Research Policy 41 (2012) 1205–1218

Contents lists available at SciVerse ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Knowledge flows – Analyzing the core literature of innovation, entrepreneurship and science and technology studies

Samyukta Bhupatiraju[a], Önder Nomaler[b], Giorgio Triulzi[a], Bart Verspagen[a,*]

[a] UNU-MERIT and Department of Economics, Maastricht University, PO Box 616, NL-6200MD Maastricht, The Netherlands
[b] Eindhoven Centre for Innovation Studies (ECIS), Eindhoven University of Technology, PO Box 513, NL-5600MB Eindhoven, The Netherlands

ARTICLE INFO

Article history:
Received 4 April 2011
Received in revised form 5 March 2012
Accepted 6 March 2012
Available online 14 April 2012

Keywords:
Innovation studies
Entrepreneursip
Science and technology studies
Citation networks
Network analysis

ABSTRACT

This paper applies network analysis to a citation database that combines the key references in the fields of Entrepreneurship (ENT), Innovation Studies (INN) and Science and Technology Studies (STS). We find that citations between the three fields are relatively scarce, as compared to citations within the fields. As a result of this tendency, a cluster analysis of the publications in the database yields a partition that is largely the same as the a priori division into the three fields. We take this as evidence that the three fields, although they share research topics and themes, have developed largely on their own and in relative isolation from one another. We also apply a so-called 'main path' analysis aimed at outlining the main research trajectories in the field. Here we find important differences between the fields. In STS, we find a cumulative trajectory that develops in a more or less linear fashion over time. In INN, we find a major shift of attention in the main trajectory, from macroeconomic issues to business-oriented research. ENT develops relatively late, and shows a trajectory that is still in its infancy.

1. Introduction

With the growing importance attached to knowledge, innovation and technological change in policy-making (e.g., Lundvall and Borrás, 2004), these topics have also become prominent in the social science literature. Fagerberg and Verspagen (2009) discussed these developments under the heading of "innovation studies", but their survey did not capture important parts of the social science literature that are pre-occupied with knowledge. In particular, the field of Science and Technology Studies (STS), which broadly deals with the social role and context of scientific research, was strongly underrepresented. In addition, the growing importance attached to entrepreneurship in the (economic) exploitation of knowledge (Braunerhjelm et al., 2010) did not appear very clearly in the Fagerberg and Verspagen approach.

This was one reason for the EXPLORE project, of which this contribution is part, to launch a detailed study of the three fields of Science and Technology Studies (STS), Innovation Studies (INN) and Entrepreneurship (ENT). These detailed studies were carried out as part of a larger bibliometric study, based on a new approach in which handbooks published for the three fields played a central role. The result of each of the three studies, as described in Martin et al. (2012), Fagerberg et al. (2012) and Landström et al. (2012) was the identification of a list of core contributions, consisting of about 125–150 references. This list was then used to provide a description of the way in which the literature emerged and an analysis of the main topics. One aim of the current paper is to perform a quantitative citation network analysis of the three lists of core contributions, and to compare the outcomes of this network analysis with the picture emerging from the three separate field studies.

An important part of the EXPLORE project was to investigate the relationships between the three fields. As they are all concerned with broadly related topics, namely the social (including economic) context and usage of knowledge, there are linkages between them. These linkages, for example, take the form of common origins (early inspirations), cross-citation, and overlap (in terms of scholars participating in more than a single field). One might even hypothesize that instead of three separate literatures, there is really only one (large) social science literature about knowledge and innovation, although the prior casual impression held by many participants in the EXPLORE project (based on their experience as practitioners in one or more of the three fields) was that this would probably not be the case. The second main aim of this paper is therefore to investigate how the three fields are related to each other. How important are their common roots? When did they start to diverge from each other, if at all? And what is the nature of the interactions between the fields once they started to be distinguishable from each other? We will again use quantitative citation network analysis to investigate these questions.

* Corresponding author.
E-mail address: b.verspagen@maastrichtuniversity.nl (B. Verspagen).

0048-7333/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.respol.2012.03.011

As Fagerberg and Verspagen (2009) argue, the definition of a scholarly discipline covers more than just the development of a literature. It also includes the establishment of specialized journals and conferences, as well as professional associations and the like. The notion of a scholarly discipline covers a web of "institutions", including shared norms and values, which are both formal and informal in nature. Our bibliometric approach cannot capture this broad base of the phenomena involved in defining disciplines, which is why we consistently use the term "fields" rather than "disciplines". Our research questions, which are elaborated in the next section, are aimed at these three fields.

Besides elaborating our research questions, Section 2 also provides a brief introduction to our methodology. In short, our methodology relies on the use of citation data, which we collect for the core contributions in the three field studies, and consists of a range of network-based techniques, including network visualization and clustering. Section 3 introduces our data, both in terms of the nature of the data used, and in terms of some brief descriptive statistics. Sections 4 and 5 elaborate on the main methodologies used: Section 4 focuses on network visualization and clustering of the network nodes (i.e., references on the lists of core literature), while Section 5 identifies the so-called 'main paths' in the database. Section 6 summarises the main conclusions of our work.

2. Research questions and introduction to the methodology

One basic aim of this paper is to explore the interactions between the three literatures that have been investigated in the EXPLORE project: science and technology studies (STS), innovation studies (INN), and entrepreneurship (ENT). We want to find out to what extent, and how, these literatures have influenced each other. Have they developed largely in isolation of each other, or has there been cross-fertilization? If interactions between the three literatures are found, are they unidirectional (one field influencing another but not vice versa), or is the interaction more "complete", with all potential linkages being present?

Although these questions could be explored by a variety of methodologies, for example a detailed survey of the content of the three fields, or an analysis of personal interactions of the main players in the field, we opted to work with one data source, which is citations between the three fields. The reason for focusing on citations is that they are indications of intellectual influence (from the cited paper to the citing paper), and therefore can be used as "paper trails" of the flow of ideas between and within the three fields. Although citation analysis is now accepted by many as a valid analytical tool, one must be careful about the limitations of the data. Citations may depend not just on the actual flow of ideas but also on specific habits and norms, which are potentially different between fields. For example, in some fields, it may be customary to have very long reference lists, while other fields have far fewer. In addition, citation patterns may be influenced by "strategic" motivations; for example, a particular paper or author may be cited (or not cited) to enhance the probability of publication, or citations may be influenced by personal friendships or dislikes. Despite all this, we will still rely on these data to analyze our research questions. Partly, this is because in our case some of the problems seem to be relatively minor (e.g., our impression from collecting the data is that the length of reference lists does not vary much between the three fields that we investigate), and partly it is because our study is the first (at least as far as we are aware) that addresses the interaction between these three fields. Thus, even if a citation analysis cannot shed full light on the topic, it should nevertheless provide a major addition to what we know.

The citation data that we use come in the form of a matrix, with citing and cited documents in the rows and columns. We employ network analysis to analyze this matrix. Documents (i.e., books, journal articles or book chapters) are the nodes of the network, and the citations between them are the linkages. Several network methods will be used. To the extent that our research questions are related to a general, impressionistic view of the three fields and their interactions, we will mainly rely on two closely related network-based methodologies: clustering (or classification) and graph visualization. The primary purpose of both these methodologies is to detect meaningful groups of documents, in which the documents within a group are more similar to each other than to those in other groups. Clustering achieves this by applying an algorithm that attempts to maximize within-group similarity and between-group dissimilarity, while visualization techniques map the documents into a low-dimensional (usually two or three) coordinate system in such a way such that more similar documents show up closer to each other. The observer then interprets the clustering tendencies of the documents by visual inspection. In other words, while clustering techniques classify documents into comparatively 'crisp sets', visualization techniques display clusters (if there are any) as rather fuzzy sets and leave decisions on the boundaries between the 'emergent' groups to the subjective judgment of the observer (i.e. the clusters emerge in the eye of the beholder).

Neither the visualization nor the classification methodology is bound by universally accepted unique criteria. All existing techniques are essentially heuristics. This is why we prefer to use both classification and visualization methods, so that we can then compare the outcomes. Within each of the two methodologies, a wide range of heuristics is available, leading to different methods. We opt for a combination of visualization and clustering methods that are rooted in the same technical principle, based on the spectrum (eigenvalue decomposition) of the (transformed) network matrix. These techniques are explained below. In addition, for the clustering method employed, an important aspect of our method is that we choose not to fix the number of clusters in advance. While many clustering methods require the researcher to do so, this is not a desirable solution from our point of view. The reason is that one subpart of our broad research question asks whether the three fields are indeed separate, or (as an extreme case) can also be considered as a single, broad field. Obviously, the choice of a particular number of clusters (e.g., three or one) would predetermine the answer to this question. This is why we want to be able somehow to determine the number of clusters in an endogenous way, and to assess the "goodness-of-fit" of a particular cluster solution.

The majority of the network methods used in the literature have been developed for networks where the links that connect nodes do not represent a direction. For example, relations such as friendship, kinship, or co-occurrence are usually characterized by a link that goes in both directions between the nodes. In contrast, knowledge flows as indicated by citations have a strongly directed nature. In addition, citations have a temporal structure, because the cited paper must have appeared before the citing paper in order to be cited. Hence, citation networks are not only directional but acyclic (i.e., if one starts at Document A and follows the citation links, one never arrives back at Document A).[1] The acyclic and directed nature of the citation network represents valuable information for network analysis, especially if the analysis aims at a historical narrative. Hence we would prefer to use methods that take these characteristics of the network into account.

To our knowledge, two main analysis methods have been proposed that can be used to investigate the historical (acyclic) nature of networks. For the clustering part of our analysis, we draw upon

[1] The acyclic nature of a citation network can be disturbed when pre-prints are available before the official publication date. This will be discussed in more detail below.

Shibata et al. (2008), where the inter-temporal dimension of the network is captured by first analyzing a network as it existed in period T, and then gradually extending the coverage to include periods $T+1$, $T+2$, etc. In this way, clusters of network nodes (documents, in our case) emerge for each historical stage of the network, and an historical impression emerges from this in a natural way. This is how our cluster analysis will be carried out below.

The other "history-friendly" network analysis methods has its origins in the work by Hummon and Doreian (1988), and was later extended by Verspagen (2007). This method is not aimed at visualization or clustering, and is therefore rather different from the methods that have been discussed so far. The aim here is to detect the so-called 'main paths' in the citation network. These main paths are those paths that represent the largest part of the knowledge flows within the network. Hence, this method relies, to a larger extent than the clustering and visualization methods, on the indirect linkages between documents (if document A cites document B and B cites C, an indirect knowledge flow goes from C to A). All citation links in the network represent knowledge flows, but the Hummon and Doreian method aims to identify the ones that (through upstream and downstream citation linkages) represent particularly strong flows. By linking such "important" linkages, a citation-linked chain of documents emerges that can be interpreted as the main path or trajectory of the literature.

In our analysis, this method serves a number of important functions. First, it enables us to assess which are the main trajectories in the large citation network, and hence whether the documents that are found on these trajectories correspond to the leading contributions as identified in the individual field studies. In this way, we hope to be able to validate both the individual studies and our citation-based analysis. This is the first major aim of our study. Second, we can investigate how the interactions between the fields that emerge from the network visualization and clustering methods are related to individual contributions (documents).

