Only in the '80s cluster literature moved from classical location theory to a deeper investigation of the role played by the industrial structure in the development process of local economies, perhaps driven by a renewed interest in the works of Perroux (1955) and Hirschman (1958). Even if <industrial cluster> appears as key-word in 1978, only during the '80s it started to be considered as a theoretical issue on its own in which industry and space were considered equally important in explaining the development patterns of local economic systems (table 2).

[17] For interpretation of the positive characteristic textual words, we studied the local textual contexts of word-types or polyforms inside the body of text (i.e. concordance analysis). For further details see Lebart et al. (1998).

*Table 2: Positive characteristic textual elements for the take-off period (1969-1989)*

| 1969 | 1973 | 1974 | 1978 | 1979 | 1982 | 1983 | 1987 | 1989 |
|---|---|---|---|---|---|---|---|---|
| *p-value <0,001* | *p-value <0,001* | *p-value <0,001* | *p-value <0,01* | *p-value <0,001* | *p-value <0,001* | *p-value <0,001* | *p-value <0,001* | *p-value <0,01* |
| properties | urbanisation | income | central | identification | underdevelopment | complexes | industrial clustering | source |
| *p-value <0,01* | *p-value <0,01* | places | *p-value <0,02* | methods | *p-value <0,02* | manufacturing | urban | dynamic |
| process | temporal | activities | region | complexes | approach | sized | pattern | *p-value <0,02* |
| | dispersion | christaller | industrial clusters | sectoral | | *p-value <0,01* | *p-value <0,01* | innovation |
| | sectoral | system | | *p-value <0,01* | | medium | disintegration | |
| | geographical | lösch | | disaggregation | | systems | centrifugal | |
| | *p-value <0,02* | principle | | aggregation | | threshold | analysis | |
| | interaction | *p-value <0,01* | | studies | | local industry | industrial production | |
| | focus | place | | spatial | | urban | distributions | |
| | central | entry | | *p-value <0,02* | | *p-value <0,02* | manufacturing | |
| | process | *p-value <0,02* | | results | | diversity | dispersal | |
| | | complementarity | | similarity | | | *p-value <0,02* | |
| | | population | | composition | | | relocation | |
| | | | | locational | | | technology | |
| | | | | | | | productive | |
| | | | | | | | functional | |
| | | | | | | | production | |
| | | | | | | | tests | |
| | | | | | | | manufacturers | |

Source: our calculations on ISI and EconLit databases.

During the “development” period (table 3), we observe the occurrence of a major methodological breakpoint: Paul Krugman enters this literature in 1991 (Krugman, 1991a, b, c) and explains the agglomeration process using a theoretical framework – based on monopolistic competition (*à la* Dixit-Stiglitz (1977) and iceberg costs, very different from the classical location approach. A new interest is raised in the economic literature on the issue of firms location and agglomeration and, as a consequence, the positive characteristic textual elements became: <increasing>, <returns>, <monopoly>, <historical>, <dominate> .

He successfully adapted the economic tools designed to describe international movements of production factors and products to regional studies. Krugman’s success into this literature is partly due to the economic formalism he introduces as a “new” method of investigation and partly due to the changing economic relations in the real world (in other words to the globalization process). During the ‘90s the reduction of communication and transportation costs, the enlargement of markets and the reduction of administrative entry barriers for foreign goods changed the world economic organization transforming nations into economic regions. The combination of decreasing transportation costs, increasing returns to scale become the key economic factors to explain the spatial allocative equilibrium and economic geography configurations (clustering vs. dispersion).

However, at the same time, a more <industrial> perspective to regional growth began to develop. The <industrial clustering> perspective of Michael Porter (1990) was affecting regional studies introducing the new concept of competitiveness. <Sectoral linkages> and <sectoral clustering> in 1994, <advantages> and <institutions> in 1995, <competitors> and <innovative> in 1996, are typical concepts within the Porter’ competitive advantage “diamond”. During the ‘90s the methodology of cluster’s investigation changed radically because of these two new approaches. Their different perspectives produced a large methodological debate resulting in a theoretical dispute during the “maturity” period.

During the “development” period, other theoretical approaches began to emerge in this literature: the French school of milieu and the Italian school of <industrial districts> (1999), both of them with deep roots in the regional science tradition[18]. They investigate the <spatial concentration> (1994) of firms with an institutionalist theoretical framework, i.e. assuming that economic relationships are

[18] Both the “*milieu innovateur*” and the “industrial districts” approaches have been developed earlier than these dates (Becattini, 1979 and 1987 are the seminal contributes of the Italian school of industrial districts; Aydalot (1986) and Aydalot and Keeble (1988) are the early works in the *milieu innovateur* approach). However it took some years before these two heterodox “schools” were accepted in journals indexed in the two bibliographic databases used in this analysis.

embedded into social networks. They differently contribute to the economic description of cluster of firms because the French school, following the Perroux-Aydalot's tradition, stressed the role of innovation as crucial for local development (<innovative> and<coordination> in 1996; <collective> in 1999). The Italian school, following the re-interpretation by Becattini of the original Marshall's contribution, identifies people (i.e. workers and entrepreneurs) as key-factors for clustering, as witnessed by word-types such as <specialized labour> (1995), <trust> (1999), <social network> (1999).

Again, in this period, we observe an outburst of industrial research issues within the cluster literature with a new emphasis on technology and innovation production and diffusion, as the positive characteristic textual elements show: <industrial clustering>, <technology>, <high-technology cluster>, <patenting>.

*Table 3: Positive characteristic textual elements for the development period (1990-1999)*

| 1990 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 |
|---|---|---|---|---|---|---|---|---|
| *p-value <0,001* | *p-value <0,001* | *p-value <0,001* | *p-value <0,001* | *p-value <0,01* | *p-value <0,001* | *p-value <0,001* | *p-value <0,001* | *p-value <0,001* |
| returns | viability | fuzzy | sectoral linkages | economies | life cycle | attraction | core | country |
| *p-value <0,01* | gains | suburban | industrial performance | policies | competitors | disadvantages | natural resources | krugman |
| increasing | efficiency | membership | porter | cross-industry | industry | statistical | data | export |
| monopolizing | dispersed | set | manufacturing | business strategy | innovative | dispersion | competence | clustered |
| enters | clustering | transnational cluster | association | specialized labour | stages | *p-value <0,01* | segmentation | manufacturers |
| *p-value <0,02* | *p-value <0,01* | culture | industrial clustering | dynamic | propensity | industrial development | factor | businesses |
| locational clusters | small firms | aggregate | *p-value <0,01* | institutions | activity | economies | techno-industrial clusters | competitive |
| monopoly | cooperate | taxonomy | national | tertiary | methods | resources | r&d | markets |
| regional economics | porter | metropolitan | spatial concentrations | *p-value <0,02* | structures | identification | competition | countries |
| agglomeration | collective | clustering | sectoral clustering | industry clusters | coexistence | function | productivity | statistic |
| industry location | hierarchical | *p-value <0,01* | urban | industries | adaptation | estimation | change | *p-value <0,01* |
| heterogeneous | method | explanations | interact | business | incumbent | equipment/research | size | producers |
| dominate | *p-value <0,02* | nations | *p-value <0,02* | advantages | authority | | *p-value <0,01* | interfirm |
| dominant | specificities | partitions | industrial policy | external | *p-value <0,01* | | techniques | interdependencies |
| regions | nation | footloose | | decisions | strategies | | metaphors | trust |
| historical | | space | | corporations | sectoral | | innovate | disadvantage |
| | | *p-value <0,02* | | innovation networks | properties | | inputs | business growth |
| | | communities | | | policies | | decision | industrial district |
| | | uniqueness | | | dominant | | equilibrium | social networks |
| | | unsuccessful | | | tacit knowledge | | companies | collective |
| | | guidelines | | | biotechnology | | location | diversification |
| | | leadership | | | spatially | | exploratory | production systems |
| | | | | | *p-value <0,02* | | representation | import |
| | | | | | exploratory | | microeconomics | advantages |
| | | | | | declining | | high-technology cluster | ties |
| | | | | | disaggregate | | path | *p-value <0,02* |
| | | | | | investigation | | *p-value <0,02* | supply |
| | | | | | coordination | | approaches | entrepreneurial |
| | | | | | | | patenting | quality |
| | | | | | | | innovating | substitution |
| | | | | | | | high-technology | standards |
| | | | | | | | determinants | biomedical |
| | | | | | | | clustering | industrial regions |
| | | | | | | | comparisons | liberalization |
| | | | | | | | porter | exporters |
| | | | | | | | | survival |
| | | | | | | | | organization |
| | | | | | | | | market |
| | | | | | | | | smes |

Source: our calculations on ISI and EconLit databases.

In the “maturity” period (table 4), more advanced methodological tools were developed to better describe the functioning and evolution of the cluster phenomenon, from a qualitative perspective, and to make the cluster concept operational for statistical and econometric analyses, from a quantitative perspective. The qualitative investigation was focused on the description of <localised clusters> and <local clusters> (2000), based on <small firm networks> (2000), <integrated> (2001), <spatial proximity> (2003), <path-dependency> (2005), <local relationships> (2006), <learning networks> (2007). Differently the quantitative investigation was focused on <econometric> analysis in order to detect regularities for locational firms’ behaviour and to forecast the effects of exogenous shocks or policy instruments. Quantitative analyses paid large attention to <local knowledge spillovers>, <internationalization> in 2005, knowledge production and technological organization, e.g. <knowledge cluster> (2003), <science> and <park> in 2005, <university-industry> (2007).

*Table 4: Positive characteristic textual elements for the maturity period (2000-2007)*

| 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
|---|---|---|---|---|---|---|---|
| *p-value <0,001* | *p-value <0,001* | *p-value <0,001* | *p-value <0,001* | *p-value <0,001* | *p-value <0,001* | *p-value <0,001* | *p-value <0,001* |
| subsidiaries | biotechnology | branches | brownfield | development | knowledge | wages | network |
| local cohesion | flexibility | subcontractors | incubator | immigrants | technological life-cycles | financial | performance |
| local spin-offs | competencies | clustering | relations | cities | creativity | industrial clusters | technological incubators |
| competition | transportation | metropolitan | indicators | economic | restructuring | internationalization | stability |
| *p-value <0,01* | statistics | plants | districtualization | knowledge dynamics | industrial areas | local industrial clusters | local cultures |
| co-operation | business start-ups | production | policy | entrepreneurship | local knowledge spillovers | business clusters | sector structure |
| linkages | nations | electronics | firms | theorising | autoregression | technology parks | social capital |
| small firm networks | *p-value <0,01* | performing | measures | technology spillovers | *p-value <0,01* | *p-value <0,01* | collective learning |
| localized clusters | integrated | human capital | *p-value <0,01* | *p-value <0,01* | path dependency | globalisation | competitiveness |
| spatial pattern | spatial concentration | r&d | model | geography | innovation capacity | innovation management | *p-value <0,01* |
| industry clusters | government laboratories | specialization | embodied | urban | science | hypothesis | learning networks |
| method | diffusion | *p-value <0,01* | developments | industrial dynamics | park | cultural | business strategies |
| diffusion | polycentric | knowledge transfer | interaction | icts | innovativeness | trajectories | local culture |
| community | nation | regional firms | government | buzz | knowledge base | foreign | organizational |
| transnational corporations | dynamic cluster | structure | university | cultural | lock-ins | district | labour productivity |
| local suppliers | european | locations | innovative | definitions | learning region | relationship | convergence |
| trade-off | competitor | regional cluster policy | knowledge clusters | geographic concentration of firms | forecasting | *p-value <0,02* | co-located |
| *p-value <0,02* | bioindustry | product development | agencies | spatial loyalty | bayesian | local relationships | social network |
| regional economies | homogenisation | biased | ict cluster | small and medium sized | diffusion of knowledge | patterns of learning | globalized |
| icts | regional policies | consistency | alliances | social capital | regional industrial identity | local institutional assets | technology policy |
| subcontracting | innovation networks | regional sectoral innovation systems | collaboration | *p-value <0,02* | microeconomic | microfoundations | venture capital |
| patent | *p-value <0,02* | integration/disintegration | *p-value <0,02* | networks | innovation | inequalities | capabilities |
| creative | growth | local system of innovation | linkages | venture capital | capability | payoffs | *p-value <0,02* |
| local clusters | concentrate | conceptualization | associations | stage | subcontractors | non-cluster | services |
| | intraregional | reciprocity | teams | debates | *p-value <0,02* | regression | networks of learning |
| | export | tests | outsourced | welfare | spin-out | agglomeration | local indicators of network |
| | | *p-value <0,02* | competitivity | evidences | spatial analysis | | industry agglomerations |
| | | coordination | autocorrelated | dispersion | communities | | geographical distribution |
| | | history | regional product | | social networks | | technological agglomeration |
| | | biotech | technological innovation | | | | homogeneity |
| | | method | technological capability | | | | knowledge absorption |
| | | characterised | spatial proximity | | | | benchmarking |
| | | acquisition | regional development paths | | | | university industry |
| | | organizations | advantages | | | | |

Source: our calculations on ISI and EconLit databases.

