The Self-Organization of the European Information Society: The Case of "Biotechnology"

Loet Leydesdorff
University of Amsterdam, Amsterdam School of Communications Research ASCoR, Kloveniersburgwal 48, 1012 CX, Amsterdam, The Netherlands. E-mail: loet@levdesdorff.net; http://www.leydesdorff.net/

Gaston Heimeriks
Social Science Informatics, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands. E-mail: heimerik@swi.psy.uva.nl; http://swi.psy.uva.nl/usr/heimeriks/home.html

Fields of technoscience like biotechnology develop in a network mode: disciplinary insights from different backgrounds are recombined as competing innovation systems are continuously reshaped. The ongoing process of integration at the European level generates an additional network of transnational collaborations. Using the title words of scientific publications in five core journals of biotechnology, multivariate analysis is used to distinguish between the intellectual organization of the publications in terms of title words and the institutional network in terms of addresses of documents. The interaction among the representation of intellectual space in terms of words and co-words, and the potentially European network system is compared with the document sets with American and Japanese addresses. The European system can also be decomposed in terms of the contributions of member states. Whereas a European vocabulary can be made visible at the global level, this communality disappears by this decomposition. The network effect at the European level can be considered as institutional more than cognitive.

Introduction

A common root of sociological and computer science traditions is the "radically constructivist" assumption. Are the abstract constructions visible in the simulations also retrievable in the complex reality of social interactions? Can the sociological reconstructions be informed from the engineering and mathematical perspectives of the computer sciences? The recombination of qualitative specification with quantitative modeling can be expected to stimulate the development of new methodologies to understand the complex social processes that characterize the development of the information society and its relation to the formation of a European identity.

The theory of self-organization or "autopoiesis," for example, can be made relevant when the reconstructing network layer and the reconstructed layer operate selectively upon each other like in a process of "mutual shaping." This process may develop a momentum of its own when the selecting and the selected dimensions can change positions and then begin to interact in a coevolution along a specific ("path-dependent") trajectory.

In the case of the European information society, two major processes of reconstruction can be distinguished, indeed. First, there is the networking of the European Union in terms of institutional addresses. Among other things, the series of Framework Programs and other special programs have stimulated both transnational collaboration and collaboration across institutional sectors. Laredo (1997) showed that university–industry–government relations have gathered institutional momentum at the European level. In this article, we address the question of whether the socio-cognitive interaction also has gained momentum at a specifically European level? Can this momentum be observed using scientometric measurements?

In parallel to the geographical (e.g., European or American) and institutional networks, the new technosciences like biotechnology, new materials, and information technologies contain another network dynamics, namely, at the cognitive and therefore potentially global level. Gibbons, Limoges, Nowtng, Schwartzman, Scott, & Trow (1994) have coined the word "Mode-2 research" for areas like biotechnology. The contexts of application have been considered constitutive of the development of these fields. Thus, the so-called "triple helix of university–industry–government relations" (Etzkowitz & Leydesdorff, 1997) can be considered as not only an institutional, but also a

Received November 2, 2000; Revised April 10, 2001; accepted June 5, 2001

substantive development. From this global perspective, the European policies construct an intermediate network layer among competing (institutional) agencies.

In summary, a double-layered construction of networks that fulfills the conditions of autopoiesis in the European case can be hypothesized. Can the self-organization of the communication in this case be measured by using literature-based indicators? In this study, we explore this question by limiting ourselves to "biotechnology" as a typical "Mode-2" field (McKelvey, 1996), and by assuming a competition at the global level between the emerging European networks, on the one hand, and the American and Japanese research efforts, on the other.[1]

Of course, there are many more countries in the world than these three blocks of advanced industrial nations. However, we know from previous research that including more countries is not always helpful from an analytical perspective. For example, including Canada makes it sometimes more difficult to distinguish between the United States and Europe because of this country's position between the United States and the United Kingdom research systems (Leydesdorff & Gauthier, 1996). In this study, we analyze biotechnology at different levels of aggregation to better understand the possible emergence of a specific European dimension (cf. Leydesdorff, 1992a).

Operationalization of the Research Question

We hypothesized above the interaction of two operating systems: the European networks among (hitherto national) institutions and the technoscience network of biotechnology. In general, networks operate like structures by selecting on the variation; the observable variation can also be considered as a result of the interaction among these selections. If the interaction (or covariation) is repeated over time, a coevolution can be induced along a trajectory. Historical trajectories can be selected recursively by a self-organizing system because the system reflexively reconstructs continuously its present operation (Dosi, 1982; Leydesdorff, 2001; Maturana & Varela, 1980, 1984).

To operationalize these concepts, we first distinguish the selecting operations analytically. The field level will here be operationalized in terms of five major biotechnology journals. These journals select upon the variation of knowledge claims using cognitive specificity as the crucial criterion (Fujigaki, 1998). In a next section, we shall explain how we made a selection of journals using aggregated journal–journal citations as listed in the *Journal Citation Reports (JCR)* of the *Science Citation Index (SCI)*. We used the latest available update at the time of the original data collection (July 1998), that is, the *JCR* of 1996. When the *JCR* for 1997 became available (November 1998), this data was used to replicate the analysis.