3. Data description

The starting point of our data collection are the lists of core contributions that emerged from the individual studies for the three fields in the EXPLORE project (Martin et al., 2012; Fagerberg et al., 2012; Landström et al., 2012).[2] These lists were compiled by analysing the reference lists in handbooks covering the respective fields, and are taken to represent the main contributions to the literature in each field.

Each of the lists has less than 160 publications on it (see Fig. 1 below), and hence it is a relatively small sample from the literatures that are each by themselves now fairly large (we do not have precise figures on the size, but we estimate each of the three fields to be composed of over 10,000 publications). The handbook methodology used in the three case studies focuses on the publications that are considered core by the expert authors that survey the literature in the handbooks. It may be the case that publications that lie somewhat "in between" the three fields are considered as rather "peripheral" by the handbook authors, and, if so, they would receive relatively few citations in the handbooks. The result would be that such "intermediate" publications between the three fields do not

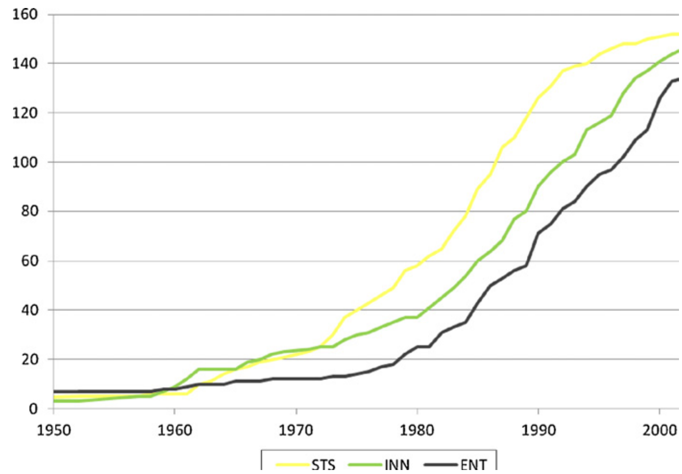

Fig. 1. Cumulative number of publications in the database, by field.

show up in our database, and hence we would underestimate the true level of interaction between the fields. This would be worth investigating in future research, but it cannot be addressed in the current paper.

For all publications on the three core lists, we collected citation information, i.e., we constructed a matrix in which the rows and columns are the publications on the three core literature lists, and a cell value is one if the row publication is cited by the column publication. We clean this matrix to eliminate cycles[3] and delete from the database all publications that are not cited and do not cite, and after this we have 412 publications: 152 in science and technology studies (STS), 146 in innovation studies (INN), and 134 in entrepreneurship (ENT). The sum is larger than 412 because some publications are found in more than one list.

Fig. 1 shows the cumulative number of publications per field, starting with 1950. Its fairly even distribution of the number of publications between the fields is easily observed, as is the somewhat uneven development over time. INN and STS both have a small take-off in the number of publications around 1960, while ENT stays fairly flat until the early 1980s. STS takes a decisive lead in the number of publications around 1960, but levels off more strongly than the two other fields at the end of the period (this may be due to sampling issues – see Martin et al., 2012).

There are 3720 citations between the 412 publications, i.e., an average of about 9 per cited or citing document. These citations are heavily concentrated within the fields, rather than between the fields, as Fig. 2 shows. The number of citations within STS and INN are about equal, with each of these fields accounting for nearly one third of total citations. Within-ENT citations are only just over half those within the two other fields. The off-diagonal values (citations between the fields) account for about one third of all citations in the database, and these are heavily dominated by the publications that are assigned to more than one field (excluding those publications, the number of off-diagonal citations falls from 1327 to 367, which is a 72% drop, while the sum of the diagonal values would fall from 3356 to 2591, a 23% drop). In the off-diagonal cells of Fig. 2, the INN field contributes relatively higher values,

[2] The list from Fagerberg, Fosaas and Sapprasert that we use is an older version of the list that they present in their paper in this special issue. They completed their most recent list after we had already finished the citation data collection process. The full list of references that we use can be found as an appendix in their early working paper version at http://ideas.repec.org/p/tik/inowpp/20100616.html. The list of references for ENT can be found in the working paper version of Landström, Harirchi and Åström at http://ideas.repec.org/p/tik/inowpp/20111005.html, and the list for STS can be found in the working paper version of Martin, Nightingale and Yegros-Yegros at http://ideas.repec.org/p/tik/inowpp/20111004.html.

[3] An example of a cycle is where paper A cited paper B, paper B cites paper C, and paper C cites paper A, or, the simplest possible cycle, paper X cites paper Y, and paper Y cites paper X. If papers had unequivocal publication dates (i.e. a single date before which they could not be cited), such cycles would not be possible. (However, the existence of pre-publication versions may mean that a paper can be cited before it is formally cited.) We want to reduce the citation matrix to such an "ideal" state, so that we will be able to apply network analysis methods for non-cyclical networks.

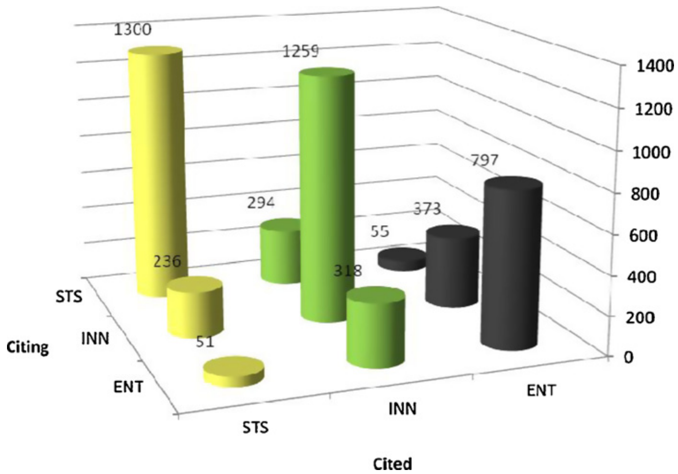

Fig. 2. Number of citations between and within fields.

i.e., this field exhibits the strongest interactions with the two other fields.

4. Clustering the network

The evidence about inter- and intra-field citation suggests that there is indeed a "clustering" in the set of all 412 documents, in the sense that particular subsets of the complete sample of documents interact more closely with each other (in terms of citation) that with other subsets. But this preliminary conclusion arises from a pre-determined grouping of the total sample into the three fields STS, INN and ENT. We now proceed to cluster the nodes (documents) in the citation network in a more open-ended way, so that we can compare the results of that exercise with the pre-determined segmentation into three fields. The degree to which the results from the clustering exercise match the pre-determined segmentation will tell us something about how "integrated" the three fields are: a close (or weak) match will point to a crisp (or fuzzy) boundary between the three fields.

At an intuitive level, the problem of how to cluster the nodes in our citation network has close similarities to the problem of how to represent the network graphically in a low-dimensional space. Nodes of the network (documents) that are in the same cluster should appear close to each other in the graphical representation. We therefore start by graphing the network, and from the method that we employ to do this, we then proceed to cluster the network.

Because citation networks have a clear arrow of time, it makes sense to use the publication year as one dimension in a graphical representation of the citation network. In order to keep the graphical representation as simple as possible, we use only one additional dimension, i.e., we will represent the citation network in 2-dimensional space. For the additional dimension, the guiding principle is that two papers that are connected by a citation should be closer to each other than two papers that do not have a citation link.

The problem of 'graph embedding' is to translate this guiding principle into an actual method for calculating coordinates. Imagine for a moment that the positions of all nodes[4] (along the 1-dimensional space) except for a focal node j, are fixed. Then, a good choice for the coordinate of node j would the average of the nodes that it is connected to. However, when applied to all nodes, we will in general be unable to find a configuration in which the "average-condition" can be satisfied for all the nodes simultaneously. The technical problem that we are faced with is therefore to satisfy the simultaneous condition "as well as possible."

This is an old problem, dating back to Kirchhoff (1847), which has been extensively studied since then. Hall (1970), and the literature that comes thereafter, formulated the graph embedding problem as a minimization problem:

$$\text{Min}\, z = \sum_{i,j} a_{ij}(y_i - y_j)^2$$

where y_i and y_j are the coordinates to be assigned to node i and node j, respectively, while a_{ij} are the elements of the adjacency matrix,[5] or some specific transformation of the matrix (which will be discussed later). Thus, the quantity z, which is to be minimized, is equal to the sum of squared distances between all nodes that have a (direct) citation link. This minimization problem, which translates our general principle into a workable method, underlies our methodology in this section and the next one. In what follows, we briefly describe some of the intuitions behind this methodology, but, because of limited space, we leave the reader to explore the original references for details.

Following Hall (1970), we rewrite z in matrix notation, as

$$z = y^r L y$$

Here, $L = D - A$ is the so-called Laplacian matrix, with A as the adjacency matrix, and D a diagonal matrix that has the row sums of A as elements (thus, the elements of D are the degree centrality of the nodes). We also realize that our minimization problem has one trivial solution that is highly undesirable, and which is to assign every node (or at least those that are connected into a single component) the very same coordinate. In order to avoid this trivial solution, the constraint $y^r y = 1$ can be imposed, which makes sure that the actual solution to the minimization problem shows a positive variance of the coordinates. In this way, the minimization problem becomes a constrained minimization problem, which can be solved by the Lagrangian method to yield the characteristic (eigenvalue) equation of the form: $Ly = \lambda y$. If this solution equation is pre-multiplied by the vector y^r, we have $y^r L y = \lambda y^r y$, which implies $z = \lambda$ since $z = y^r L y$ and $y^r y = 1$. Thus, the eigenvector associated with the smallest non-zero eigenvalue[6] minimizes z and provides the optimal coordinate vector.

In cases where the degree distribution is not close to uniform, the coordinates obtained from the spectra of the Laplacian matrix (as explained above) tend to lead to a collapse of the majority of nodes into a very narrow area, with the rest of the map occupied by a few sporadically distributed nodes. The well-accepted means to overcome to this problem is to use the spectrum of the degree-normalized Laplacian, which is $L_N = D^{-1} L$ (i.e., each element of the Laplacian matrix is divided by the degree of the node). This is the method that we use to map the citation network of the three fields.

The results of the visualization procedure are displayed in Fig. 3. Along the horizontal dimension, we see a gradual increase in the node-density of the graph, indicating the growing size of the three fields. In the vertical dimension, we see a clear separation of colors/shades. Because the colors/shades of the nodes indicate the field to which they belong (publications belonging to more than one field are diamond-shaped), this means that the publications belonging to a single field appear close together on the vertical dimension, which indicates closeness based on citations. This corresponds to

[4] We refer to the papers in the database as 'nodes' of the network from now on.

[5] Which, in our case, is the symmetrized citation matrix, which is max(c_{ij},c_{ji}), with c_{ij} equal to 1 if j cites i and 0 otherwise.

[6] The smallest eigenvalue is 0, where the associated eigenvector corresponds to the trivial solution in which all nodes are assigned exactly to the same coordinate.