In addition the coexistence of different economic paradigms to investigate the cluster phenomenon has contributed to increase theoretical basis and to deepen the empirical analysis from both spatial and industrial perspectives. The regional perspective has focused on local specific factors as <buzz> and <cultural> in 2004, <innovation capacity>and <path-dependency> in 2005, <local culture>and <social capital> in 2007, while the industrial perspective has deepened the role of innovation trajectories to territorial performance adopting the resource-based view of firm: <competencies> (2001), <knowledge transfer> (2002), <technology spillovers> (2004), <science>, <park> (2005). It is important to notice that in the most recent years, textual segments embodying a strong economic meaning enriched the vocabulary: <innovation capacity>, <knowledge dynamics>, <innovation management>, <learning networks>. This is an important step of theoretical strengthening because the community, or a part of it, shares a larger vocabulary with word crystallization to describe/explain the cluster phenomenon.

To complement the longitudinal analysis of the cluster literature, we calculated the IT index (Bolasco and Canzonetti, 2003), an index used to identify trends and life-cycles of words in the evolution of a given textual corpus over time. To define trends, we selected word-types with a dispersion index between 0,1 and 0,8 to rule out hapax and very frequent words, i.e. the basic vocabulary. So we obtained 721 word-types on which it has been possible to apply the following expression:

$$IT = \left\{ \left[ \prod_{y=2}^{n} \left[ \left( Occ_y - Occ_M \right) \cdot \left( Occ_{y-1} - Occ_M \right) \right] \right] \Big/ \left| \prod_{y=2}^{n} \left( Occ_y - Occ_M \right) \cdot \left( Occ_{y-1} - Occ_M \right) \right| \right\} - 1 \Big/ 2 \qquad (1)$$

where, for each word-type, we consider the deviation between the number of normalized occurences for each year (or word tokens, $Occ_y$) and the equidistributed occurences (i.e. its mean value, $Occ_M$).

The IT index varies between -1 and 1 and allows the identification of "extinct" words ($IT = -1$), new words ($IT = 1$), and different typologies of trends[19]. Given that our textual corpus developed significantly during the last ten years (so that we have positive word-tokens at the starting year for around 60% of the 721 word-types) we constrained this analysis (and the calculation of the IT index) only to the period ranging from 1998 to 2007 in order to detect meaningful word-type trends in the last 10 years (see tables 5 and 6 and figure 3 for some examples of different types of trends)[20].

[19] It should be noticed that trend identification depends on the frequency gap between the limit values of the time period (Bolasco and Canzonetti, 2003).

[20] The stability of the IT index crucially depends on the corpus distribution over the years. Given the concentration of our corpus in the last years, trends identification can change by changing the limit values. If we chose 1997 as starting point, only the 24% of the selected textual units had a positive frequency at the beginning of period; while opting for 1999, the value increased (83%). A high number of word-types starting with zero frequency and finishing with a positive value are categorized as neologism.

*Table 5: Evolutionary trends of word-types (1998-2007)*

| Trend | Interval | IT value | Word-type | % |
|---|---|---|---|---|
| Obsolescent | | -1 | 27 | 3,7 |
| Very decreasing | | -1...-0,5 | 96 | 13,3 |
| Decreasing | | -0,5...0 | 49 | 6,8 |
| Bimodal | Occ10 – OccM > 0 | 0 | 172 | 23,9 |
| Unimodal | Occ10 – OccM < 0 | 0 | 133 | 18,4 |
| Growing | | 0.... 0,5 | 54 | 7,5 |
| Very growing | | 0,5 ... 1 | 9 | 1,2 |
| Neologism | | 1 | 132 | 18,3 |
| Non identifiable | | | 49 | 6,8 |
| **Total** | | | **721** | **100** |

Source: our calculations on ISI and EconLit databases.

As shown in table 5 most word-types used to build up the theoretical framework of the cluster concept, show a bimodal or unimodal trends (42%). This means that some words have attracted scholars' attention in particular years (unimodal), while others present an "up and down" trend suggesting they have irregularly contributed to the scientific debate. We can suppose that these particular trends depend on occasional participation of new scholars to this topic, or derive from the exploration of new theoretical paths.

In addition table 5 shows that both the very decreasing (13,3%) and neologism (18,3%) trends are very significant for the evolution of the cluster concept. In our opinion, some word-types are very decreasing because the theoretical focus changed during time, while others have been subsumed in other theoretical categories (e.g. <places> became <economies>). The relevance of the neologism group (18,3%) suggests that the cluster topic has enriched its lexicon during the period at study.

Table 6 displays some significant word-types for each group identified in table 5, while graphical examples of the main profiles are portrayed in Figure 3.

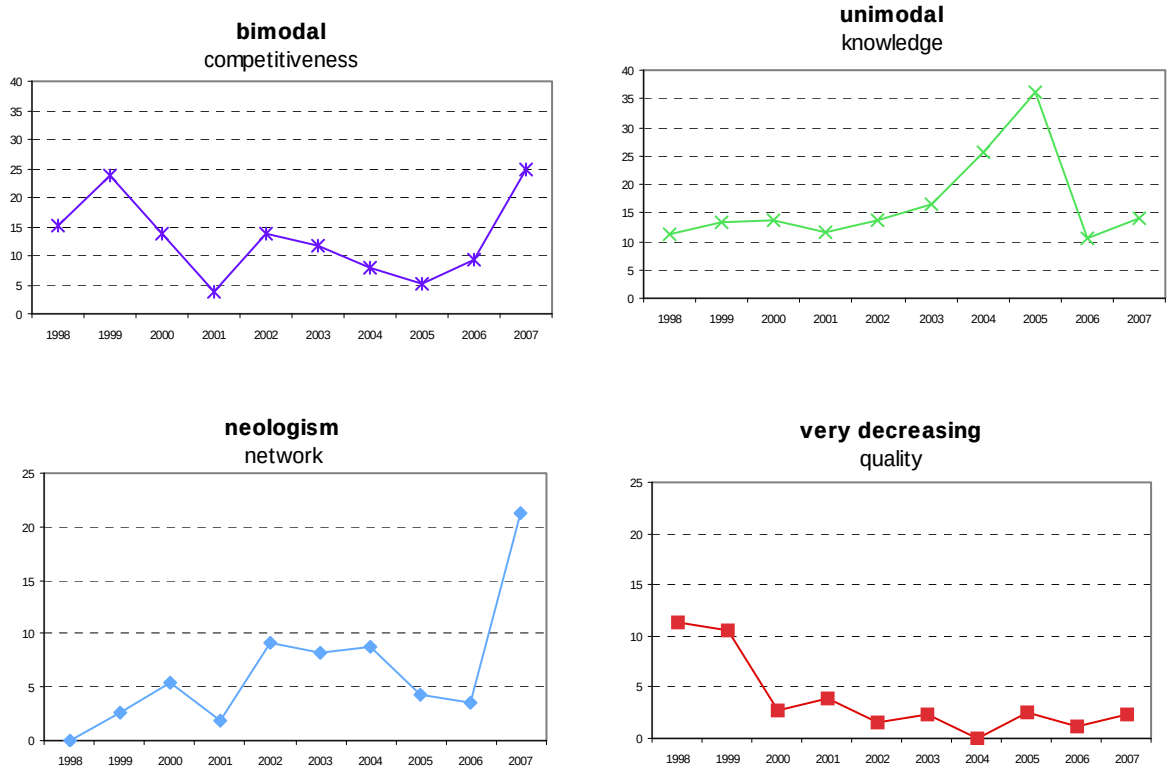
It is worth noting that the evolution of specific word-types must be considered together with other different linguistic "inflections" that the same concept acquired over time. <Evolution> displays a unimodal pattern with a peak in 2006 but <evolutionary> acts as a neologism with an increasing use in the last years; <technological> is bimodal, <technology> decreasing but <technologies> is growing and <biotechnologies> a neologism; <institutions> is bimodal but <context> is growing and <governance> is a neologism.

*Table 6: A sample of characteristic textual elements (normalised frequencies)*

| Characteristic textual elements | IT value | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Obsolescent** | | | | | | | | | | | |
| developments | **-1,00** | 4 | 0 | 0 | 2 | 2 | 8 | 4 | 3 | 0 | 0 |
| specialization | **-1,00** | 4 | 5 | 3 | 0 | 11 | 1 | 1 | 3 | 4 | 0 |
| distance | **-1,00** | 4 | 3 | 3 | 6 | 0 | 0 | 4 | 5 | 2 | 0 |
| **Very decreasing** | | | | | | | | | | | |
| quality | **-0,78** | 11 | 11 | 3 | 4 | 2 | 2 | 0 | 3 | 1 | 2 |
| infrastructure | **-0,52** | 11 | 0 | 0 | 2 | 3 | 4 | 3 | 1 | 2 | 1 |
| local firms | **-0,52** | 4 | 3 | 0 | 0 | 2 | 6 | 5 | 3 | 2 | 1 |
| **Decreasing** | | | | | | | | | | | |
| process | **-0,44** | 15 | 11 | 11 | 16 | 11 | 20 | 15 | 13 | 13 | 6 |
| located | **-0,23** | 4 | 0 | 5 | 4 | 3 | 1 | 0 | 3 | 6 | 2 |
| technology | **-0,17** | 15 | 11 | 3 | 10 | 18 | 11 | 12 | 13 | 9 | 11 |
| **Bimodal** | | | | | | | | | | | |
| technological | **0,00** | 15 | 8 | 11 | 4 | 15 | 7 | 4 | 10 | 12 | 11 |
| competitive advantage | **0,00** | 11 | 0 | 8 | 2 | 2 | 4 | 8 | 1 | 7 | 7 |
| institutions | **0,00** | 8 | 0 | 8 | 4 | 5 | 6 | 4 | 9 | 7 | 8 |
| **Unimodal** | | | | | | | | | | | |
| growth | **0,00** | 8 | 26 | 38 | 41 | 21 | 16 | 28 | 24 | 28 | 19 |
| evolution | **0,00** | 0 | 11 | 5 | 4 | 6 | 6 | 9 | 8 | 12 | 4 |
| sector | **0,00** | 8 | 0 | 8 | 16 | 21 | 7 | 14 | 16 | 13 | 5 |
| **Growing** | | | | | | | | | | | |
| global | **0,11** | 8 | 11 | 27 | 12 | 24 | 8 | 12 | 15 | 20 | 19 |
| results | **0,43** | 11 | 5 | 19 | 10 | 9 | 14 | 8 | 15 | 15 | 18 |
| management | **0,48** | 4 | 0 | 3 | 2 | 6 | 1 | 5 | 5 | 2 | 8 |
| **Very growing** | | | | | | | | | | | |
| performance | **0,52** | 8 | 5 | 14 | 18 | 6 | 6 | 5 | 13 | 18 | 29 |
| linkages | **0,59** | 4 | 18 | 30 | 6 | 5 | 20 | 16 | 7 | 0 | 12 |
| **Neologism** | | | | | | | | | | | |
| biotechnology | **1,00** | 0 | 0 | 0 | 51 | 12 | 14 | 12 | 5 | 5 | 13 |
| capabilities | **1,00** | 0 | 3 | 8 | 6 | 6 | 2 | 5 | 9 | 4 | 13 |
| processes | **1,00** | 0 | 3 | 11 | 8 | 6 | 7 | 10 | 8 | 4 | 15 |

Source: our calculations on ISI and EconLit databases.

*Figure 3: Examples of trends typologies*




Source: our calculations on ISI and EconLit databases.

### 3.3 "Industry" vs. "Space" in the evolution of clusters literature

The most diffused and quoted definitions of clusters contain references to the interaction between an industrial and a spatial element. Clusters are: "a geographic concentration of interconnected companies and institutions in a particular field" (Porter, 1998); "a large group of firms in related industries at a particular location" (Swann and Prevezer, 1998) or "a spatial and sectoral concentrations of firms" (Bresnahan et al., 2001)[21].

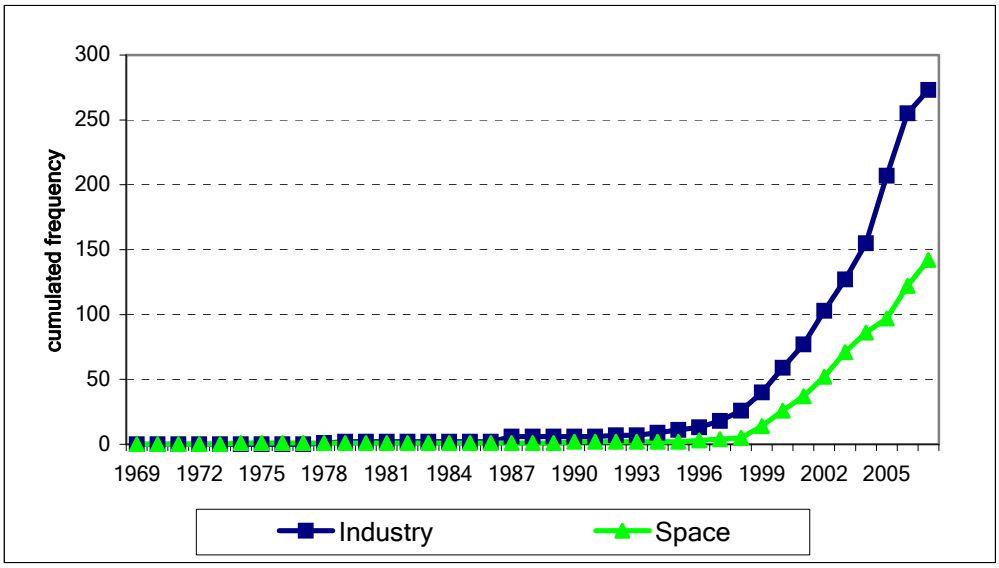
Based on these very primitive and intuitive findings, we thought about testing the possible interactions between these two elements (space and industry). In order to operationalise such an enquiry we searched the textual corpus for two groups of concepts.

The first – labelled as "industry" (Ind) and containing the following words: industrial cluster, industrial clusters, industry cluster, industry clusters, industrial clustering, industry clustering – recorded 273 entries in the period 1969-2007; the second – labelled as "space" (Spa) and containing the following words: spatial clusters, spatial clustering, regional cluster, regional clusters, regional clustering, local cluster, local clusters, local clustering, geographical cluster, geographical clusters, geographical clustering, geographic clustering, locational clusters, localized clusters – recorded 142 entries in the same period.

The temporal evolution of both groups of words, as illustrated in figure 4, seems to follows a common non-linear S-Shaped pattern with a slow start, an intermediate rapid growth and) a final period of (slightly) lower growth, although with some variation between the two groups.

[21] The works containing the above mentioned quotation record over 6.000 entries on Google Scholar on April the 16th 2009.

*Figure 4: Cumulated entries of "Industry" vs. "Space" groups of words*




Source: our calculations on ISI and EconLit databases.

We therefore decided to model the evolution of both groups of words within the theoretical framework of population ecology[22] and, in particular, to the co-evolution of two interacting populations as a system of differential equation as follows:

$$\begin{cases} \dfrac{dn_1}{dt} = \left(a_1 n_1 + a_{11} n_1^2 + a_{12} n_2\right) \\ \dfrac{dn_2}{dt} = \left(a_2 n_2 + a_{22} n_2^2 + a_{21} n_1\right) \end{cases} \qquad (2)$$

where the growth of each population, in any moment of time, is a non linear (quadratic) function of its own stock and the stock of the other interacting population.

In the phenomenon at study we can model the growth of two distinct "sides" of the cluster concept (the spatial and the industrial side) within the textual corpus represented by the titles and abstract of the papers included in our original dataset presented in section 2.

From the patterns depicted in figure 4 for each population ("industry" and "space") we expect the sign of $a_1$ and $a_2$, i.e. the coefficient on its own stock, to be positive (otherwise we could not observe any growth), while $a_{11}$ and $a_{22}$, the coefficient of the quadratic term, to be negative (therefore assuming that a congestion effects emerges after a given level is reached). This would mean that the use of a "space" ("industry") related word in the title or abstract of a published paper in a given year should firstly increase (and after a threshold is reached, decrease) the likeness that similar words are used in the following year in the textual corpus due to two concurring phenomena.

The first is related to the behaviours of authors. Scientists are known to follow fads and trends[23]. When a given approach is appearing as dominant in a given field, authors tend to write papers

[22] Which has been previously applied by one of the author (Maggioni, 2002; Maggioni and Riggi, 2008) to the analysis of the development of high-tech cluster in the US.

[23] See, among others, Bikhchandani et al. (1992); van Dalen and Klamer (2005); Khalil (2006);.

according to that very approach in order to maximise the probability of their papers being published. However this trend, based on increasing return mechanisms, is counteracted by the desire to emerge as original contributors to a particular field and the temptation to reduce the competitive pressure by establishing his/herself within a small and protected "academic niche".

The second is related to the behaviours of journals editors. Editors are concerned with the relevance of their own journal (which is increasingly measured through "objective bibliometric indexes" as the impact factor); therefore they have an incentive in publishing papers belonging to "rising trends" in order to maximise the probability that the articles published in their journal are cited within other "relevant" journals. However also these self-reinforcing dynamics have a concave shape since fads and trends in science, as in fashion, tend to have a limited life span. Thus a "good" editor, trying to jump from one fad to another, could, in theory, forecast the end of a trend and cause it effectively with his/hers own decision.

A different reasoning concerns the interaction coefficients, $a_{12}$ and $a_{21}$, which, in theory may assume different values (and signs) corresponding to very different relations between the two concepts[24]. We therefore tested an empirical version of equation (2) and estimated it using a Poisson estimation which is designed to cope with count data (Long, 1997). Table 7 shows the empirical results.

*Table 7: The ecological model of "industry" and "space" concepts*
*Poisson estimations*

| | dInd/dt | dSpa/dt |
|---|---|---|
| Ind | 0,0742*** | 0.0012 |
| $Ind^2$ | -0,0001*** | |
| Spa | -0,061*** | 0.0586*** |
| $Spa^2$ | | -0.0003*** |
| Wald | 3010,89 | 997.41 |
| Log likelihood | -72,4379 | -71.3095 |
| n. obs | 39 | 39 |

*** significant at 1%; ** significant at 5%

The results show the existence of an amensalistic relation between "industry" and "space" where the most diffused concept ("industry") develops along a logistic pattern (as defined by the positive coefficient on its own stock and a negative one on its own stock square) which is counteracted by the existing stock of space-related words in the corpus (column 2); while the opposite does not hold (column 3). "Space" follows a similar logistic development pattern, in isolation, but is left undisturbed by the existing stock of the "rival concept".

Within the biological realm, amensalism occurs when one organism exudes a chemical compound as part of its normal metabolism that is detrimental to another organism. The most common examples of this kind of relations are Penicillium, secreting penicillin, a chemical that kills bacteria, and the black walnut tree (*Juglans nigra*), which secrete juglone, a chemical that harms or kills some species of neighboring plants, from its roots. This kind of interaction thus increase the fitness

[24] By attributing positive, negative or null values to these coefficients, it is possible to describe very different interactive behaviours ranging from synergetic to competitive, from predation to amensalism (for a complete discussion, see Maggioni, 2005)

of the non-harmed organism (in our case: the “space-related” set of words) through the reduction of competition thus allowing an easier access to a given set of scarce resources (the total amount of pages published each year by all journals included in our database). In this sense the impeding species can be said to be negatively affected by the other's very existence, making it very similar to a predatory interaction[25].

A first explanation for these results refers to the fact that, in the cluster literature (contained in our sampled corpus), the use of “space-related” words is rather limited as compared to “industry-related” terms. Thus its development goes unnoticed, and therefore unharmed, by the development of “industry-related” terms; they are not really competing on the same resource set (the total textual corpus). A second explanation refers to the lower carrying capacity of “space-related” words. It may well be the case the further development of “space-related” words are already inhibited by the “decreasing returns” phase of their inner development pattern to be “disturbed” by the rival’s stock.

## 4. The clustering of cluster literature

With this expression we refer to the analysis of the “geographical” location (and co-location) of people contributing to this literature, i.e. writing on clusters. We thought it would be interesting to know whether people working on clusters are effectively clustered. However, in order to measure location and co-location, it is necessary to define the meaning of the word “geography” in this particular context.

Since we are dealing with a meta-analysis based on what has been written by people working on clusters, we thought that our analysis should be based on both a “physical” geography (where each author is assigned to a specific point on the earth surface where his/her institution is located) and a “textual” geography (where each author, or better the set of authors belonging to the same country, is assigned to a specific point in a multidimensional hyperspace composed by all the words contained in the vocabulary of our textual corpus). The following sections deal with these different perspectives.

### 4.1 “Physical” geography

In section 3.1 we analysed how the behaviour of scholars in general (i.e. average number of authors per paper) evolved over time, and we focused on “New” and “Old” authors’ behaviour, irrespectively of the role played by geographical proximity. However it seems interesting to know whether or not people writing on clusters do agglomerate in particular geographical areas, because of the existence of agglomeration economies, market-pull effects or other forces.

Hence in the following two paragraphs we study the role played by geographical proximity/dispersion of authors in shaping the evolution of the “cluster literature”.

In particular, in section 4.1.1, we detect how the average geographical distance among scholars writing on clusters is changing over time, while, in section 4.1.2, we focus on the evolution of the average distance within networks of co-authors. Thus, while in the former paragraph we detect a measure of a geographical “attribute” of authors (each author has been geo-coded according to his/her affiliation); in the latter paragraph we introduce a measure of “relations” based on people writing together a paper on clusters (each paper written in co-authorship constitutes a tiny network where each co-author could be geo-coded, hence the average distance among co-authors could be computed to detect the role played by geographical proximity/distance in determining the likelihood of co-authoring).

[25] In which the value of the interaction coefficients should be both significantly different form zero and, in particular, positive for the predator and negative for the prey.

### 4.1.1 Where do people writing on clusters locate?

In this section we analyse how the geographical proximity/distance among scholars writing on clusters evolved over time, to detect the presence (or not) of clustering driving forces that influence the locations and productivity of people studying this phenomenon.

To detect these forces, we adopted two complementary procedures, by calculating the average geographical distance per each year and the Moran'I index per each statistical period, that capture different geographical attributive aspects of authors' locations. According to the first procedure, we calculate the average geographical distance among people writing on clusters; while according to the second procedure we compute a measure of clustering (Moran'I) that identifies the presence, or absence, of spatial autocorrelation.

As described in section 2, we firstly geo-coded each affiliations appearing in the database according to the city, hence for each year we computed the geographical distances among couplets of cities[26] mentioned in the affiliations, irrespective to the ranking (primary vs. secondary affiliations)[27]. Figure 5 reports the evolution of the average distance among cities in kilometres and it clearly shows that geographical distance is increasing over time, in particular after 1994, signalling that there is a worldwide diffusion pattern followed by scientists writing on this topic.

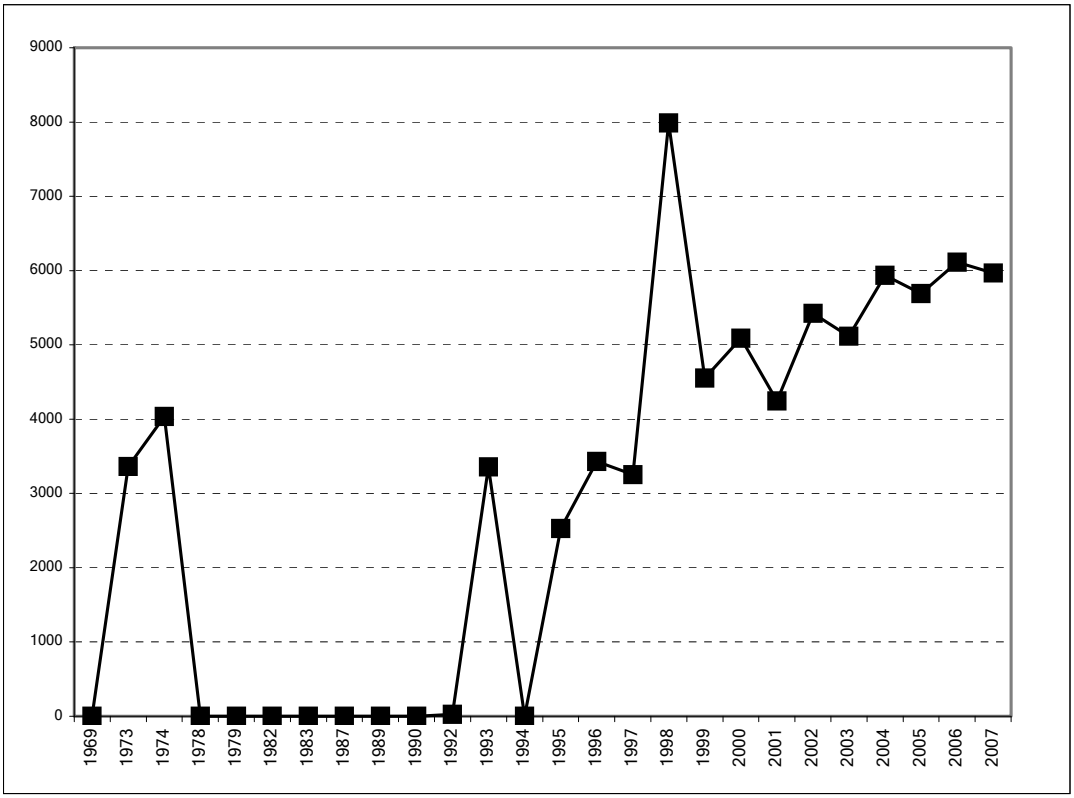
During the first years, due to the extremely small number of articles and authors involved[28], the average distance among authors is generally equal to 0, but after 1994 there is a continuous upward trend of geographical distance, identifying an extreme delocalisation of people writing on clusters.

[26] Once identified the latitude and longitude of 282 cities mentioned in the affiliations of researchers, we computed all geodesic distances (Jenness, 2005).

[27] ISI and EconLit database do not distinguish between primary and secondary affiliations. Hence due to this reason, we maintained both information and treated them equally.

[28] Indeed table 1 shows that the number of papers per year is very low (1 article per year) and similarly the average number of authors per paper is equal to 1. Hence by definition the average distance is zero.

*Figure 5: Average geographical distance (km)*




Source: our calculations on ISI and EconLit databases.

If we collapse the yearly data by using the "historical" criterion, it emerges that the average geographical distance extremely increased over time: from 740 km on average in the first period, to 3.142 km in the second period to 5.448 km in the last period (a value 7 times larger than the value of the first period).