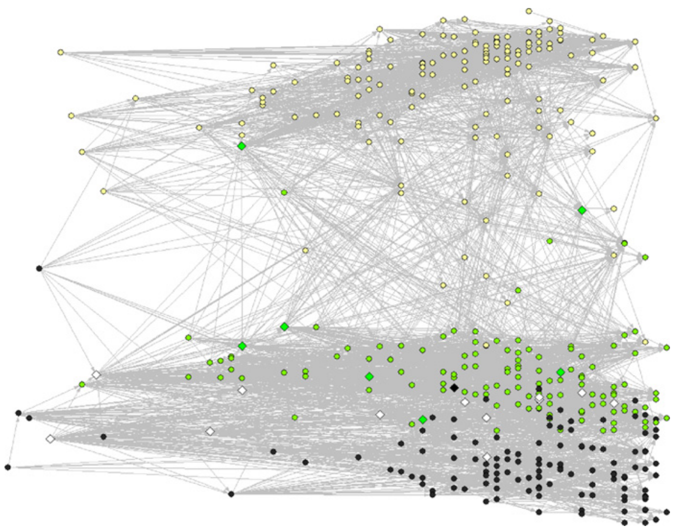

Fig. 3. Visualization of the citation network. The horizontal axis represents time (early years on the left), the vertical dimension represents closeness based on citation patterns; colors/shades of the nodes indicate field (yellow/light for STS, green/intermediate for INN, black/dark for ENT); diamond-shaped (white) nodes are documents that are assigned to more than a single field. The horizontal time-scale is ordinal, i.e., newer nodes are placed further to the right, but horizontal distances do not reflect precise differences in years. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

the expectation that may stem from the relative unimportance of between-field citations (Fig. 2).

Moreover, we see that the INN field is positioned in the middle, corresponding to its higher tendency to be involved in between-field citations than the other two fields. INN and ENT are also generally closer than INN and STS, and there is actually a fair amount of overlap between INN and ENT on the vertical axis. The overlap between INN and STS is much weaker, represented by only a handful of documents, including some that are classified into multiple fields (diamond shaped).

The impression that this gives is that the complete citation network indeed consists of subfields, which largely correspond to the three fields under study. In other words, it makes sense to talk about three fields, rather than a single large field that is concerned with knowledge and its societal and economic relevance. The pattern of knowledge flows, as indicated by citations, suggests that this broad area has been developing largely in the form of a separate entities, although this is a more adequate description of the interactions between STS and the two other fields, than between INN and ENT. And even between STS and the rest of the dataset, some linkages exist. We will now put this visual impression to the test in a more rigorous clustering analysis.

Before the introduction of the 'modularity' concept by Newman and Grivan (2004), clustering methods had traditionally required the user to pre-determine the number of clusters generated by the algorithm. Newman and Grivan (2004) then provided a meaningful criterion to judge the "goodness of fit" of the clustering partition reached. In their interpretation (which we follow), a good cluster partition is one in which the elements of each cluster are better connected to each other than to elements of other clusters. Newman and Girvan benchmark the connectivity within the clusters against that of a random network of equal size, requiring that a good clustering result yields higher-than-random connectivity within the clusters. Accordingly, the "modularity" can be defined as the number of links (citations) falling within clusters (of documents), minus the expected number in an equivalent network with links placed at random. For a partition with R clusters, the modularity index is defined as

$$Q = \frac{1}{2M}\sum_{ij}^{N}\sum_{r}^{R}\left(A_{ij} - \frac{k_i k_j}{M}\right) S_{ir}S_{jr}$$

where k_i (k_j) is the degree of node i (j), M is the number of links in the network, N the number of nodes, A_{ij} is the associated element of the adjacency matrix, and S_{ir} (S_{jr}) a binary variable which (given the partition) is 1 if node i (j) belongs to cluster r and 0 otherwise. Modularity Q can be either positive or negative, but only positive values indicate the possible presence of a "community structure" (i.e., linkages are stronger within clusters than between them). Thus, the aim of our cluster analysis will be to find a cluster partition that maximizes Q. Finding this optimal partition is essentially an integer programming problem, but, unfortunately, finding the global maximum for Q is impossible within a reasonable computing time. Hence, following the literature, we will resort to a heuristic method to locally maximize Q.

The chosen heuristic operates on the same spectrum (eigenvector) as the visualization method. This follows White and Smyth (2005), who show that the eigenvectors corresponding to the largest (non-trivial) eigenvalue of the row-normalized adjacency matrix (i.e., following our earlier notation, $A_N = D^{-1}A$) provides a good basis for a modularity-maximizing clustering algorithm. White and Smyth (2005) show that this algorithm can produce better clustering (i.e., partitions with higher modularity scores) than the one suggested by Newman (2004).

In addition to the acclaimed advantages of the White and Smyth (2005) algorithm in terms of local modularity maximization, our choice also draws on the close proximity of the spectral (eigenvector) basis of this algorithm to that of our visualization procedure. It has been shown that the spectrum (i.e., eigenvalues and eigenvectors) of the row-normalized adjacency matrix is identical to the spectrum of the Laplacian matrix as was used in the visualization procedure in the previous section. However, whereas in the previous section we wanted to minimize the quantity z, now we are interested in maximizing Q. In other words, the problems of clustering and visualization are dual to each other, and while we previously wanted to look at the smallest non-zero eigenvalue, we now want to look at the largest non-trivial eigenvectors of the row-normalized adjacency matrix.[7]

Obviously, the more variables we use to cluster the nodes of the network, the more information we have about their similarity or dissimilarity. This is the main degree of freedom that we have to (locally) maximize modularity. In our case, the variables are eigenvectors of the row-normalized adjacency matrix, which we rank from largest to smallest. Our heuristic is to use the first $F-1$ non-trivial leading eigenvectors (we set $F=50$), and for each $f=1,2,\ldots,F-1$, and then to use a k-means clustering algorithm to compute the partition that gives $f+1$ clusters on the basis of the first f non-trivial eigenvectors. For each of the values of f, we have a cluster partition, for which we calculate the modularity index Q. We then pick the partition (i.e., f) where the modularity turns out to be greatest.

So far, we have not paid any attention to any methodological aspects that are related to the temporal structure of the citation network. We draw on Shibata et al. (2008) to develop this further. The basic novelty that they add is to look at subparts of the complete network, corresponding to periods in time. Specifically, we start by carrying out a clustering analysis for the network that consists of

[7] It can be shown (e.g., Hall, 1970) that the eigenvector corresponding to the largest eigenvalue is trivial, since all elements in it are equal to each other, and hence this eigenvector provides no information for the clustering procedure.

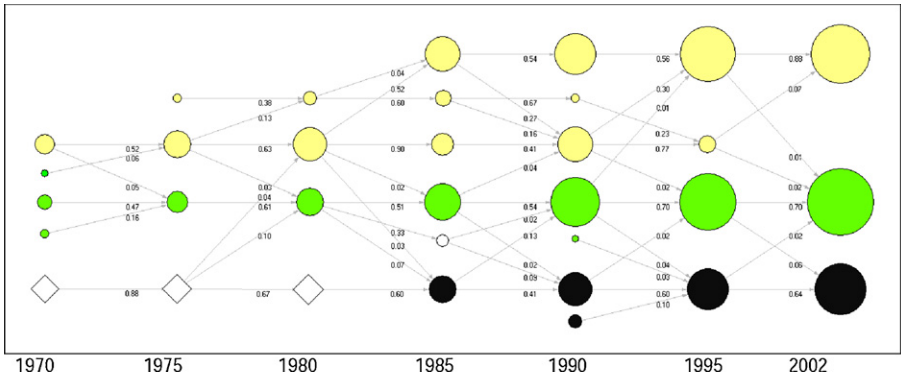

Fig. 4. Cluster–time profile. The circles represent clusters of documents; the vertical dimension is loosely derived from Fig. 3, with the circle size indicating the number of documents in the cluster, and the colors/shades indicating the main content of the clusters in terms of field (as in Fig. 3). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

all documents that appear in 1970 or before. This represents the citation network as it existed at the end of 1970. Then we proceed to carry out the same analysis, but for the period up to and including 1975, i.e., five years of additional data. By comparing the two sets of results, we are able to assess the development of the network over these five years. We repeat this in steps of five years, i.e., for 1980, 1985, 1990, 1995 and 2002 (note that 2002 is the last year in our database, and we prefer this seven-year step instead of also including the intermediate year 2000).

This procedure yields a cluster partition (that maximizes modularity) for each of the years 1970, 1975, etc. For each cluster *j* at time *t*, we compute the number of nodes (documents) that are common with each cluster *i* at time $t-1$, and express this number as a share of the number of documents in cluster *j* at time *t*.

Fig. 4 displays these results. On the left, the clustering partition in 1970 is displayed. We find five clusters, two of which are relatively small (with just two and three members). In terms of cluster content, two of the three large ones broadly correspond to STS and INN. The STS-cluster has 17 members, of which 15 are solely in the STS field, one is jointly classified as STS/INN, and one is in INN. The INN-cluster has nine members, of which seven are solely INN, and the other two are joint INN/STS and INN/ENT. The other large cluster (diamond shaped) contains a mixture of INN and ENT documents (of the 15 members, eight are classified as ENT, three as INN, three as INN/ENT and one as INN/STS). This suggests that in 1970, the entrepreneurship field did not yet exist very distinctly (this is broadly in line with the conclusions in Landström et al., 2012), while the two other fields, INN and STS, were already rather more established. In addition, the intellectual roots of ENT that did exist at that time were more closely related to INN, indicating that from the early stages it has been closely associated with INN.

For the next two time observations, in 1975 and 1980, this situation remains largely unchanged. We have two large and more or less distinct INN and STS clusters, and one mixed INN–ENT cluster. These clusters are continuations of the previously existing ones, as indicated by the large fraction of documents that stem from the corresponding cluster in the previous year.[8] In this early phase, some movement between clusters of different colors/shades does take place, but this is limited to relatively small flows (at most 10%) between the white, diamond-shaped nodes and the INN clusters.

In 1985, we see, for the first time, a distinct ENT cluster emerging. In terms of our analysis, this represents the birth of ENT as a separate field. This cluster has 30 members, of which 25 are classified as solely ENT, while just four are joint INN/ENT and one is INN. The majority of documents in this cluster (60%) come from the previously mixed INN/ENT cluster. From this point onwards, the network remains separated into three large clusters, together with a number of smaller ones, and each of those clusters is clearly associated with one of the three fields. Thus, along with the birth of ENT in 1985, the separation into three distinct fields is also now more or less complete.

The final partitioning into three clusters mimics the original classification into the three fields very closely. In this respect, the 19 documents that belong to more than a single field are a somewhat special case. All of these are classified into the "correct" field, i.e., one of the two (or three) that they belong to. But in 17 out of 19 cases, this is the INN field. Thus, in the large majority of cases where the "original" classification does not unequivocally point to a single field, the clustering algorithm assigns the document to the INN field.

Of the documents that belong only to a single list of core literature, 29 (corresponding to 7.4%) are classified in a different field than the one they belong to (according to our database). Here again, INN is the largest attractor of such "cross-classifications": 28 of the 29 are assigned to INN, when they actually belong to a different field (18 to ENT and 10 to STS). In a number of these cases, good arguments can be made that the document in question actually (also) belongs to INN, although it did not appear on the list of core INN literature. Thus, overall, the clustering results seem to provide convincing evidence that, according to the citation patterns, a division of the large literature that our database covers into STS, INN and ENT makes a lot of sense. The fields can truly be characterized as essentially separate fields.

5. Development of the literatures and the role of individual documents

5.1. Methodology

In this section, we turn our attention to the role of individual documents (or nodes) in the citation network. Our aim is two-fold. First, we want to obtain an impression of the main developments of the literatures in the three fields, and to contrast this impression with that from the individual field studies that were the source of our core literature lists. Second, we want to investigate how the main developments in the three separate fields interacted.

Our methodology is based on Hummon and Doreian (1988) and Verspagen (2007). It is aimed at finding a set of so-called main paths in the citation network. A main path is a chain of citation-linked documents representing a major stream of ideas and insights that flows through the network. The intuitive interpretation of a main path is that it summarizes the main ideas (or more precisely, the

[8] Note that the sum of incoming links is smaller than unity because new documents are continuously added to the network.

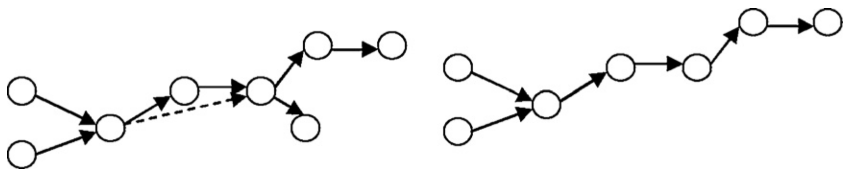

Diagram 1. Complete network (left) and main path (right) in a Hummon–Doreian calculation.

documents in which these ideas are expressed) in the literature under analysis, as well as their historical relationships (i.e., how these main ideas build upon each other). Note that our analysis does not aim to identify a single main path (as Hummon and Doreian did[9]), but instead takes every single "start point" (a document that cites no other documents in the database, but is cited itself) as a potential starting point of a main path.

By adopting such a "multiple" main path perspective, we hope to make sure that the summary view that we provide is not too crude (as might be the case with a single main path). We will see that in many cases, the multiple main paths that we identify converge on each other, while in other cases they do not. This provides valuable information of how the many developments that take place within a literature at any given point in time, are related to each other when this literature develops over time. Therefore, the convergence–divergence dynamics of main paths will be a main topic in the discussion of our results.

The analysis starts by assigning weights to each citation in the network. In raw form, the citation matrix only consists of zeros and ones. The Hummon and Doreian method assigns weights on the basis of how many upstream and downstream documents are connected by the citation link. Diagram 1 illustrates the calculation.

The citation weight is equal to the number of downstream nodes multiplied by the number of upstream nodes.[10] For example, in the left part of Diagram 1, the dotted arrow has three downstream links, and four upstream links, which yields a weight of 12. Note that this way of calculating the weights tends to favor links that lie in the middle of the network. For example, the rightmost link in Diagram 1 has only one upstream node, and seven downstream links, yielding a weight of 7. The two leftmost links have a weight of 6. The method also penalizes direct links between nodes that are far apart on the horizontal dimension (i.e. time, in our case). Compare, for example, the weight of 12 of the dotted link with the weight of the two links directly above it. The left one of those has a weight equal to 15, and the right one has a weight of 16 (this is also the maximum value found in the diagram). Despite the fact that the dotted link serves a similar function in the network to the two links directly above it, its weight is less than the average of these two links on the "upper path". Thus, the calculation method tends to favor small, incremental links, rather than long steps that stretch far into the past. In terms of the outcomes of the analysis, we can thus expect to obtain a map of the direct, historical dependencies of the documents in the network, rather than some ex post interpretation of the linkages.

The next step in the main path method is to find paths in the network that locally maximize the sum of the weights. This is done by starting at a given "start point" of the network, and then following the outward link with the highest weight. At the other side of this link (the citing document), the process if repeated, and again the path with the highest outward weight is chosen. This is repeated until an "end point" is reached (an "end point" is a point with no outward connections; in Diagram 1 we have two "end points").

The right-hand part of Diagram 1 shows the main paths that are found in the complete network.[11] In this case, these main paths cover a large part of the complete network (only two links and one node are dropped), and hence it may appear as if the proposed method does not actually reduce the complexity of the full network by much. However, this is not generally the case in analysis on real datasets, where the networks are much larger and more complex. In those empirical cases (including ours, as will be seen below), the set of main paths provides a very significant reduction of the network.

5.2. Results

Following Verspagen (2007), we perform the main path analysis in a similar way to that in which the clustering exercise was done, i.e., we start with a main path search for the period up to and including 1970, and then repeat this for the period up 1980, 1990 and 2000. (We use 10-year periods here rather than 5-year periods simply in order to save space.) The network of main paths for the period up to 1970 is displayed in Fig. 5.[12]

We observe the presence of one large and two smaller weakly connected components.[13] The main component is very heterogeneous, comprising publications from all three fields of STS, INN and ENT, as well as hybrid ones classified into more than one field. However, we may identify two main thematic groups in the main component. One group, at the bottom of the figure, is a mixed INN/ENT path, mainly composed by the three seminal books of Schumpeter, drawing on Cantillon (1755) and Marshall (1890), and continuing in turn to Rogers (1962) and Kotler (1976). This chain of publications forms what might be labeled as a *proto-INN* group (and perhaps also a *proto-ENT* group). The other element in the main structure is clearly an STS-group. The connection between the two groups is only due to Kuhn's (1962) "The structure of scientific revolutions", which is a key publication that influenced several academic fields (indeed Kuhn's book has the highest ISI citation index of our whole dataset, with an average of more than 400 citations received each year).

The STS-group in the main component consists of several paths that, as of 1970, all converge on Rose and Rose (1969). One cluster of paths originates from the top of the figure with a group of references cited by Barber (1952), namely Weber (1930), Merton (1938), Bernal (1939) and Bush (1945). This is consistent with the

[9] Hummon and Doreian's (1988) database is much smaller, and they identify, on the basis of their detailed knowledge of the field, a single document as the starting point of the main path. See Verspagen (2007) for a more detailed discussion of this issue, as well as how a "top main path" could be identified among the multiple main paths that we find.

[10] Hummon and Doreian (1988) propose several ways of calculating the weights. Here we use their SPNP measure.

[11] Note that in Diagram 1, the right hand side represents two main paths, one emanating from each "start point", although these main paths overlap after the second node. Note also that, in line with the above discussion, the dotted link from the left hand side is not on a main path, because it has a lower weight.

[12] At first sight, it would seem paradoxical that the network of main paths in 1970 contains one reference with a publication date later than 1970 (Kotler, 1976), but this is because there were several editions of this textbook, which have been unified here under a single date.

[13] The term "weakly connected component" refers to a group of nodes that appear connected, but in which not all nodes are reachable from other nodes. For example, in Fig. 5, Kotler (1976) is an "end point", but it cannot be reached from, e.g., Merton (1938), which is a "start point", even if the two appear as indirectly connected. Of course the direction of the links is responsible for the non-reachability.

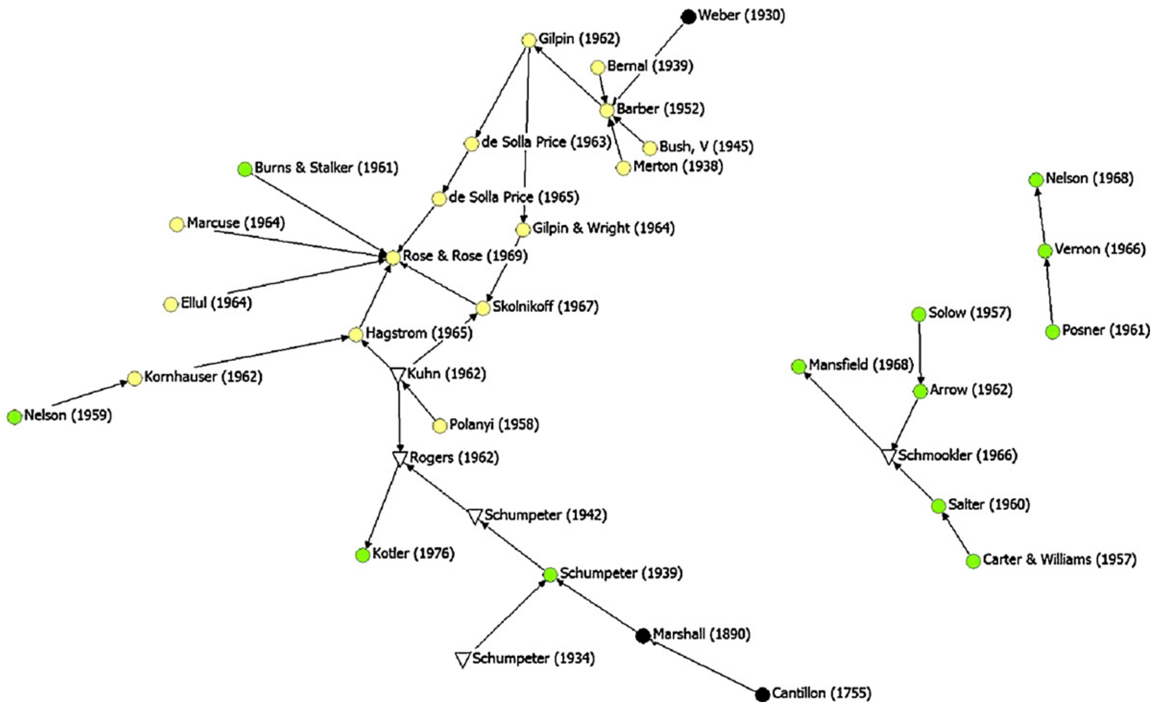

Fig. 5. Network of main paths for the entire network up to and including 1970. Colors/shades indicate the field that a paper belongs to. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

observation by Martin et al. (2012) that Merton (1938) and Bernal (1939), together with Fleck (1935),[14] helped to provide the foundations for the emerging field of STS. In particular Bernal's 1939 "The social function of science" pioneered the idea of looking at the social function of (natural) science from a social science perspective, while Merton's 1938 book looked at the historical and sociological development of science and technology in England in the 17th Century. Hence, as Martin et al. (2012) suggest, it would seem that Barber (1952) did indeed begin "*to lay the groundwork for the integration of the sociology and history of science*".