These results are confirmed adopting the statistical subdivision of the whole period: the average geographical distance increases continuously, although values are very different since the number of articles included in periods influences the average values. In particular during the first period the average distance is equal to 1.981 km, in the second period the distance is 4.928, in the third period is equal to 5.813, and finally, during the last period, it is equal to 6.041 km.

This evolution of the geographical distance shows that, at the very beginning, the "community" of scientists writing on clusters were mostly located in neighbouring countries; but later on, the diffusion of ICTs, the tremendous reduction of transport costs, and the increase of international mobility of researchers, contribute to diffuse clusters literature worldwide[29]. Hence the reduction of these costs enabled people, physically located at longer distances respect to the original scientific community, to came across, discuss about and, possibly, publish in "established" journals. For example, nowadays thanks to the diffusion of user-friendly and easily accessible search engines it is easier to access papers (mimeo and draft versions in particular) containing new ideas and topics irrespective to the location of their authors.

In addition to this simple analysis of average distance – to identify the presence/absence of clustering forces – we computed the Moran'I index of spatial autocorrelation (Anselin, 1988)

[29] Thanks to the diffusion of user-friendly and easily accessible search engines on the Internet it is now easier to access papers (mimeo and draft versions in particular) containing new ideas and topics irrespective to the location of their authors than it was 10 years ago.

relative to the distribution of authors, located in 282 cities, for 4 statistical periods[30]. In addition we replicated this calculus limiting the scope of the analysis to European cities and American cities[31].

Moran'I index individuates the presence of overall spatial correlation among observations distributed over the space[32] and is calculated respect a weights matrix[33], in this case being the inverse of geographical distance matrices among cities.

The results obtained are stable and are replicated at continental levels (Europe and America). While in the first period (1969-2000) spatial autocorrelation emerges (the Moran'I index is significantly different form zero at 5% l.o.s, positive and equal to 0.041), thus suggesting that geographical spillovers may be at play; the existence of spatial autocorrelation disappears in all subsequent periods (Moran'I indexes, respectively equal to 0.021, 0,004, and -0,003 in the second, third and fourth statistical periods) are no longer significant[34].

A similar analysis has been performed on the number of paper written each year in every city included in the database. This analysis confirms the results based on the number of authors and shows a significant and positive spatial autocorrelation for the first and second periods.

The results on the clustering behaviour confirm the outcomes of the geographical distance analysis: until year 2000, when the literature on clusters was in the take-off period (figure 1), people should be located nearby to take advantage from their physical proximity and agglomeration, but later on, physical proximity loosed its importance in determining the productivity of a scientist and the possibility of being exposed to new "trends".

The "death of distance" hypothesis seems thus to be confirmed for our sample: after year 2000 clustering forces are no longer influencing the location of people writing on clusters. However, these results are based on an attributive perspective that underscores the role played by relational factors. In particular this perspective does not capture the role played by geography in influencing scientific collaborations, as it will be described in next section.

---

30 Since the calculus of a Moran'I index requires a significance test, related to the total number of observations, we did not computed it for single years and for the division of the whole period according to the skewed historical procedure. The relatively small number of observations would have strongly affected the results.

31 We computed the same indexes for the Asian continent and for the Oceania but, probably linked to the small number of observations, the Moran'I values were always not significant for all periods considered.

32 Moran'I index is an extension of a Pearson correlation index and individuates the presence of overall spatial correlation among all observations (Anselin, 1988). This index ranges between -1 (negative spatial autocorrelation, i.e. random distribution across space) and +1 (positive spatial autocorrelation, i.e. like values tend to be located nearby) and is computed as follows:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(a_i^t - \bar{a}^t\right)\left(a_j^t - \bar{a}^t\right)}{\sum_{i=1}^{n}\left(a_i^t - \bar{a}^t\right)^2}$$

where $w_{ij}$ indicates the element of the weights matrix , $a^t_i$ and $a^t_j$ indicate the number articles, counted as number of cities affiliations (either double counting and not double counting co-authors living in the same city) written in sub-period $t$ in cities $i$ and $j$, $\bar{a}^t$ is the mean value of articles written during a sub-period $t$, $n$ is the number of cities ($n$ = 282 in the whole population, $n$ = 143 in the European continent and $n$ = 89 in the American continent), $S_0$ is the sum of all $w_{ij}$, $S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$.

33 Weights matrix are the matrices used to consider the role played by geographical space in determining the diffusion of a socio-economic phenomenon. Usually weights matrices are distance matrices, either binary or not. In some cases it is possible to identify a weights matrix calculated respect to other relations, not exclusively geographical – i.e. the presence of search collaborations, internal migration – that influence socio-economic social phenomena (see Maggioni et al. 2007; Cracolici and Uberti 2009).

34 In Europe Moran'I values are respectively 0,028 (significant at 5%), 0,003, 0,003 and -0,022 not significant; in America Moran'I values are 0,078 (significant at 5%), 0,009, -0,003 and 0,02 not significant.

### 4.1.2 Is geographical distance relevant for co-authorships?

As already mentioned in section 3.1, scientific collaborations are increasing over time (table 1). The purpose of this paragraph is to identify the role played by geographical distance in influencing this behaviour.

In order to perform this analysis, we applied Social Network Analysis (SNA) techniques[35] relative to bipartite graphs to treat records contained in the database (Albert and Barabasi, 2002; Newman, 2003).

Each record is treated as a bipartite graph, i.e. two-mode network, $N_t = (Ar_t, Au_t, E_t, w_t)$ where $Ar_t$ is the set of top nodes and indicates article written at time $t$, $Au_t$ is the set of bottom nodes and indicates the author(s) writing $Ar_t$, $E_t \subseteq Ar_t \times Au_t$ is the set of links (edges) at time $t$ and $w_t$ is a weight, $w_t : E_t \to \Re$ and we are able to assign a network matrix $W = [w_{Au,Ar,t}]$ with elements $w_{Au,Ar,t} = w(Au_t, Ar_t)$ for $(Au_t, Ar_t) \in E_t$ and $w_{Au,Ar,t} = 0$ otherwise[36].

Since these are bipartite graphs, the presence of a link identifies the article and its author(s) and no links exist within the top and bottom sets. Hence to identify scientific collaborations we covert $N_t$ into a one-mode networks $N_{1,t} = (Au_t, E_{1,t}, w_{1,t})$ where $E_{1,t}$ and $w_{1,t}$ are determined by the symmetric network matrix $W_{(1),t} = WW^T$, $w_{(1)Ar,Au,t} = \sum_{k \in Ar} w_{Ark} \cdot w^T_{kAu}$ and $W^T$ is the transpose of matrix W. There is an edge $\{Ar, Au\} \in E_{1,t}$ in $N_{1,t}$ iff $N_t(Ar) \cap N_t(Au) \neq \emptyset$.

Out of a total of 499 articles, 264 were written in co-authorship, involving a total of 683 authors (see tables 1 and 8). Figure 6 synthesises, for each year, 4 different types of $N_{1,t}$ sub-networks: isolated authors (people writing on their own); diads and triads (sub-networks composed by two or three authors); and small networks (composed by more than 3 authors).

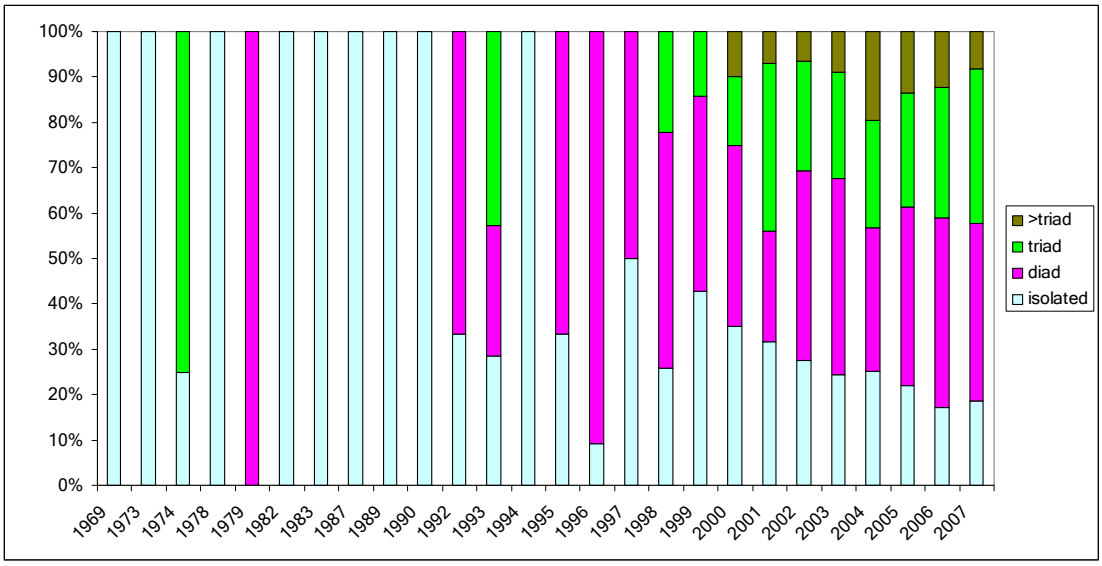
Figure 6 confirms the results described in section 3.1, since scientific collaboration (involving two or more authors per paper) is increasing over time, while the number of isolated authors is continuously reducing.

Further SNA perspective allows to investigate the collaboration behaviour into greater details: while the share of articles written by two persons (diads) is constant over time; triads and sub-networks started diffusing in 1998, with continuously increasing percentages. This behaviour could be explained by the fact that writing a paper involves more and more a division of labour reflecting the competencies of different scholars, and that a paper is more and more the final result of different theoretical and empirical competencies.

[35] Wasserman and Faust (1994) and Goyal (2007) constitute two valuable textbooks on social network analysis techniques.

[36] In this case the weight is equal to 1.

*Figure 6: Isolated authors, Diads, Triads and Small groups*




Source: our calculations on ISI and EconLit databases.

To investigate the role of geography in shaping the scientific collaboration behaviour we computed and monitored how the average distance among co-authors evolved since 1974, since before all papers in the database are single authored (see figure 6). Hence for each sub-network, identifying a scientific collaboration between two or more scientist in a paper, we computed the average geographical distance. In general the average distance among collaborations is quite stable over time, with an average value nearly equal to 1.000 km and contained within 2.000 km. But analysing the evolution of its maximum values, it clearly emerges that these values are increasing over time.

Therefore these results confirm that the reduction of communication, transportation and search costs due to the globalisation process and to the diffusion of ICTs improved the possibility of establishing scientific collaborations at longer distance, surprisingly achieving the slogan of the "death of distance". But these could also be the effect of the increasing worldwide mobility of researchers that, thanks to the reduction of transport costs, could maintain multiple affiliations in different cities.

In table 8 we identified 4 distance-related groups to summarize these results. Firstly, to capture the role played by face-to-face collaborations, we identified a group whose geographical distance is equal to zero (i.e. co-authors are placed in the same city); the second group of geographical distances ranges between 1 and 500 kilometres (which may identify the possibility to daily commute from one author's to another author's location); the third group identifies a long distance, between 501 and 3000 kilometres; the last group where the geographical distance is greater than 3000 kilometres identified very long distance.

Statistics in table 8 show that the number of papers written at zero distance is increasing over time despite the possibility to travel at longer distances and to communicate using low-price technologies. Thus the relevance of urban *buzz* in the process of transformation of tacit into codified knowledge and its replication and diffusion by word of mouth it is still very important (Storper and Venables, 2004).

*Table 8: Number of articles in co-authorships per km*

| Publication Year | 0] | 0-500] | 500-3000] | >3000 | Total articles in co-authorship | Total articles |
|---|---|---|---|---|---|---|
| 1974 | 0 | 0 | 1 | 0 | 1 | 2 |
| 1979 | 1 | 0 | 0 | 0 | 1 | 1 |
| 1992 | 0 | 1 | 0 | 0 | 1 | 2 |
| 1993 | 1 | 0 | 0 | 0 | 1 | 4 |
| 1995 | 1 | 0 | 0 | 0 | 1 | 2 |
| 1996 | 2 | 2 | 0 | 2 | 6 | 8 |
| 1997 | 2 | 0 | 0 | 0 | 2 | 6 |
| 1998 | 5 | 0 | 1 | 3 | 9 | 17 |
| 1999 | 4 | 5 | 1 | 1 | 11 | 30 |
| 2000 | 7 | 1 | 1 | 2 | 11 | 26 |
| 2001 | 7 | 5 | 2 | 2 | 16 | 37 |
| 2002 | 11 | 4 | 3 | 2 | 20 | 38 |
| 2003 | 13 | 8 | 6 | 6 | 33 | 59 |
| 2004 | 18 | 7 | 5 | 8 | 38 | 70 |
| 2005 | 21 | 8 | 9 | 5 | 43 | 75 |
| 2006 | 19 | 7 | 6 | 5 | 37 | 57 |
| 2007 | 20 | 5 | 4 | 4 | 33 | 55 |
| **Total** | **132** | **53** | **39** | **40** | **264** | **489** |

Source: our calculations on ISI and EconLit databases.