A few years after Barber, Kuhn's seminal book (1962), drawing on Polanyi (1958), integrates the philosophy of science into the discourse. This is another part of the STS component of the main cluster, showing that the three previously rather separate streams of research on the history, philosophy and sociology of science were successfully brought together (Martin et al., 2012). The Polanyi–Kuhn path then branches in two directions, both of which later converge on Rose and Rose (1969). One of these corresponds to the sociological perspective of Hagstrom (1965), which integrates the philosophy of science of Polanyi and Kuhn with works by Nelson (1959) and Kornhauser (1962), a branch that analyzes the interplay between science, economics and industry. The other involves Skolnikoff (1967), a publication which brings together the Kuhn-branch with the path that starts from the Barber-centered stream of work. In our interpretation, this path, which goes via Gilpin (1962) and Gilpin and Wright (1964), represents a distinct non-Mertonian approach to STS, which focused on the relationships between science and policy-making, strongly criticizing the linkages between science (in particular physics) and defense-oriented (and funded) projects (Martin et al., 2012). Finally, there is a path from Gilpin (1962) to Rose and Rose (1969) going through two publications by De Solla Price (1963,1965), reflecting the quantitative approach to STS that the latter pioneered.

[14] Fleck (1935) shows up in the 1990 main paths.

All the various paths within the 'emergent STS cluster' eventually converge on Rose and Rose (1969). This is the last publication in the figure of the main paths as of 1970, a book that synthesized the principal contributions and intellectual movements which characterized STS in its early years. In this, the authors summarize the criticisms to the traditional view of science as a process of unveiling the deterministic laws of the natural world, so it is perhaps little surprise that, as we shall see in the 1980 figure, later trajectories of STS were to draw heavily on this influential publication.

Apart from the above mentioned publications, all of which were integrated into the main component of the network, we find two smaller, disconnected components, both exclusively made up of publications from INN. Both these components represent the early emergence of the economics of technological change in the U.S., which would later become a part of the larger INN field. The larger component, which converges on Mansfield (1968), deals with issues of economic growth, technology and productivity. The other component, converging on Nelson (1968), deals with issues of international trade and innovation across countries.

From the 1970 figure, we can conclude that at this time, INN and ENT were far from being developed fields of analysis. INN was developing in an embryonic form as the economics of technological change, but ENT only existed in terms of various publications that would later come to provide the foundations of the field. These findings are in accordance with the detailed descriptions of INN by Fagerberg et al. (2012), and of ENT by Landström et al. (2012). STS was clearly somewhat more developed, in that we see an interdisciplinary STS 'proto-paradigm' emerging, which, according to Martin et al. (2012), and in line with the main paths in Fig. 5, had a "*distinctive emphasis on unmasking the external (*i.e. *extra-scientific) social factors behind the processes of science but also the content of science*". In addition, at this stage, STS seems to have provided the 'glue' to keep the citation network together.

The network of main paths in 1980, in Fig. 6, shows a further development of the STS field, while neither INN nor ENT yet show up very prominently. We find two paths that appear to form the

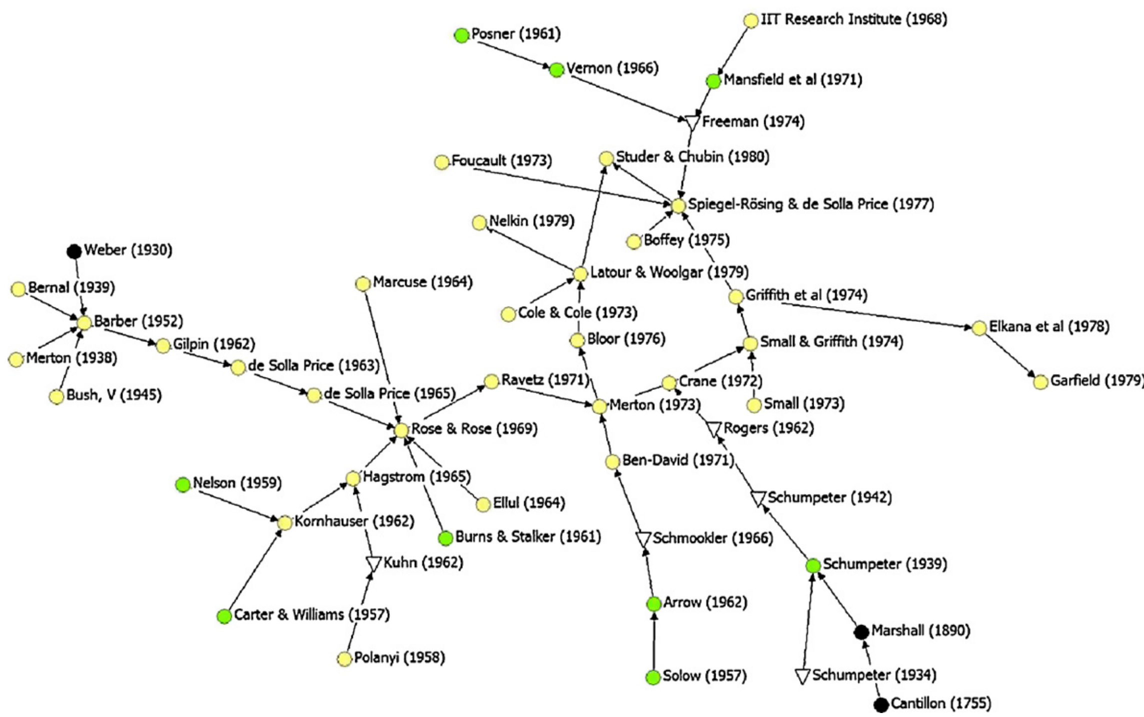

Fig. 6. Network of main paths for the entire network up to and including 1980. Colors/shades indicate the field that a paper belongs to. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

main 'backbone' of the network, and which are almost exclusively STS in nature. There are some INN and ENT nodes that feed into this backbone, but in terms of main paths, INN and ENT are still dominated by STS, which has by far the most connected nodes. The left part of the figure reproduces almost identically the main component of the network of main paths in 1970, which was discussed in connection with Fig. 5. Rose and Rose (1969) is still a main attractor in this part of the figure, and this particular path continues with the publications of Ravetz (1971), Merton (1973), Bloor (1976), Latour and Woolgar (1979), and Studer and Chubin (1980). Following the historical description of the evolution of STS by Martin et al. (2012), we interpret this as a stream of literature that builds upon Bernal (1939), Barber (1952), Kuhn (1962) and Gilpin (1962) and their development of a non-'traditional' (i.e. opposing the traditional view of science that was dominant up to the 1960s) approach to STS, and which in this period is beginning to evolve into the idea of the "social construction of science". It is not surprising that nearly all of these publications have the term 'sociology', 'sociological' or 'society' along with 'science' or 'scientific' in their title. Ravetz (1971) analyzes social problems of scientific knowledge, while Merton's book (1973) is a theoretical and empirical study of the sociology of science, and Bloor's book (1976) is one of the most influential works from the Edinburgh school. Latour and Woolgar (1979) investigate the social construction of scientific facts. Finally, Studer and Chubin (1980), studying the social contexts of biomedical research, ask themselves if "*ideas are independent or dependent variables for the sociology of science*" (p. 6).

There are several short branches that feed into this part of the main path. Besides the pre-1970s literature already discussed, three new STS publications make a contribution: Nelkin (1979), Cole and Cole (1973) and Ben-David (1971). The last of these, which analyzes the role of scientists in society, is interesting in that it links up with the INN-based works belonging to the previously separate 'economics of growth and technology' path apparent in 1970, i.e., the sequence from Solow (1957) to Arrow (1962) and then Schmookler (1966).

Another, parallel part of the backbone of the network in Fig. 6 starts with the INN and ENT path that was already apparent in the 1970 figure (Cantillon, Marshall, Schumpeter and Rogers), and continues through Crane (1972) to Small and Griffith (1974) and Griffith et al. (1974). This path, together with the additional feed-in node of Small (1973) and the spin-off node of Elkana et al. (1978) and Garfield (1979), forms a thematic cluster focusing on the use of science indicators such as citations and co-citations, among other things for the identification of scientific communities. This group of studies can therefore be labeled the '*Scientometric cluster*'. As argued by Martin et al. (2012), science indicators were an important part of the STS community up to the late 1970s, and this is consistent with the prominent position of this component of the main path in the figure for 1980. Afterwards scientometric researchers started to drift apart from STS, ultimately becoming more of a separate community. However, in 1980, the scientometric cluster is still well connected to the main STS paradigm, and indeed flows into Spiegel-Rösing and de Solla Price (1977), which is where all the main paths converge in the 1980 figure.

The identification of Spiegel-Rösing and de Solla Price (1977) as the main point of convergence in Fig. 6 is a significant finding, since this publication is the first STS handbook[15] to be published. The fact that this draws on all the paths emerging from the 1980 picture of the evolution of the three fields indicates how wide-ranging STS was up until that time. The fact that Freeman, a central figure in the development of innovation studies, has a chapter in this handbook further illustrates the point.[16]

Fig. 7 continues the analysis up to 1990. Here we see a major break with the previous figure in that we now see an INN group emerging that contains a significant number of nodes as well as forming a more or less coherent set of main paths. One part of this newly emerging INN field is a component that is disconnected from the rest of the network, and which mainly consists of a set

[15] The second STS handbook was published in 1995.

[16] We thank one of the referees for reminding us about this.

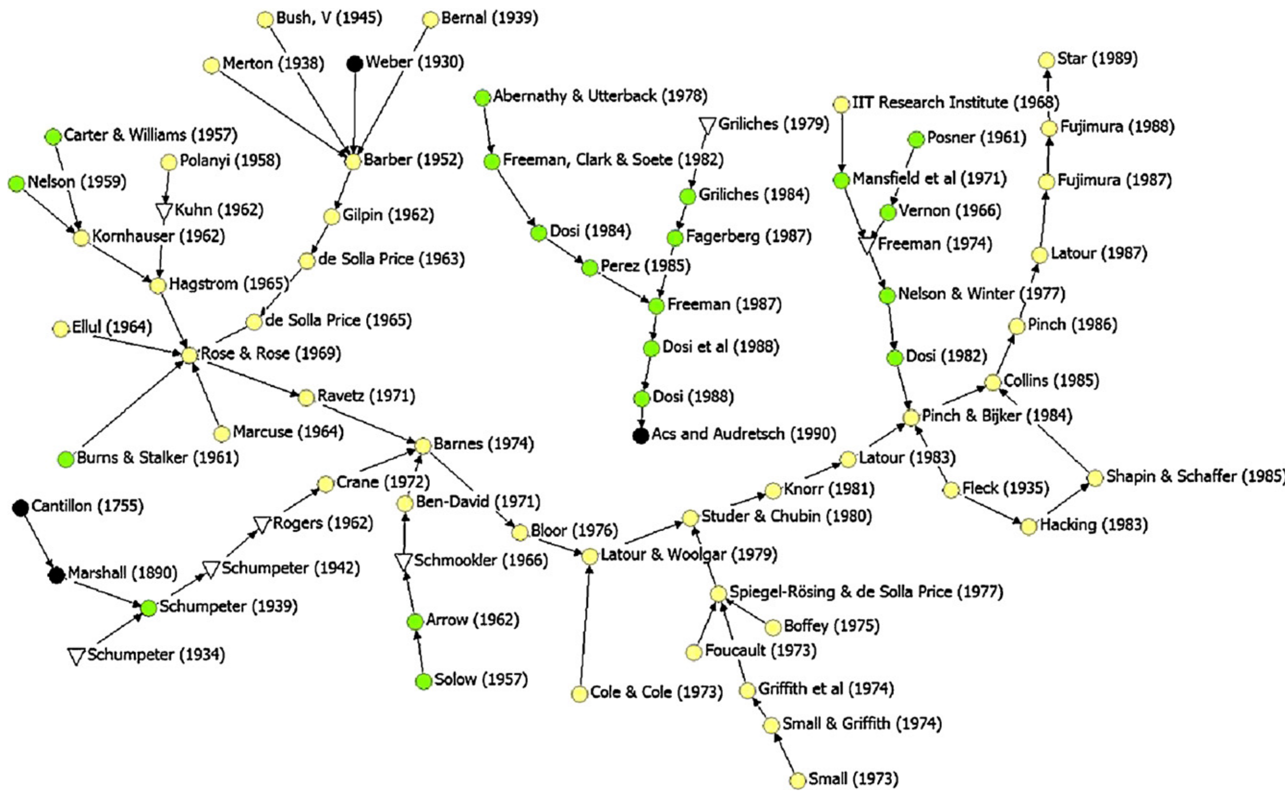

Fig. 7. Network of main paths for the entire network up to and including 1990. Colors/shades indicate the field that a paper belongs to. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

of European-based authors in the neo-Schumpeterian tradition – namely Freeman et al. (1982), Dosi (1984), Perez (1985), Freeman (1987), Fagerberg (1987) and Dosi et al. (1988). This corresponds to the SPRU-based tradition described in Fagerberg et al. (2012), although the component also includes three U.S.-based authors in the form of Abernathy and Utterback (1978) and Griliches (1979, 1984). This INN component covers a variety of topics. Abernathy and Utterback (1978) analyze the patterns of industrial innovation, while Freeman (1987) is the first publication to introduce the concept of a "system of innovation", drawing both from the macro aspects dealt with in the above mentioned papers, as well as from Griliches (1979, 1984) and Fagerberg (1987). Dosi et al. (1988) played a role in furthering the development of evolutionary economics, as can be seen from the trajectory. This publication can, to a certain extent, be viewed as an early handbook, which took stock of the field at that time (Fagerberg et al., 2012). It is interesting to note that this newly emerging INN component later converges on Acs and Audretsch (1990), which, as we shall see, would later form part of the ENT field. Apart from this, ENT is only visible through early publications such as Schumpeter (1939, 1942) which would eventually form part of its foundations.

Despite the fact that a number of INN nodes are now disconnected from the rest of the network, beginning to "start" their own core field, there are also still some INN nodes that are connected to the main part of the network (i.e., to the STS literature). In particular, the Mansfield–Posner–Vernon–Freeman cluster that was visible in the previous figure is now extended with Nelson and Winter (1977) and Dosi (1982). This branch is still connected to the STS field rather than to the new and now somewhat separate INN network (i.e., it still has its strongest citation impact through STS). This is an interesting result, which supports our decision to focus on multiple main paths (rather than a single one). It suggests that, in the early stages, the SPRU tradition (as reflected in its citation impact) was as much one of the "feeders" into STS (through Dosi, 1982) as it was a part of the emergence of the new INN field.

In the major component of the network in Fig. 6, which is still STS-dominated, we note that Spiegel-Rösing and de Solla Price (1977), which was the publication that the 1980 network converged into, now appears as something of a side stream. This is somewhat surprising in the light of what we concluded from our analysis of the 1980 figure. One possible explanation is that the previously important scientometrics cluster had by now begun to fall out of favor among many STS researchers. This component is now much less prominent than in Fig. 6, which is not to imply that its development stopped, merely that it occurred increasingly outside the main STS and INN communities. Only the 'non-traditional' component of the pre-1980 STS main path is carried on directly into the following decade with publications such as those from the Edinburgh school (Shapin and Schaffer, 1985), and by other sociologists such as Collins (1985), Latour (1983 and 1987) and Pinch (1986) among the elements of the STS main path up to 1990. Finally, another reason of the lack of direct linkages between the more recent STS developments and Spiegel-Rösing and de Solla Price's handbook (1977) is that the INN path which flowed into it in 1980 had since moved away, now reconnecting directly to the newer developments of STS.

Two other interesting publications on the 1990 STS main path are the works by Pinch and Bijker (1984) and Fleck (1935). The former, as testified by its citation to Dosi (1982), opened the way for the integration of technology and its economic analysis into the STS discourse, which was no longer focused so exclusively on science. The linkage from Dosi (1982) to Pinch and Bijker (1984) now fed the "old" sequence of Posner (1961) – Vernon (1966) – Freeman (1974), as well as Nelson and Winter (1977) into the STS network. One may also note that this particular link in the main path network may reflect a certain amount of "disagreement" between Dosi's essentially "positivistic" notion of technology and innovation, on the one hand, and Pinch and Bijker's social constructivist view, on the other. It may appear somewhat strange that the main path links together two publications that represent opposing sides of a debate that later became quite heated. The citation analysis, of course, does not recognize the exact nature of the citation and whether it points to agreement or disagreement. But in any case, even disagreement indicates a certain influence, and in that sense the citation link is undoubtedly important. Possibly, this particular linkage may also be the result of Bijker's local interaction in Maastricht with the newly established MERIT institute, which under the guidance of its director, Luc Soete, who was a member of SPRU staff at the time of his PhD work and shortly after that, continued the SPRU INN tradition on mainland Europe. Fleck (1935), on the right hand side of the figure, is a 'new entrant' into the network of main paths. According to Martin et al. (2012), Fleck (1935) was one of the works from the first half of the 20th century that was to prove crucial in the development of a distinct non-Mertonian approach to STS. However, it is only with the later work by Hacking (1983) and Pinch and Bijker (1984) that Fleck became recognized in this way.

Fig. 8 shows the network of main paths for the complete citation network up to 2002. This is the first time that, in addition to an STS and an INN component, we can now identify a distinct ENT component, although it is still small compared to the other two fields. For STS, we see a continuation of the main path as identified before. Latour (1987) is the reference from which the main STS path continues, with the publications by Fujimura (1987, 1988) and Star (1989) no longer in the main path. From Latour (1987), there is a simple (non-branching) continuation of the main path through Traweek (1988), Haraway (1989, 1991), Law (1991), Pickering (1995) and Knorr (1999), forming the backbone of the STS component. That these newer STS additions to the network follow in a single path from earlier developments of the fields, building on a backbone that was already fairly well-defined in the previous decade, might suggest that a relatively mature STS paradigm has perhaps begun to emerge.

There appears to be rather more turbulence with regard to the INN component of Fig. 8. One important change compared to Fig. 7 is that the branch that includes Nelson and Winter (1977) and Dosi (1982) now links up not with the STS component as previously, but with the INN component that emerged for the first time in the previous figure. This switch suggests a growth in the maturity of the INN field at this time, with this group of publications now having their largest citation impact not through STS but through the emerging INN field. While in previous decades, the majority of the connections of these documents were to the STS literature, now they have amassed a larger number of downstream connections in INN. Interestingly, this is not the case, however, for the very early U.S.-based publications on the economics of technological change, such as Solow (1957), Arrow (1962) and Nelson (1959), which still appear as early inputs to the main STS component.

The second major development in the INN component is a switch from the macroeconomic and industrial economics topics of the SPRU-tradition, to a more business- and management-oriented direction. This starts with Porter (1990), who exemplifies the switchover, being a scholar rooted in the management literature[17] but dealing with policy-related issues surrounding the competitive advantage of nations. The following node, Bygrave and Timmons (1992) is (somewhat surprisingly, perhaps) from ENT, while later publications such as Leonard-Barton (1995) and Teece et al. (1997) are in the tradition of strategic management. This is in line with the conclusion of Fagerberg et al. (2012) that the INN field has seemingly undergone a major change in direction since the early 1990s.

In ENT, by contrast, we find a relatively simple linear main path that starts with Birch (1979), who is considered one of the "insiders" in the field, and ends with Klepper (2001). Two of the intermediate nodes, namely Kent et al. (1982) and Sexton and Smilor (1986), are both handbooks from a series. Gartner (1988) and Aldrich (1990) deal with the creation of organizations. Gartner in particular is regarded as a "domain-defining" work, which moved the field away from the earlier "behavioral approach" to entrepreneurship (Landström et al., 2012). The last part of the main ENT path, i.e., Shane (2000), Shane and Venkataraman (2000) and Klepper (2001), focuses on the creation of new firms and markets.

What emerges clearly from the network of main paths in 2002 is the almost complete separation of the three fields, based around the emergence of three largely separate bodies of literature. Interactions among STS, INN and ENT are now rather rare and most often the legacy of earlier years. The lack of recent interactions between the three fields is somehow at odds with the common roots that they shared in the past. This would tend to support the interpretation of Martin et al. (2012) that, although STS and INN shared certain commonalities in the past, STS has since diverged from what Martin labels 'Science Policy and Innovation Studies' (SPIS), the publications of which are largely covered in our dataset for the INN field.

A final point should be noted about the nature of the documents on the main path of the network, as described by Figs. 5–8. While these figures include many of the publications that were identified as crucially important in the three field studies, they also exclude certain publications that a casual observer might identify as particularly crucial ones. Table 1 lists the top-10 publications in our database on the basis of citations received and citations made. These papers are ranked on "throughput", which is defined as the number of (direct) citations that a paper makes (to other papers in the database) multiplied by the number of citations that the paper receives (by other papers in the database). The throughput measure is somewhat similar to the weights that the Hummon and Doreian procedure assigns, the main difference being that only direct citations are taken into account.

In Table 1, we see that three of the top-10 publications never appear in any of the main paths. The most striking omission is Nelson and Winter (1982), which appears in all three core lists of literature (indeed, it is the only publication that does so). It is also the publication with the highest throughput. Yet it does not appear in any of the main paths. Part of the reason may be that a predecessor publication, namely Nelson and Winter (1977), does appear. The 1982 book receives many citations (both from within our database, and from outside, i.e., from the general literature), providing a number of 'shortcuts' in the citation network. Nelson and Winter (1982) is undoubtedly a crucial reference in the literature, but with regard to our main paths, it is not "path-defining", while Nelson and Winter (1977) evidently is.

The second reference that appears in Table 1 but not in any of the main paths is Bijker et al. (1987). This is certainly a very important publication in STS, as it sets the scene for the appearance of the idea of social construction of technology (SCOT). Interestingly, here we again have a precursor publication in the form of Pinch and Bijker (1984), which does appear as an important node on the main path in STS. Thus, the situation is perhaps somewhat similar to Nelson and Winter, with the earlier publication obtaining priority in the main path of the literature.

Rosenberg (1982) is the last case that appears in Table 1 but not in any of the main paths. Here we do not have any precursor that is in the main paths, although this Rosenberg volume again contains many chapters that had been previously published. On their own,

[17] As one referee pointed out, Porter also worked in industrial economics before moving into management. However, all of Porter's publications in our database are clearly management-oriented.