As the geographical distance increases (see columns 3, 4, 5 in table 8), the number of co-authored papers per year increases, but at much more slower rate. This signals that, although transportation and communication costs have been declining over the last years, this is not a sufficient condition to activate scientific collaborations over longer distances because co-location (even temporary co-location) seems to be a necessary condition to start a scientific collaboration. Co-authorships at very long distances (more than 3,000 km) are also increasing, but at a lower rate (Frenken, 2008; Baldi, 1998).

These results suggest that while new communication and information technologies enable relationships over longer distance being maintained at very low costs, and the drop of transportation costs allowed a greater mobility of people around the globe; spatial proximity, that enforces personal relationships, is still essential in the process of knowledge creation.

### 4.2 "Textual" geography

Since we are dealing with a meta-analysis based on a bibliographic database it seemed sensible to further exploit the textual corpus looking for national characteristics "hidden" in the vocabulary used in titles and abstracts.

As explained in paragraph 3, the research interests of a scientist are affected by the social preferences of his/her scientific community. Interactions in the scientific community are crucial for growing competencies, exploiting scale economies in the production of scientific knowledge, accessing funding sources and research facilities, etc. In the previous sections we showed that the collaborative behaviours and strategies of scientists are sensitive to location of scholars, while, at

the same time, knowledge diffusion and cross-fertilization of "paradigms" can be possible through alternative "mobility" opportunities such as: seminars, conferences, scientific partnerships at international level.

The balance/trade-off between local and global interactions, between "local buzz and gobal pipelines" (Bathelt et. al., 2004), is one of the crucial issue in science and technology policy. For this reason we applied correspondence analysis (CA) to the study of the evolution over time of different research communities (scientists belonging to the same countries and participating in the cluster literature), to detect specialization patterns and the presence of convergence/divergence dynamics.

CA is a multivariate statistical technique that reduces distances (proximities) of dimensional spaces (i.e. rows and columns of a matrix) into a simple visual representation using a chi-square measure (Lebart et al., 1998)[37]. In our case, CA graphically represents associations among word-types (rows) and the paper authors' nationality (columns)[38].

This graphical representation of the vocabulary contingency table allows us to identify a lower-dimensional space approximating the original distributions of word-types. In the following graphics (figures 7, 8, 9) the origin of the principal axes represents the mean profiles of our analysis (i.e. the marginals of the table of frequencies), while points in the representative space tend to cloud in such a way that distances (i.e. dispersion or, technically, moment of inertia) produce different profiles within both scholars' nationality space and the <cluster> topics space in order to detect the existence of meaningful relations between <cluster> research topics and the space of scholars' nationality. The size of triangles represents the relative contribution of each national community of scientists to eigenvalues; while the size of circles reflects the importance of specific words in determining the "textual" distance among national communities (see figures from 7 to 10).

In this section we present the outcomes of CA of textual longitudinal data for the first two eigenvalues (the most relevant ones) that measure the variances along each principal axis and that, cumulatively, absorb more than the 50% of inertia for each period. We run CA for each statistical period and for nine most relevant countries in order to observe the evolution over time of the major national scientific communities contributing to the cluster literature. Given that the distribution of paper of our textual corpus approximates a logistic function, we chose the statistical distribution of papers to have enough textual variability for each time period.

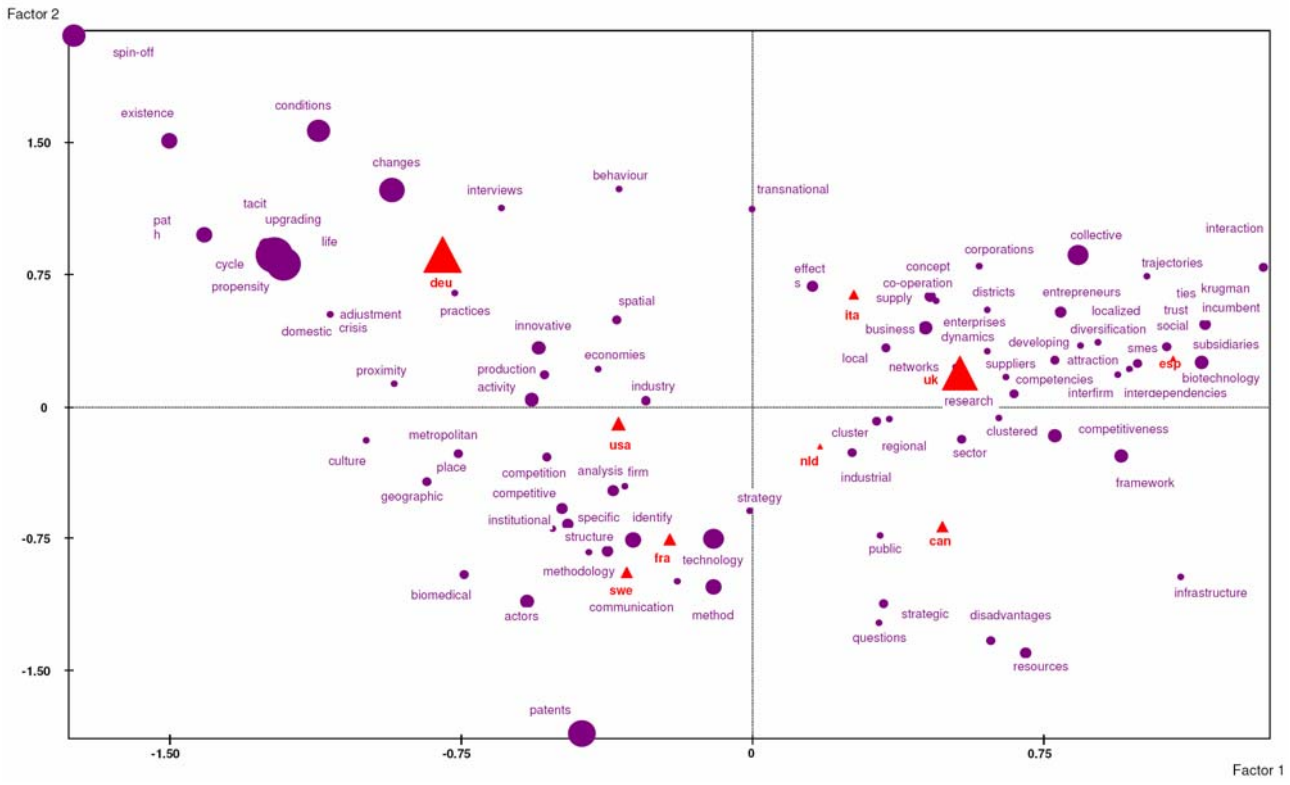
In Figure 7, referring to the first statistical period 1969-2000 (108 papers), the configuration of research topics presents four separated clusters: starting from the third quadrant, we observe that scientists communities in Sweden (SWE) and France (FRA) focus on similar topics in both in industrial investigation, addressing <technology> (e.g. <patents>, <biomedical>), and territorial analysis (<place>, <metropolitan>, <geographic>). A bridging position between these two national scientific communities and the German (DEU) one is played by US scientists (USA) which focus their research on the relationships between territory (e.g. <proximity>, <economies>, <spatial>, <activity>) and the industrial economic dynamics (e.g. <spin-off>, <life cycle>, <conditions>, <changes>). A different research approach is adopted by British scientists (UK), and, by a smaller extent, by Italians (ITA) and Spanish (ESP). These communities look for key factors of local development (e.g. <interdependencies>, <trust>, <networks>, <social>, <collective>) within a "regional science" theoretical framework (<cluster>, <regional>, <local>, <district>). Dutch scientists (NLD) seem to play an bridging role between the communities located in the first and third quadrants by focussing on the core terms of this literature: <cluster>, <strategy>, <industrial>,

[37] Remember that CA is an explanatory method and not a confirmatory one. This means that it is used not to test hypotheses but to describe data and to get an idea of the structure of variables in order to develop hypotheses. For an exhaustive presentation of CA see Greenacre (1989), Greenacre and Blasius (1994), Lebart, et al. (1998).

[38] As explained in section 2, we limited our analysis to the most "relevant" nine countries in our database, accounting for about 90% of the total number of papers.

<regional>. A similar, but more peripheral position, is occupied by Canadians (CAN) whose research interests are focused on: <infrastructure>, <resources>, <public>, <disadvantages>.

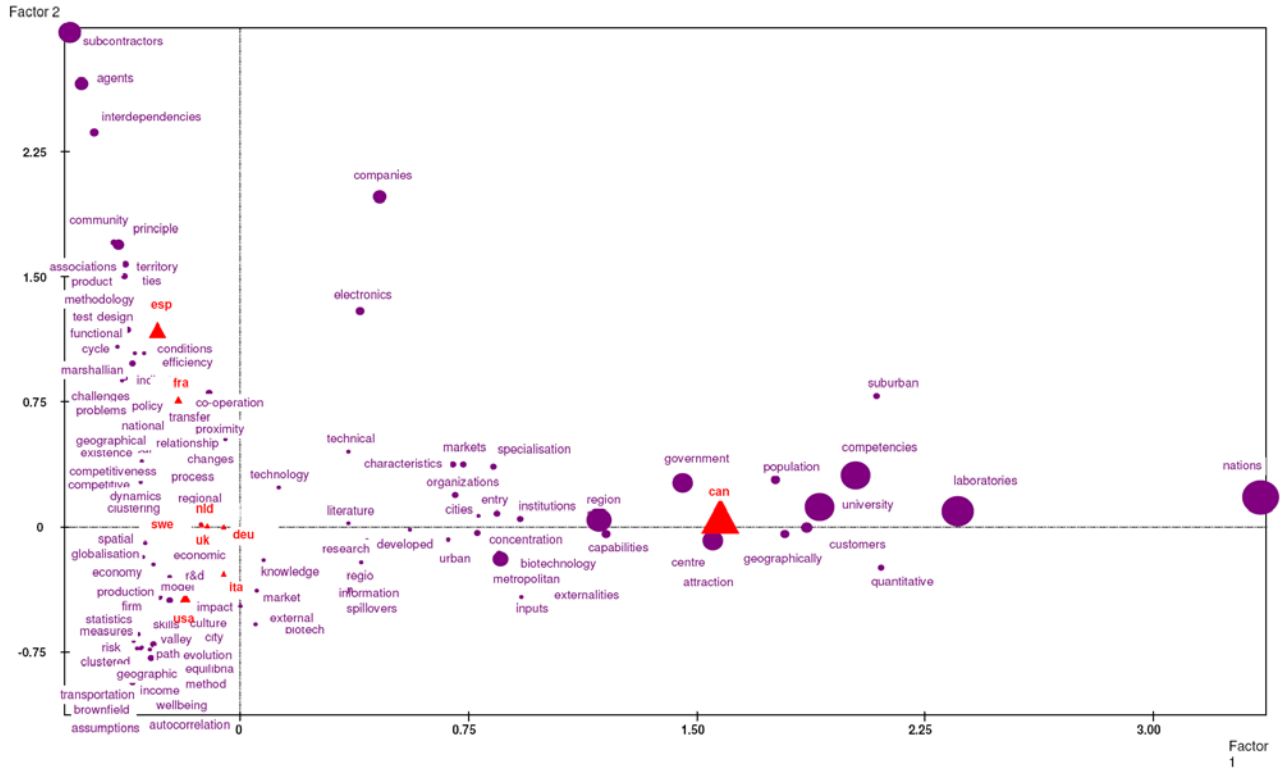
*Figure 7: CA for the first period, 1969-2000*




Source: our calculations on ISI and EconLit databases.

In Figure 8, referring the second statistical period 2001-2003 (134 papers), it is noticeable a strong dominance of the Canadian research community devoted to the analysis of local innovation systems in terms of relationships between <university> and <laboratories>. Despite the fact that the remaining countries are "squeezed" in the left side of the diagram, it is possible to identify two distinct research approaches: a quantitative and formalised one (<autocorrelation>, <assumptions>, <equilibria>) performed by US scientists and an institutionalist approach (<subcontractors>, <interdependencies>, <agents>, <community>, <product>) put forward by Spanish and French communities. In this period ITA, NLD, SWE, UK contribute very little to the definition of the first two eigenvalues even if they are relevant other three factors explaining the total inertia.

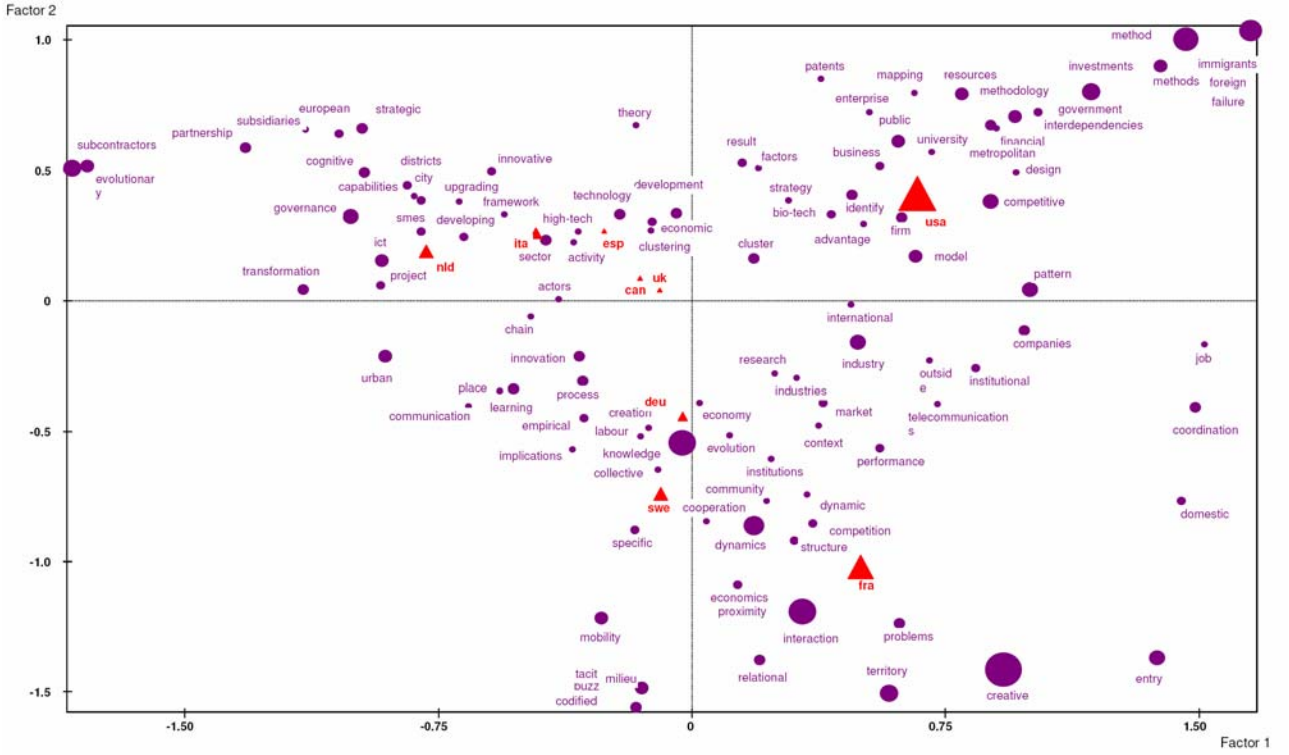
*Figure 8: CA for the second period, 2001-2003*




Source: our calculations on ISI and EconLit databases.

Figure 9, referring to the third statistical period 2004-2005 (145 papers), presents a more equilibrated (triangular) configuration, and a higher number of national research communities which equally contribute to the "cluster literature". SWE and DEU (third quadrant) share investigation in the process of knowledge production (<learning>, <innovation>, <process>, <mobility>) and social communication (<tacit>, <buzz>, <codified>) as well as <labour> market and <mobility>. Moving to the second quadrant, we observe a "cluster" of countries: NLD, ITA, ESP, CAN, UK. These communities seems to be focused on an <evolutionary> perspective applied to cluster analysis. Technology is considered as a major driver of cluster development (<ict>, <high-tech>, <technology>) and there is lot of interest in the emergence of co-ordination procedures and structures (<governance>, <upgrading>, <transformation>, <partnership>, <strategic>). The first quadrant hosts only the US community, specializing in quantitative investigation with a strong emphasis on the methodological side (<methodology>, <method>, <methods>, <model>) where the key factors of cluster development are several actors (<university>, <enterprise>, <companies>, <industry>), mobilising capitals (<foreign> <investments>, <financial>, <resources>). FRA community stands alone in the fourth quadrant. Its research is focused on the relationships between territory, institutions and social contexts (<coordination>, <interaction>, <relational>, <proximity>, dynamics>, <community>, <cooperation>).

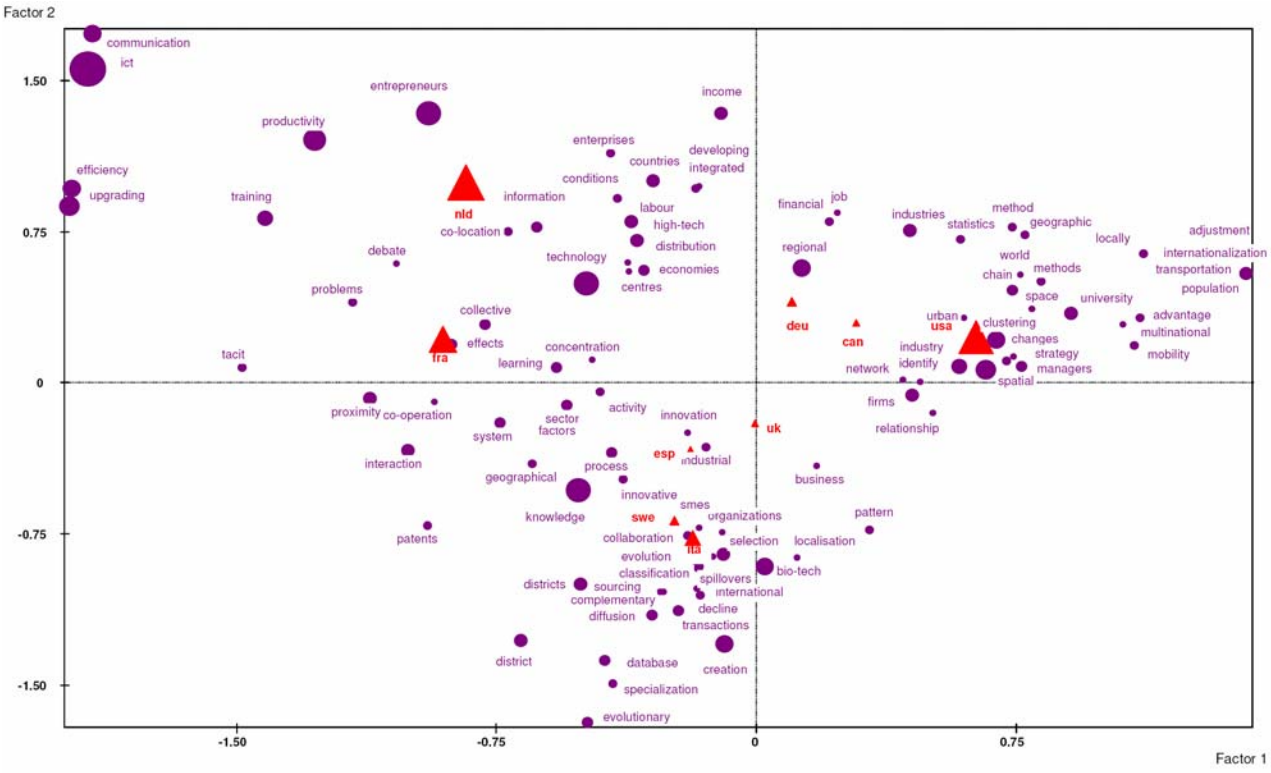
*Figure 9: CA for the third period, 2004-2005*




Source: our calculations on ISI and EconLit databases.

In Figure 10, referring to the fourth statistical period 2006-2007 (112 papers), ITA, ESP and SWE (third quadrant) appear to have focused on the interaction of industrial and organisational concepts (<district>, <knowledge>, <organization>) in explaining the crucial role played by clusters and their internal structures of interdependences (<proximity>, <co-operation>, <complementarities>) in shaping the innovation process (<bio-tech>, <diffusion>, <spillover>, <selection>, <smes>). NLD and FRA share common interests for advanced technologies (<ict>, <high-tech>) and the study of the spatial conditions (<economies>, <concentration>, <co-location>) that sustain <co-operation> and produce <efficiency> and <productivity>. In this period, USA share with DEU and CAN the interests for methodological issues (<method>, <statistics>) and for the analysis of the spatial re-allocation of production due to globalisation dynamics (<internationalization>, <multinational>, <mobility>, <strategy>, <job>, <regional>). UK scientists play a minor role in the literature with a small contribution to peculiar research topics as defined by the first two eigenvalues of CA.

*Figure 10: CA for the fourth period, 2006-2007*




Source: our calculations on ISI and EconLit databases.

Looking at the evolution of these four diachronic CA graphs, it is possible to conclude that, in general, each national scientific community shows a significant degree of continuity in the choice of research topics. The contribution of Canadian scientists is based on one hand on innovation systems and policies, on the other on the process of knowledge diffusion; attention is paid to the role of public authorities in the process of knowledge production, and to the relationships between University and Industry. The research path of US scientists is dominated by the investigation of a quantitative-assiomatic methodology for studying clusters and clustering. North-American scholars focus on theoretical models and econometric analyses aimed at interpreting and explaining the performance of clusters in terms of knowledge production and diffusion. This focus is partially shared by Canadian and German scientists.

The existence of a common "European Research Area"[39] does not appear to have shaped the research activity of scholars contributing to the cluster literature. ITA and ESP deepened their investigation over time on the Marshallian concept of "industrial district" through theoretical and empirical analyses on technology, competitiveness, social capital and co-ordination devices.

FRA has developed a specific research model, rooted in Perroux's thought. Analyses are focused on the relationships between space and technology (e.g. growth poles) assuming, as basic ingredients, cognitive and institutional dimensions. In the first period NLD showed some similarities with CAN but, more recently, it moved its attention to knowledge production and regional evolution. SWE appears to be very focused on knowledge production and diffusion while UK combines the interest for both neo-classical and institutionalist approaches. DEU is the European country closer to USA with respect to research interests; however, it also cultivates research questions, more similar to SWE, concerning the spatial dynamics of local economies, and the dynamics of knowledge production and diffusion.

[39] To quote the title of the famous communication of the European commission (European Commission, 2000).

In conclusion, CA shows that specialization and differentiation dynamics in the choice of research trajectories by different national scientists' communities coexist. Research communities build their own social preferences (qualitative vs. quantitative methods; neoclassical vs. istitutionalist-evolutionary schools; cognitive vs. behavioural approaches, etc.) and their dynamic is strictly connected with these theoretical choices.
However, at the same time, scientists are often called to confront different (if not opposite) positions in workshops, conferences and, more rarely, journals, and, from these interactions, the robustness of the analyses is increased and theoretical concepts are further refined. This method of comparison of different ideas and knowledge diffusion process produces, in some cases, a strong homogenization of research topics (which is labelled, in our empirical exercise, as "convergence of research interest") as well as a segregation of original thoughts which cannot be easily integrated in the dominant paradigm. This process may be further re-enforced by the oligopolistic structure of the demand side of the "research market" (i.e the academic journals), which is increasing over time.

### 4.2.1 Inter-country distance in correspondence analysis

The previous section described the use of CA as a tool for locating authors (grouped according to their affiliation-based "nationality") within a "textual geography" defined by the use of specific word-types. CA is indeed a powerful descriptive instrument to show the existence of similarities and differences in the approach followed by scientists of different countries when studying the same (or, better, similar) topics. However, in order to produce sharper results, CA must reduce a complex multidimensional world into a two dimensional Cartesian space based on two factors (eigenvalues).

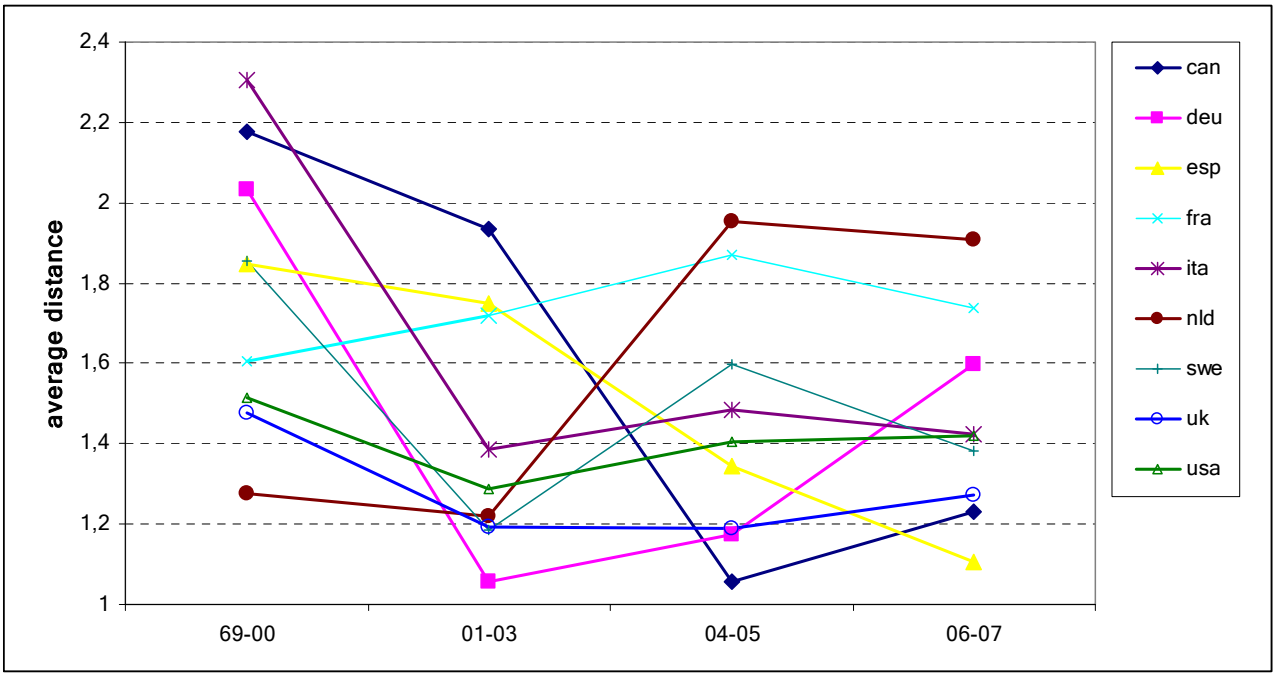
Looking for a less reductionistic approach we built a "hyper-spatial textual distance matrix" where, in each period, each nationality is assigned to a specific point within a 5-dimensional hyperspace which is composed by the factors accounting for about 75% of the "moment inertia" of the vocabulary contained in the textual corpus and where distance between each and every nationality, *i* and *j*, is calculated according to the following expression:

$$d_{ij}^{t} = \sqrt{\sum_{n=1}^{5} \left(i_n - j_n\right)^2} \qquad (3)$$

where *n* are the 5 largest CA factors and *t* is the "statistical" period.