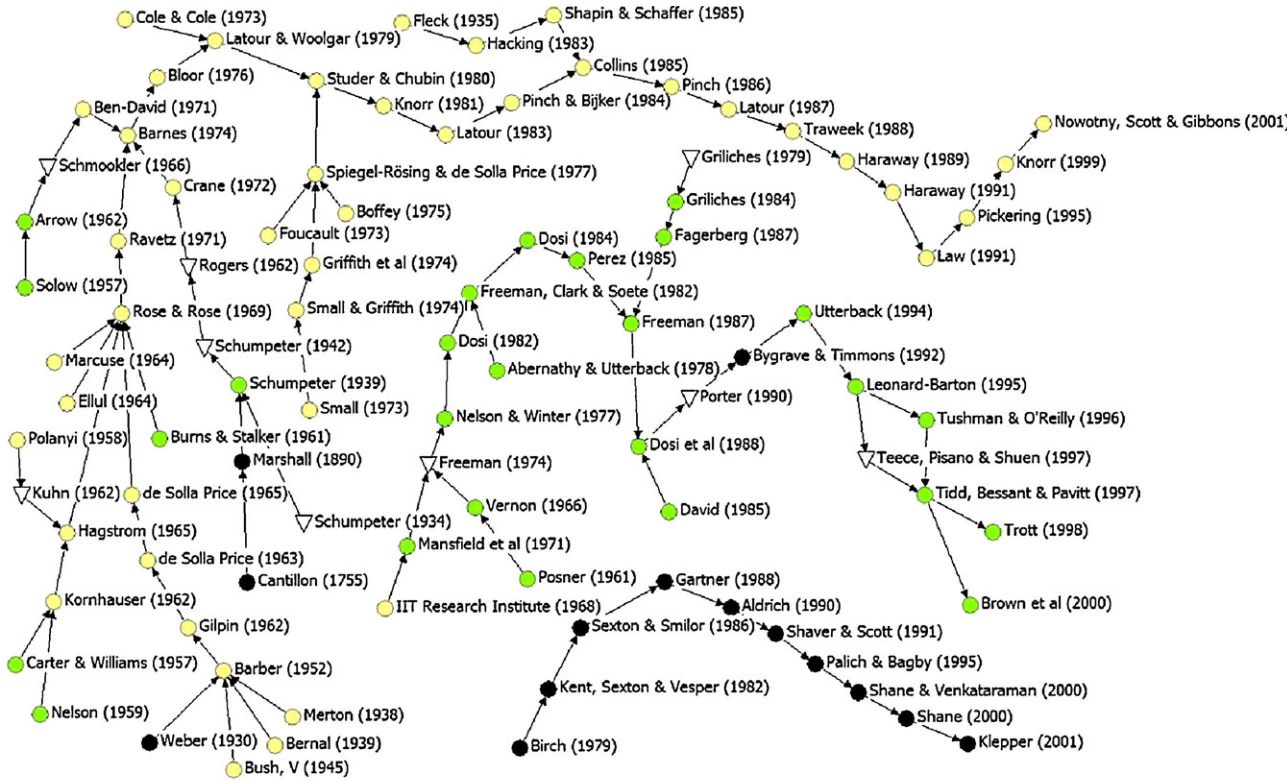

Fig. 8. Network of main paths for the entire network up to and including 2002. Colors/shades indicate the field that a paper belongs to. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 1
Top-10 documents in terms of "throughput" (entire database).

Document	Field	In main path network of	#Citations received	#Citations made	Throughput (cit. made × cit. received)
Nelson and Winter (1982)	INN/ENT/STS	Never	50	28	1400
Latour and Woolgar (1979)	STS	1980, 1990, 2002	55	15	825
Dosi et al. (1988)	INN	1990, 2002	12	59	708
Bijker et al. (1987)	STS	Never	19	37	703
Dosi (1988)	INN	1990	13	41	533
Rosenberg (1982)	INN	Never	28	19	532
Collins (1985)	STS	1990, 2002	24	22	528
Porter (1990)	INN/ENT	2002	26	20	520
Latour (1987)	STS	1990, 2002	30	17	510
Freeman (1974)	INN/STS	1980, 1990, 2002	39	13	507

these chapters did not make it to the database, and despite the fact that the overall volume has a high throughput value, it does not figure on any of the main paths.

6. Conclusions

The overall main conclusion from our analysis is that the social science literature on knowledge, technological change and innovation has developed in a progressively more compartmentalized manner. In terms of their citation profiles, the three fields of science and technology studies (STS), innovation studies (INN) and entrepreneurship (ENT) now appear as largely distinct, not as part of a strongly connected field. This can be seen from the relatively low importance of between-field citations between the core publications of the three fields. Our citation-based cluster analysis of all references in the entire database of core publications for the three fields yields a breakdown that is very similar to the actual division between the fields. INN assumes a somewhat special role in this respect, as it seems to be positioned between the other two fields. This is evident both from the between-field citation pattern, and from the classification exercise, where the few "mis-classifications" are all cases of publications being "wrongly" assigned to innovation studies.

This is most apparent at the end of the period that we consider (the early 21st Century). Further back in time, it has not always been the case that STS, INN and ENT appear as clearly separated fields. The three fields share a large number of common roots in terms of their early sources of inspiration. In the 1960s and 1970s, INN and STS still overlapped to a considerable degree, while ENT did not yet exist as a separate body of literature. This is evident from our citation network analysis, which shows that, in the 1960s and 1970s, the "main paths" in the networks (i.e., the citation chains that correspond to the strongest knowledge flows) consisted of documents coming from both STS and INN. The interpretation of this finding is that STS and INN used to cite each other quite frequently, and that a logical and consistent back and forth flow of ideas can be seen from a chain of documents that contains both STS and INN publications.

The period before 1970, which can be characterized as the early beginnings of the social science of knowledge, is particularly fluid in this respect. Around 1970, we observe a situation in which many of the contributions that would later become characteristic for one of the three fields, would still interact (in terms of forward and backward citations) not just with each other but also with a broad range of publications from the other two fields. This is a stage in which the state of the social science knowledge in this field was still fluid, and the main trajectories had not yet been completely established.

From the early 1970s onwards, STS developed into a field in which a more cumulative path of knowledge developed around the sociological perspective on societal influences and the social embedding of science and knowledge in general. This main trajectory of the literature can be characterized as one that is critical about the "ideal picture" of science as a value-free and "pure" quest for knowledge. Instead, STS looks at science as a social process in itself, in which differing values and norms may lead the scientific process in different directions.

Somewhat later, INN also established a main trajectory. Here the development appears to be less continuous, with our analysis suggesting a major shift of focus in the early 1990s. Before this breakpoint, macroeconomics and industrial dynamics, two topics strongly influenced by heterodox economics, had dominated, but by the start of the 1990s, business and management studies were beginning to assume a position of prominence in the field of innovation studies, while economics as a discipline disappeared from the main trajectory in the field.

Entrepreneurship emerges as the youngest of the three fields. Although the roots of this field go back in time at least as far as those for the other two fields, its development seems much more in its infancy. A clear trajectory appears to have developed only around the turn of the millennium.

These descriptions of the main trajectories of knowledge development seem broadly consistent with the conclusions from the case studies carried out for each of the fields, and from which we draw our lists of core contributions. Thus, the quantitative analysis reported in this paper appears to work relatively well in terms of sharpening the conclusions from the case studies in the larger project, as well as identifying the specific interactions between the three emerging fields.

In terms of broader conclusions, our analysis provides a number of lessons about the development of social science. First, it shows how an initially "fluid" group of scholars studying a broadly related subject may over time evolve into different fields with only limited interaction. Partly, this is a process that is driven by specialization, focusing, for example, on the economic and later the management aspects of knowledge as opposed to the sociological aspects. However, we would also suggest that the process of diminishing interaction may be caused in part by a set of evolving values and norms that are specific to particular subgroups of scholars identified in our larger network. This is a conclusion that is very much in line with what STS scholars would contend about science in general. Our impression from the case studies of STS, INN and ENT is that such a process may also be at work here. Social scientists studying the societal and economic impact of knowledge gradually cluster into distinct subgroups that are driven by specific norms, beliefs and values that evolve in each subgroup. The results of our quantitative analysis are consistent with such a view, but only more specific qualitative research can provide further support for such a hypothesis.

Second, our results show that groups of social science scholars who study a set of closely related topics can develop in rather different ways. The STS field, as one example, has grown over time to become one with a large degree of internal consistency that results from a strongly cumulative development of the main building blocks of knowledge in the field. INN, on the other hand, has developed in a less cumulative, and more disruptive way. In the INN field, a major shift of focus is observed that corresponds to a change in focus (from macroeconomics and industrial dynamics to management and business related topics). How the ENT field will develop in this respect still remains to be seen. More generally, it would seem that the social science of knowledge and innovation is an interesting field that merits further study of its internal dynamics.

Acknowledgements

We thank participants in the EXPLORE project, three anonymous referees, the editors of the special issue, and participant at the DIME conference in Maastricht, April 2011, for useful comments and suggestions. This research was undertaken in the DIME Network of Excellence, sponsored by the European Commission's 6th Framework Programme.

References

Abernathy, W., Utterback, J., 1978, June/July. Patterns of industrial innovation. Technology Review, 40–47.
Acs, Z.J., Audretsch, D., 1990. Innovation and Small Firms. MIT Press, Cambridge, MA.
Aldrich, H., 1990. Using an ecological perspective to study organizational founding rates. Entrepreneurship: Theory and Practice 14, 7–24.
Arrow, K., 1962. The economic implications of learning by doing. Review of Economic Studies 29, 155–173.
Barber, B., 1952. Science and the Social Order. Free Press, Glencoe, IL.
Ben-David, J., 1971. The Scientist's Role in Society: A Comparative Study. Prentice Hall, Englewood Cliffs, NJ.
Bernal, J.D., 1939. The Social Function of Science. The Macmillan Company, New York.
Bijker, W.E., Hughes, T., Pinch, T.J., 1987. The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology. MIT, Press, Cambridge, MA.
Bloor, D., 1976. Knowledge and Social Imagery. Routledge & Kegan Paul, London.
Braunerhjelm, P., Acs, Z., Audretsch, D., Carlsson, B., 2010. The missing link: knowledge diffusion and entrepreneurship in endogenous growth. Small Business Economics 34, 105–125.
Birch, D., 1979. The Job Generation Process. MIT Program for Neighborhood and Regional Change, Cambridge, MA.
Bush, V., 1945. Science: The Endless Frontier. U.S. Government Printing Office, Washington.
Bygrave, W., Timmons, J., 1992. Venture Capital at the Crossroads. Harvard Business School Press, Boston, MA.
Cantillon, R., 1755. Essai sur la nature de commerce en géneral (Essay on the nature of trade in general) (H. Higgs (Ed.) 1931). Macmillan, London.
Cole, J.R., Cole, S., 1973. Social Stratification in Science. University Chicago Press, Chicago.
Collins, H.M., 1985. Changing Order: Replication and Induction in Scientific Practice. Sage, London.
Crane, D., 1972. Invisible Colleges: Diffusion of Knowledge in Scientific Communities. University of Chicago Press, Chicago.
De Solla Price, D.J., 1963. Little Science, Big Science. Columbia University Press, New York.
De Solla Price, D.J., 1965. Networks of scientific papers. Science 149, 510–515.
Dosi, G., 1984. Technical Change and Industrial transformation. Macmillan, London.
Dosi, G., 1982. Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. Research Policy 11, 147–162.
Dosi, G., 1988. Sources, procedures and microeconomic effects of innovation. Journal of Economic Literature 26, 1120–1171.
Dosi, G., Freeman, C., Nelson, R., Silverberg, G., Soete, L. (Eds.), 1988. Technical Change and Economic Theory. Pinter, London.
Elkana, Y., Lederberg, J., Merton, R.K., Thackray, A., Zuckerman, H., 1978. Toward a Metric of Science: The Advent of Science Indicators. John Wiley, New York.
Fagerberg, J., 1987. A technology gap approach to why growth rates differ. Research Policy 16, 87–101.
Fagerberg, J., Fosaas, M., Sapprasert, K., 2012. Innovation – exploring the knowledge base. Research Policy 41 (7), 1132–1153.
Fagerberg, J., Verspagen, B., 2009. Innovation studies – the emerging structure of a new scientific field. Research Policy, 218–233.
Fleck, L., 1935. Genesis and Development of a Scientific Fact. University of Chicago Press, Chicago.
Freeman, C., 1974. The Economics of Industrial Innovation. Penguin Modern Economics Texts, London.
Freeman, C., 1987. Technology Policy and Economic Performance. Lessons from Japan, Pinter, London.
Freeman, C., Clark, J., Soete, L., 1982. Unemployment and Technical Innovation: A Study of Long Waves and Economic Development. Frances Pinter, London.
Fujimura, J., 1987. Constructing "do-able" problems in cancer research: Articulating alignment. Social Studies of Science 17, 257–293.
Fujimura, J., 1988. The molecular biological bandwagon in cancer research: Where social worlds meet. Social Problems 35, 261–283.