We then calculated $\overline{d_{i\bullet}^{t}} = \frac{\sum_{j=1}^{8} d_{ij}}{8}$ (i.e. the average distance of each nationality from all the other ones for each statistical period) and, finally, we plotted the time evolution of this average distance for each nationality. Results on the evolution of these national distances are plotted in figure 11.

*Figure 11: Evolution of textual distance between nationalities*




Source: our calculations on ISI and EconLit databases.

Figure 11 shows that, despite the presence of a wide variety of non monotonic evolution, there is a general declining trend of the average textual distance, with the exception of France and the Netherlands, which signal a progressive convergence in the vocabulary of the authors belonging to 9 major countries writing on clusters[40].

This result could alternatively be interpreted either as the positive and progressive emergence of a well connected and compact international research community, or as the negative effect of a reduction of national specificities induced by a "consensus based" scientific publication systems in which a limited number of scientific journals determine the academic careers of scientists all over the world (Frey, 2003).

## 5. Conclusion

This paper, based on an original database, developed by the authors on the basis of two major independent bibliographic databases (ISI and EconLit), contains a meta-analysis of the "cluster literature" in order to describe the evolution of the concept of cluster in major academic journals and to identify the existence of trends and life-cycles in both the textual corpus and the authorships.

In particular, we identified different phases of growth of this literature, from the birth of the concept to the take off phase, from the development to the current maturity. We performed a series of analyses aimed at identifying the existence of a life-cycle within the authorships of the "cluster literature" and we discovered the existence of a substitutability relation between different collaborative behaviours ("autarkic" vs. "invisible colleges" and "mentoring vs. mixed effects"). The IT index allowed us to distinguish between different trends in the use of specific word-types.

We applied a population ecology growth models – along the lines of Swann (1998) and Maggioni (2002) – to understand the intertwined relationships existing between a "spatial" and an "industrial" approach within the textual corpus of cluster literature. The empirical results showed that a sort of a

[40] In addition the ratio between the average distance of the most isolated and the most connected nationality decreased from 1.8 to 1.7.

predatory interaction between these two approaches, where the “spatial” approach predate the “industrial” one.

We performed several “physical geographic” analyses aimed at detecting the existence of clustering behaviours in the location of authors working on clusters and at measuring the influence of geographical distance in co-authorship. We also performed a “textual geographic” exercise searching for country-specific vocabulary in the cluster literature and testing the hypothesis of a convergence dynamic acting over time.

This paper is also a research agenda for the authors. Several interesting issues have been briefly analysed while they would deserve specific in depth analyses. We are convinced that through meta-analyses one could gather in depth knowledge of a given stream of literature (its nature, history and evolution) and that complementary empirical techniques (from characteristics textual analysis to population ecology, from social network and geographical analysis to correspondence analysis) could be jointly used in order to achieve this aim.

**References**

ALBERT R., BARABASI A.L. (2002) Statistical mechanics of complex networks, Rev Modern Phys 74: 47–99.

ANSELIN L. (1988) Spatial Econometrics: Methods and Models. Dordrecht: Kluwer Academic Publishers.

ASHEIM B., COOKE P., MARTIN R. (Eds), (2006) Clusters and Regional Development. Critical Reflections and Explorations, London: Routledge.

AYDALOT P. (1986) Milieux innovateurs en Europe, Paris: Gremi.

AYDALOT P., KEEBLE D. (Eds) (1988) High Technology Industry and Innovative Environment, London: Routledge.

BALDI S. (1998) Normative versus social-constructivist processes in the allocation of citations: A network-analytic model, American Sociological Review 63, 829-846.

BARABASI A. L., JEONG H., RAVASZ E., NEDA Z., SCHUBERT A., VICSEK T. (2001) Evolution of the social network of scientific collaborations, PHYSICA A, 3-4, 590-614.

BATHELT H., MALMBERG A., MASKELL P. (2004) “Clusters and Knowledge: Local Buzz, Global Pipelines and the Process of Knowledge Creation”, Progresss in Human Geography, 28, 31-56.

BECATTINI G. (1979) Dal 'settore' industriale al 'distretto' industriale: alcune considerazioni sull'unità d'indagine dell'economia industriale, Rivista di economia e politica industriale, 1, 1-79.

BECATTINI G. (Ed.) (1987) Mercato e forze locali: il distretto industriale, Bologna: Il Mulino.

BIKHCHANDANI S., HIRSHLEIFER D., WELCH I., (1992), “Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades,” Journal of Economic Perspectives, 12, pp. 151-170, 1998.

BOLASCO S. (1999) Analisi Multidimensionale dei Dati. Metodi, strategie e criteri d’interpretazione, Roma: Carocci.

BOLASCO S., CANZONETTI A. (2003) Il text mining come chiave di lettura del linguaggio degli anni Novanta su “La Repubblica”. Convegno Scientifico S.I.S. 2003 sul tema: Analisi statistica multivariate per le Scienze Economico-Sociali, le Scienze Naturali e la Tecnologia, Università di Napoli “Federico II”, 9-11 giugno 2003.

BRAUNERHJELM P., FELDMAN M. (Eds) (2006) Cluster Genesis. Technology-based Industrial Development, Oxford: Oxford University Press.

BRESCHI S., MELERBA F. (Eds) (2005) Clusters, Networks, and Innovation, Oxford: Oxford University Press.

BRESNAHAN T. F., GAMBARDELLA A., SAXENIAN A. (2001) 'Old Economy' inputs for 'New Economy' Outcomes: Cluster Formation in the New Silicon Valleys, Industrial and Corporate Change, 4, 835-860.

BRESNAHAN T. F., GAMBARDELLA A. (Eds), (2004), Building high-tech clusters: Silicon Valley and beyond, Cambridge: Cambridge University Press.

CHRISTALLER W. (1933) Die zentralen Orte in Süddeutschland. Jena- Gustav Fischer. (Translated (in part), by Charlisle W. Baskin, as Central Places in Southern Germany. Prentice Hall 1966).

CRACOLICI M.F., UBERTI T.E (2009) Geographical Distribution of Crime in Italian Provinces: A Spatial Econometric Analysis, Jahrbuch für Regionalwissenschaft-Review of Regional Research, 29, 1-27.

CRANE D. (1972) Invisible colleges: diffusion of knowledge in scientific communities, Chicago: University of Chicago Press.

DIXIT A. K., STIGLITZ J. E., (1977) Monopolistic Competition and Optimal Product Diversity, American Economic Review, 67, 297-308.

DURIAU V., REGER R.K., PFARRER M.D. (2007) A Content Analysis of the Content Analysis Literature in Organization Studies- Research Themes, Data Sources, and Methodological Refinements, Organizational Research Methods, 10, 5-34.

EUROPEAN COMMISSION (2000), http://europa.eu/scadplus/leg/en/lvb/i23010.htm, accessed march 2009.

FRENKEN K. (2008) Geography of science: a proximity approach, Utrecht: Utrecht University, mimeo.

FREY B. S (2003) Publishing as Prostitution? Choosing between One's Own Ideas and Academic Success, Public Choice, 116, 205-23.

GAMBAROTTO F., BRAMANTI A. (2009) Local policies require cognitive alignment for Firms and Institutional actors: an explorative analysis for the Italian eyewear system in Veneto, Workshop "Globalization, Local Development and Emerging Powers", Bologna, Dep. of Economic Sciences, 6-7 February.

GAMBAROTTO F., WALTER S. (2005) What social preferences prevail in the recruitment process? Differences and similarities of the Italian and German labour markets, 17th SASE Conference, Budapest, 30 June-2 July.

GLÄNZEL W, SCHUBERT A. (2004) Analyzing scientific networks through co-authorship, in MOED HFM, GLÄNZEL W, SCHMOCH U, (Eds) Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies on S&T Systems, pp. 257-276. Dordrecht: Kluwer Academic Publishers.

GOYAL S. (2007) Connections: An Introduction to the Economics of Networks, Princeton: Princeton University Press.

GREENACRE M. J. (1989) Theory and Applications of Correspondence Analysis, London: London Academic Press.

GREENACRE M. J., BLASIUS J. (1994), Correspondence Analysis in the Social Sciences, San Diego: San Diego Academic Press.

HIRSCHMAN A. (1958) The Strategy of Economic Development, New Haven: Yale University Press.

JENNESS J. (2005) Distance Matrix (dist_mat_jen.avx) extension for ArcView 3.x, v. 2. Jenness Enterprises, available at: http://www.jennessent.com/arcview/dist_matrix.htm.

KARLSSON C. (Ed.) (2008) Handbook of Research on Clusters: Theories, Policies and Case studies, Cheltenham: Edward Elgar.

KHALIL E. L. (2006) Nonlinear Thermodynamics and Social Science Modeling: Fad Cycles, Cultural Development and Identificational Slips, American Journal of Economics and Sociology, 54 4, 423-438

KRETSCHMER H. (1994) Coauthorship networks of invisible colleges and institutional communities, Scientometrics, 30, 363-369.

KRUGMAN P. (1991a) Geography and Trade, Cambridge (Mass.): MIT Press.

KRUGMAN P. (1991b) Increasing Returns and Economic Geography, Journal of Political Economy, 99, 483-499.

KRUGMAN P. (1991c) History vs. Expectations, Quarterly Journal of Economics, 106, 651-667.

LEBART L., SALEM A., BERRY L. (1998), Exploring Textual Data, Dortrecht: Kluwer Academic Press.

LONG J. S. (1997) Regression Models for Categorical and Limited Dependent Variables, Thousand Oaks, CA: Sage.

LÖSCH, A. (1940) Die räumliche Ordnung der Wirtschaft, Jena- G. Fischer. English translation (of the 2nd rev. ed.)- The Economics of Location. New Haven: Yale University Press, 1954.

MAGGIONI M. A. (2002) Clustering Dynamics and the Location of High-Tech Firms, Heidelberg: Springer Verlag.

MAGGIONI M. A. (2005) "The Dynamics of High Tech Clusters: Competition, Predation, Synergies", in Quadrio Curzio A., Fortis M., Research and Technological Innovation, The Challenge for a New Europe, pp. 110-127, Heidelberg: Springer.

MAGGIONI M. A., NOSVELLI M., UBERTI T. E. (2007) Space versus networks in the geography of innovation: A European analysis, Papers in Regional Science, 86, 471-493.

MAGGIONI M. A., RIGGI M. R. (2008) High-Tech Firms and Innovative Industrial Clusters, in KARLSSON C. (Ed.) Handbook of Research on Clusters: Theories, Policies and Case studies, pp. 55-78, Cheltenham: Edward Elgar.

MALMBERG A., POWER D. (2006) True Clusters: a severe case of conceptual headache, in ASHEIM B., COOKE P., MARTIN R. (Eds) Clusters and Regional Development. Critical Reflections and Explorations, pp. 50-68 London: Routledge.

MARTIN R., SUNLEY P. (2003), Deconstructing clusters: chaotic concept or policy panacea?, Journal of Economic Geography, 3, 5-35.

MASKELL P., KEBIR L. (2006) What Qualifies as a Cluster Theory?, in ASHEIM B., COOKE P., MARTIN R. (Eds), Clusters and Regional Development. Critical Reflections and Explorations, pp. 30-49, London: Routledge.

NEWMAN M. E. J. (2003), The Structure and Function of Complex Networks, SIAM Rev., 45: 167-256.

NEWMAN M. E. J. (2004) ‘Who is the best connected scientist? A study of scientific coauthorship networks, in BEN-NAIM E., FRAUENFELDER H., TOROCZKAI Z. (Eds) Complex networks, pp. 337–370, Berlin: Springer.

PERROUX, F. (1955) Note sur la notion de pôle de croissance, Économ. Appliquée, VII, 307-320.

PORTER M. (1990) The competitive advantage of nations, London: Macmillan.

PORTER M. (1998), Clusters and the New Economics of Competition, Harvard Business Review, 76: 77-90.

STORPER M., VENABLES A. J. (2004) Buzz: face-to-face contact and the urban economy, Journal of Economic Geography, 4, 351-370.

SWANN G.M.P. (1998), Towards a Model of Clustering in High-Technology Industries, in SWANN G.M.P., PREVEZER M., STOUT D. (Eds), The Dynamics of Industrial Clustering, pp. 52-76, Oxford: Oxford University Press.

SWANN G.M.P., PREVEZER M. (1998), Introduction, in SWANN G.M.P., PREVEZER M., STOUT D. (Eds), The Dynamics of Industrial Clustering, pp. 1-12, Oxford: Oxford University Press.

TUZZI A. (2003) L’Analisi del Contenuto. Introduzione ai metodi e alle tecniche di ricerca, Roma: Carocci.

VAN DALEN H. P., KLAMER A. (2005) Is Science A Case of Wasteful Competition?, Kyklos, 58, 395-414.

WASSERMAN S., FAUST K. (1994), Social Network Analysis. Methods and Application, Cambridge: Cambridge University Press.

**NOTE DI LAVORO DELLA FONDAZIONE ENI ENRICO MATTEI**

**Fondazione Eni Enrico Mattei Working Paper Series**

**Our Note di Lavoro are available on the Internet at the following addresses:**

http://www.feem.it/Feem/Pub/Publications/WPapers/default.htm
http://www.ssrn.com/link/feem.html
http://www.repec.org
http://agecon.lib.umn.edu
http://www.bepress.com/feem/

**NOTE DI LAVORO PUBLISHED IN 2009**

| | | |
|---|---|---|
| SD | 1.2009 | Michael Hoel: Bush Meets Hotelling: Effects of Improved Renewable Energy Technology on Greenhouse Gas Emissions |
| SD | 2.2009 | Abay Mulatu, Reyer Gerlagh, Dan Rigby and Ada Wossink: Environmental Regulation and Industry Location |
| SD | 3.2009 | Anna Alberini, Stefania Tonin and Margherita Turvani: Rates of Time Preferences for Saving Lives in the Hazardous Waste Site Context |
| SD | 4.2009 | Elena Ojea, Paulo A.L.D. Nunes and Maria Loureiro: Mapping of Forest Biodiversity Values: A Plural Perspective |
| SD | 5.2009 | Xavier Pautrel : Macroeconomic Implications of Demography for the Environment: A Life-Cycle Perspective |
| IM | 6.2009 | Andrew Ellul, Marco Pagano and Fausto Panunzi: Inheritance Law and Investment in Family Firms |
| IM | 7.2009 | Luigi Zingales: The Future of Securities Regulation |
| SD | 8.2009 | Carlo Carraro, Emanuele Massetti and Lea Nicita: How Does Climate Policy Affect Technical Change? An Analysis of the Direction and Pace of Technical Progress in a Climate-Economy Model |
| SD | 9.2009 | William K. Jaeger: The Welfare Effects of Environmental Taxation |
| SD | 10.2009 | Aude Pommeret and Fabien Prieur: Double Irreversibility and Environmental Policy Design |
| SD | 11.2009 | Massimiliano Mazzanti and Anna Montini: Regional and Sector Environmental Efficiency Empirical Evidence from Structural Shift-share Analysis of NAMEA data |
| SD | 12.2009 | A. Chiabai, C. M. Travisi, H. Ding, A. Markandya and P.A.L.D Nunes: Economic Valuation of Forest Ecosystem Services: Methodology and Monetary Estimates |
| SD | 13.2009 | Andrea Bigano, Mariaester Cassinelli, Fabio Sferra, Lisa Guarrera, Sohbet Karbuz, Manfred Hafner, Anil Markandya and Ståle Navrud: The External Cost of European Crude Oil Imports |
| SD | 14.2009 | Valentina Bosetti, Carlo Carraro, Romain Duval, Alessandra Sgobbi and Massimo Tavoni: The Role of R&D and Technology Diffusion in Climate Change Mitigation: New Perspectives Using the Witch Model |
| IM | 15.2009 | Andrea Beltratti, Marianna Caccavaio and Bernardo Bortolotti: Stock Prices in a Speculative Market: The Chinese Split-Share Reform |
| GC | 16.2009 | Angelo Antoci, Fabio Sabatini and Mauro Sodini: The Fragility of Social Capital |
| SD | 17.2009 | Alexander Golub, Sabine Fuss, Jana Szolgayova and Michael Obersteiner: Effects of Low-cost Offsets on Energy Investment – New Perspectives on REDD – |
| SD | 18.2009 | Enrica De Cian: Factor-Augmenting Technical Change: An Empirical Assessment |
| SD | 19.2009 | Irene Valsecchi: Non-Uniqueness of Equilibria in One-Shot Games of Strategic Communication |
| SD | 20.2009 | Dimitra Vouvaki and Anastasios Xeapapadeas: Total Factor Productivity Growth when Factors of Production Generate Environmental Externalities |
| SD | 21.2009 | Giulia Macagno, Maria Loureiro, Paulo A.L.D. Nunes and Richard Tol: Assessing the Impact of Biodiversity on Tourism Flows: A model for Tourist Behaviour and its Policy Implications |
| IM | 22.2009 | Bernardo Bortolotti, Veljko Fotak, William Megginson and William Miracky: Sovereign Wealth Fund Investment Patterns and Performance |
| IM | 23.2009 | Cesare Dosi and Michele Moretto: Auctioning Monopoly Franchises: Award Criteria and Service Launch Requirements |
| SD | 24.2009 | Andrea Bastianin: Modelling Asymmetric Dependence Using Copula Functions: An application to Value-at-Risk in the Energy Sector |
| IM | 25.2009 | Shai Bernstein, Josh Lerner and Antoinette Schoar: The Investment Strategies of Sovereign Wealth Funds |
| SD | 26.2009 | Marc Germain, Henry Tulkens and Alphonse Magnus: Dynamic Core-Theoretic Cooperation in a Two-Dimensional International Environmental Model |
| IM | 27.2009 | Frank Partnoy: Overdependence on Credit Ratings Was a Primary Cause of the Crisis |
| SD | 28.2009 | Frank H. Page Jr and Myrna H. Wooders (lxxxv): Endogenous Network Dynamics |
| SD | 29.2009 | Caterina Calsamiglia, Guillaume Haeringer and Flip Klijnb (lxxxv): Constrained School Choice: An Experimental Study |
| SD | 30.2009 | Gilles Grandjean, Ana Mauleon and Vincent Vannetelbosch (lxxxv): Connections Among Farsighted Agents |
| SD | 31.2009 | Antonio Nicoló and Carmelo Rodríguez Álvarez (lxxxv): Feasibility Constraints and Protective Behavior in Efficient Kidney Exchange |
| SD | 32.2009 | Rahmi İlkiliç (lxxxv): Cournot Competition on a Network of Markets and Firms |
| SD | 33.2009 | Luca Dall'Asta, Paolo Pin and Abolfazl Ramezanpour (lxxxv): Optimal Equilibria of the Best Shot Game |
| SD | 34.2009 | Edoardo Gallo (lxxxv): Small World Networks with Segregation Patterns and Brokers |
| SD | 35.2009 | Benjamin Golub and Matthew O. Jackson (lxxxv): How Homophily Affects Learning and Diffusion in Networks |

| | | |
|---|---|---|
| SD | 36.2009 | Markus Kinateder (lxxxv): Team Formation in a Network |
| SD | 37.2009 | Constanza Fosco and Friederike Mengel (lxxxv): Cooperation through Imitation and Exclusion in Networks |
| SD | 38.2009 | Berno Buechel and Tim Hellmann (lxxxv): Under-connected and Over-connected Networks |
| SD | 39.2009 | Alexey Kushnir (lxxxv): Matching Markets with Signals |
| SD | 40.2009 | Alessandro Tavoni (lxxxv): Incorporating Fairness Motives into the Impulse Balance Equilibrium and Quantal Response Equilibrium Concepts: An Application to 2x2 Games |
| SD | 41.2009 | Steven J. Brams and D. Marc Kilgour (lxxxv): Kingmakers and Leaders in Coalition Formation |
| SD | 42.2009 | Dotan Persitz (lxxxv): Power in the Heterogeneous Connections Model: The Emergence of Core-Periphery Networks |
| SD | 43.2009 | Fabio Eboli, Ramiro Parrado, Roberto Roson: Climate Change Feedback on Economic Growth: Explorations with a Dynamic General Equilibrium Mode |
| GC | 44.2009 | Fabio Sabatini: Does Social Capital Create Trust? Evidence from a Community of Entrepreneurs |
| SD | 45.2009 | ZhongXiang Zhang: Is it Fair to Treat China as a Christmas Tree to Hang Everybody's Complaints? Putting its Own Energy Saving into Perspective |
| SD | 46.2009 | Eftichios S. Sartzetakis, Anastasios Xepapadeas and Emmanuel Petrakis: The Role of Information Provision as a Policy Instrument to Supplement Environmental Taxes: Empowering Consumers to Choose Optimally |
| SD | 47.2009 | Jean-François Caulier, Ana Mauleon and Vincent Vannetelbosch: Contractually Stable Networks |
| GC | 48.2009 | Massimiliano Mazzanti, Susanna Mancinelli, Giovanni Ponti and Nora Piva: Education, Reputation or Network? Evidence from Italy on Migrant Workers Employability |
| SD | 49.2009 | William Brock and Anastasios Xepapadeas: General Pattern Formation in Recursive Dynamical Systems Models in Economics |
| SD | 50.2009 | Giovanni Marin and Massimiliano Mazzanti: Emissions Trends and Labour Productivity Dynamics Sector Analyses of De-coupling/Recoupling on a 1990-2005 Namea |
| SD | 51.2009 | Yoshio Kamijo and Ryo Kawasaki (lxxxv): Dynamics, Stability, and Foresight in the Shapley-Scarf Housing Market |
| IM | 52.2009 | Laura Poddi and Sergio Vergalli: Does Corporate Social Responsibility Affect the Performance of Firms? |
| SD | 53.2009 | Valentina Bosetti, Carlo Carraro and Massimo Tavoni: Climate Change Mitigation Strategies in Fast-Growing Countries: The Benefits of Early Action |
| GC | 54.2009 | Alireza Naghavi and Gianmarco I.P. Ottaviano: Firm Heterogeneity, Contract Enforcement, and the Industry Dynamics of Offshoring |
| IM | 55.2009 | Giacomo Calzolari and Carlo Scarpa: On Regulation and Competition: Pros and Cons of a Diversified Monopolist |
| SD | 56.2009 | Valentina Bosetti, Ruben Lubowski and Alexander Golub and Anil Markandya: Linking Reduced Deforestation and a Global Carbon Market: Impacts on Costs, Financial Flows, and Technological Innovation |
| IM | 57.2009 | Emmanuel Farhi and Jean Tirole: Collective Moral Hazard, Maturity Mismatch and Systemic Bailouts |
| SD | 58.2009 | Kelly C. de Bruin and Rob B. Dellink: How Harmful are Adaptation Restrictions |
| SD | 59.2009 | Rob Dellink, Michel den Elzen, Harry Aiking, Emmy Bergsma, Frans Berkhout, Thijs Dekker, Joyeeta Gupta: Sharing the Burden of Adaptation Financing: An Assessment of the Contributions of Countries |
| SD | 60.2009 | Stefania Tonin, Anna Alberini and Margherita Turvani: The Value of Reducing Cancer Risks at Contaminated Sites: Are More Heavily Exposed People Willing to Pay More? |
| SD | 61.2009 | Clara Costa Duarte, Maria A. Cunha-e-Sá and Renato Rosa: The Role of Forests as Carbon Sinks: Land-Use and Carbon Accounting |
| GC | 62.2009 | Carlo Altomonte and Gabor Békés: Trade Complexity and Productivity |
| GC | 63.2009 | Elena Bellini, Gianmarco I.P. Ottaviano, Dino Pinelli and Giovanni Prarolo: Cultural Diversity and Economic Performance: Evidence from European Regions |
| SD | 64.2009 | Valentina Bosetti, Carlo Carraro, Enrica De Cian, Romain Duval, Emanuele Massetti and Massimo Tavoni: The Incentives to Participate in, and the Stability of, International Climate Coalitions: A Game-theoretic Analysis Using the Witch Model |
| IM | 65.2009 | John Temple Lang: Article 82 EC – The Problems and The Solution |
| SD | 66.2009 | P. Dumas and S. Hallegatte: Think Again: Higher Elasticity of Substitution Increases Economic Resilience |
| SD | 67.2009 | Ruslana Rachel Palatnik and Roberto Roson: Climate Change Assessment and Agriculture in General Equilibrium Models: Alternative Modeling Strategies |
| SD | 68.2009 | Paulo A.L.D. Nunes, Helen Ding and Anil Markandya: The Economic Valuation of Marine Ecosystems |
| IM | 69.2009 | Andreas Madestam: Informal Finance: A Theory of Moneylenders |
| SD | 70.2009 | Efthymia Kyriakopoulou and Anastasios Xepapadeas: Environmental Policy, Spatial Spillovers and the Emergence of Economic Agglomerations |
| SD | 71.2009 | A. Markandya, S. Arnold, M. Cassinelli and T. Taylor: Coastal Zone Management in the Mediterranean: Legal and Economic Perspectives |
| GC | 72.2009 | Gianmarco I.P. Ottaviano and Giovanni Prarolo: Cultural Identity and Knowledge Creation in Cosmopolitan Cities |
| SD | 73.2009 | Erik Ansink: Self-enforcing Agreements on Water allocation |
| GC | 74.2009 | Mario A. Maggioni, Francesca Gambarotto and T. Erika Uberti: Mapping the Evolution of "Clusters": A Meta-analysis |

(lxxxv) This paper has been presented at the 14th Coalition Theory Network Workshop held in Maastricht, The Netherlands, on 23-24 January 2009 and organised by the Maastricht University CTN group (Department of Economics, http://www.feem-web.it/ctn/12d_maa.php).