Garfield, E., 1979. Citation Indexing: Its Theory and Application in Science, Technology and Humanities. Wiley, New York.
Gartner, W., 1988. Who is an entrepreneur?" Is the wrong question. American Journal of Small Business 12, 11–32.
Gilpin, R., 1962. American Scientists and Nuclear Weapons Policy. Princeton University Press, Princeton, NJ.
Gilpin, R., Wright, C., 1964. Scientists and National Policy-making. Columbia University Press, New York.
Griffith, B.C., Small, H., Stonehill, J.A., Dey, S., 1974. The structure of scientific literatures. II: toward a macro- and microstructure for science. Science Studies 4, 339–365.
Griliches, Z., 1979. Issues in assessing the contribution of research and development to productivity growth. Bell Journal of Economics 10, 92–116.
Griliches, Z., 1984. R&D, Patents, and Productivity. University of Chicago Press, Chicago.
Hacking, I., 1983. Representing and Intervening: Introductory Topics in the Philosophy of Natural Science. Cambridge University Press, Cambridge.
Hagstrom, W.O., 1965. The Scientific Community. Basic Books, New York.
Hall, K., 1970. R-dimensional quadratic placement algorithm. Management Science 17, 219–229.
Haraway, D., 1991. Simians, cyborgs, and women: The reinvention of nature. Routledge & Kegan Paul, New York.
Haraway, D., 1989. Primate visions: Gender, race, and nature in the world of modern science. Routledge & Kegan Paul, New York.
Hummon, N.P., Doreian, P., 1988. Connectivity in a citation network: the development of DNA theory. Social Networks 11, 39–63.
Kent, C.A., Sexton, D., Vesper, K. (Eds.), 1982. Encyclopedia of Entrepreneurship. Prentice-Hall, Englewood Cliffs, NJ.
Kirchhoff, G., 1847. Uber die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird. Annals of Physical Chemistry 72, 497–508.
Klepper, S., 2001. Employee startups in high-tech industries. Industrial and Corporate Change 10, 639–674.
Knorr, K., 1999. Epistemic Cultures: How the Sciences Make Knowledge. Harvard University Press, Cambridge, MA.
Kornhauser, W., 1962. Scientists in Industry: Conflict and Accommodation. University of California Press, Berkeley.
Kotler, P., 1976. Marketing Management: Analysis, Planning, Implementation and Control. Prentice Hall International, Eaglewood Cliff, NJ.
Kuhn, T., 1962. The Structure of Scientific Revolution. Chicago University Press, Chicago.
Landström, H., Harirchi, G., Åström, F., 2012. Entrepreneurship – exploring the knowledge base. Research Policy 41 (7), 1154–1181.
Latour, B., 1983. Give me a laboratory and I will raise the world. In: Knorr Cetina, K., Mulkay, M. (Eds.), Science Observed: Perspectives on the Social Study of Science. Sage, Beverly Hills, CA, pp. 141–170.
Latour, B., 1987. Science in Action: How to Follow Scientists and Engineers through Society. Harvard University Press, Cambridge, MA.
Latour, B., Woolgar, S., 1979. Laboratory Life: The Social Construction of Scientific Facts. Sage, Beverly Hills, CA.
Law, J., 1991. A Sociology of Monsters: Essays on Power, Technology and Domination. Routledge & Kegan Paul, London.
Leonard-Barton, D., 1995. Wellsprings of Knowledge: Building and Sustaining the Sources of Innovation. Harvard Business School Press, Boston.
Lundvall, B.-Å, Borrás, S., 2004. Science, technology, and innovation policy. In: Fagerberg, J., Mowery, D.C., Nelson, R.R. (Eds.), The Oxford Handbook of Innovation. Oxford University Press, New York, pp. 599–631.
Mansfield, E., 1968. Industrial Research and Technological Innovation. Norton, New York.
Marshall, A., 1890. Principles of Economics. MacMillan, London.
Martin, B.R., Nightingale, P., Yegros-Yegros, A., 2012. Science and technology studies: exploring the knowledge base. Research Policy 41 (7), 1182–1204.
Merton, R.K., 1938. Science, technology and society in seventeenth century England. Osiris 4, 360–632.
Merton, R.K., 1973. The Sociology of Science: Theoretical and Empirical Investigations. University of Chicago Press, Chicago.
Nelkin, D., 1979. Controversy, Politics of Technical Decisions. Sage, Beverly Hills, CA.
Nelson, R., 1959. The simple economics of basic scientific research. Journal of Political Economy 67, 297–306.
Nelson, R., 1968. A ‘diffusion’ model of international productivity differences in manufacturing industry. American Economic Review 58, 1219–1248.
Nelson, R., Winter, S., 1977. In search of a useful theory of innovation. Research Policy 5, 36–76.
Nelson, R., Winter, S., 1982. An Evolutionary Theory of Economic Change. The Belknap Press of Harvard University Press, Cambridge, MA.
Newman, M., Girvan, M., 2004. Finding and evaluating community structure in networks. Physical Review E 69, 026113.
Newman, M.E.J., 2004. Modularity and community structure in networks. Proceedings of the National Academy of Sciences United States of America 103, 8577–8582.
Perez, C., 1985. Microelectronics, Long Waves and the World Structural Change: New Perspectives for Developing Countries. World Development, 13, pp. 441–463.
Pickering, A., 1995. The Mangle of Practice: Time, Agency and Science. University of Chicago Press, Chicago.
Pinch, T., 1986. Confronting nature: the sociology of solar-neutrino detection. Reidel, Dordrecht, the Netherlands.
Pinch, T., Bijker, W.E., 1984. The social construction of facts and artifacts, Or how the sociology of science and the sociology of technology might benefit each other. Social Studies of Science 14, 399–441.
Polanyi, M., 1958. Personal Knowledge: Towards a Post-Critical Philosophy. University of Chicago Press, Chicago.
Posner, M., 1961. International trade and technical change. Oxford Economic Papers 13, 323–341.
Porter, M., 1990. The Competitive Advantage of Nations. Macmillan, London.
Ravetz, J.R., 1971. Scientific Knowledge and its Social Problems. Clarendon Press, Oxford.
Rogers, E.M., 1962. Diffusion of Innovations. The Free Press, New York.
Rose, H., Rose, S., 1969. Science and Society. Allen Lane, London.
Rosenberg, N., 1982. Inside the Black Box. Cambridge University Press, Cambridge.
Schmookler, J., 1966. Invention and Economic Growth. Harvard University Press, Cambridge, MA.
Schumpeter, J.A., 1939. Business Cycles: A Theoretical, Historical and Statistical Analysis of the Capitalist Process, Vols. 2. McGraw-Hill, New York.
Schumpeter, J.A., 1942. Capitalism, Socialism, and Democracy. Harper, New York.
Sexton, D., Smilor, R. (Eds.), 1986. The Art and Science of Entrepreneurship. Ballinger, Cambridge, MA.
Shane, S., 2000. Prior knowledge and the discovery of entrepreneurial opportunities. Organization Science 11, 448–469.
Shane, S., Venkataraman, S., 2000. The promise of entrepreneurship as a field of research. Academy of Management Review 25, 217–226.
Shapin, S., Schaffer, S., 1985. Leviathan and the Air-pump: Hobbes, Boyle and the Experimental Life. Princeton University Press, Princeton, NJ.
Shibata, N., Kajikawa, Y., Takeda, Y., Matsushima, K., 2008. Detecting emerging research fronts based on topological measures in citation networks of scientific publications. Technovation 28, 758–775.
Skolnikoff, E.B., 1967. Science, Technology and American Foreign Policy. MIT Press, Cambridge, MA.
Small, H., Griffith, B.C., 1974. The structure of scientific literatures I. Identifying and graphing specialties. Science Studies 4, 17–40.
Small, H., 1973. Co-citation in the scientific literature: a new measure of the relationship between two documents. Journal of the American Society for Information Science 24 (4), 265–269.
Solow, R., 1957. Technical change and the aggregate production function. Review of Economics and Statistics 34, 312–320.
Spiegel-Rösing, I.S., de Solla Price, D.J., 1977. Science, Technology and Society: A Cross-disciplinary Perspective. Sage, London.
Star, S.L., 1989. Regions of the Mind: Brain Research and the Quest for Scientific Certainty. Stanford University Press, Stanford, CA.
Studer, K.E., Chubin, D.E., 1980. The Cancer Mission: Social Contexts of Biomedical Research. Sage, Beverly Hills, CA.
Teece, D.J., Pisano, G., Shuen, A., 1997. Dynamic capabilities and strategic management. Strategic Management Journal 18, 524–526.
Traweek, S., 1988. Beamtimes and Lifetimes: The World of High Energy Physicists. Harvard University Press, Cambridge, MA.
Vernon, R., 1966. International investment and international trade in the product cycle. Quarterly Journal of Economics 80, 190–207.
Verspagen, B., 2007. Mapping technological trajectories as patent citation networks: a study on the history of fuel cell research. Advances in Complex Systems 10, 93–115.
Weber, M., 1930. The Protestant Ethic and the Spirit of Capitalism. Scribner's, New York.
White, S., Smyth, P., 2005. A spectral clustering approach to finding communities in graphs. In: Kargupta, H., Srivastava, J., Kamath, C., Goodman, A. (Eds.), Proceedings of the 5th SIAM International Conference on Data Mining, Society for Industrial and Applied Mathematics. Philadelphia, pp. 274–285.