[1] In a related project, comparable events in the case of information science are analyzed (Van den Besselaar & Heimeriks, 2000; cf. Van den Besselaar & Leydesdorff, 1996).

The European dimension can be operationalized in terms of the institutional addresses of the documents in the journals under study. All articles with an address in one of the 15 member states are selected. Similarly, an American and Japanese dataset was generated for the sake of comparison. In the case of internationally coauthored articles we attributed the article in full to each of the implied nations (so-called "integer counting").[2] Note that by using this design, the cognitive selection and the institutional selections are analytically independent: we study a sample of documents that fulfill two selection criteria, notably belonging to the field of biotechnology (in terms of its major journals) *and* containing an institutional address in the geographical areas under study.

In summary, the research question can be formulated as follows: how can this data be used in order to show self-organization as a complex dynamics in the case of both social and cognitive transformation processes? As noted, we wish to study the interaction between the ongoing reorganization of the European S&T system by processes of transnational collaboration, and the transformation of the sciences involved in "biotechnology" by the global transition towards Mode-2 research. The two dynamics, however, are expected to interact. Can this interaction be considered as a "feed-back" or as a "feed-forward" loop? Is the European intervention advantageous to the development of this field of technoscience, and if so, at which level? Can "loosers" and/or "winners" be indicated?

In other words: is it possible to show an effect in the European case which is different from the American and Japanese cases? Whereas the global dynamics in these latter cases are firmly stabilized within "national systems of innovation" (Lundvall, 1992; Nelson, 1993), the inter-European dynamics provides the participants with an additional degree of freedom (Frenken, 2000). Does this additional degree of freedom make a difference that can be measured?

Methods and Data

Delineation of the Domain

The definition of "biotechnology" itself has changed in recent decades, and it varies among goegraphical regions (OECD, 1988). If one follows the actors historically (Latour, 1987), one is able to obtain a reflexive understanding about how these definitions have changed (Nederhof, 1988). With hindsight, however, one expects that some of these historical elements have been carried over into the current understanding, while other elements may have faded away. The evolutionary perspective is obtained by focussing on the *operation* of the system in the present. The discursive

[2] Integer counting has been distinguished in the scientometric literature as an alternative to fractional counting (Narin, 1976). In the latter case, a relative weight is attributed in accordance with the number of coauthorship relations. See for a discussion: Anderson, Collins, Irvine, Isard, Martin, Narin, & Stevens (1988).

system under study is reflexive with reference to its historical manifestations. The history of the system is continuously rewritten and therefore the disciplinary discourse cannot be identified other than as the (theoretically informed) hypothesis of a cognitive dynamics.

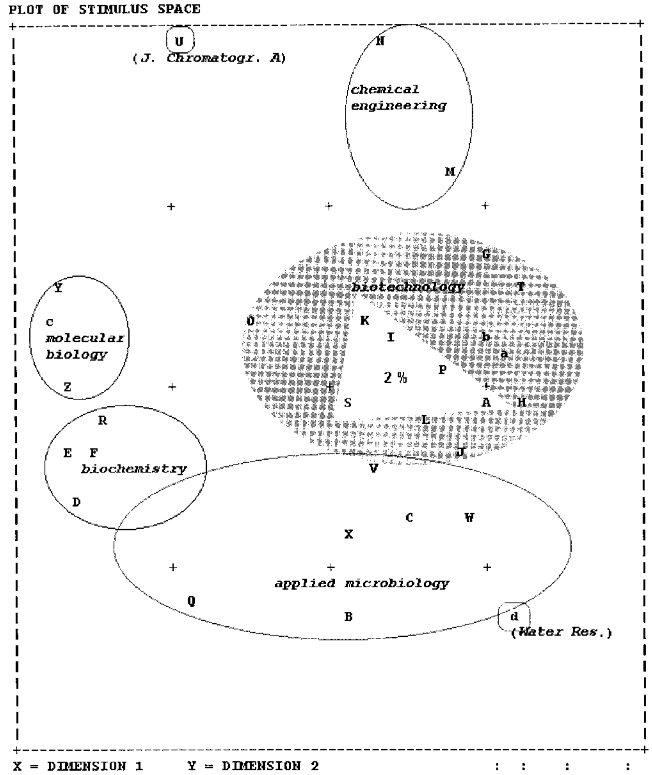
On the basis of previous research (Leydesdorff & Gauthier, 1996), we conjectured that *Biotechnology and Bioengineering* could be considered as the leading journal in this field. Using this seed journal as an entrance to the *Journal Citation Reports* of the *Science Citation Index* for 1996, a relevant journal environment was delineated by taking all journals into consideration that cited this journal or were being cited by it to the extent of one percent of its total citations. These aggregated citation relations were organized in a matrix that was then factor analyzed.

Both in the case of using a 1% and in the case of using a 2% cutoff rate, an eigenvector representing "biotechnology" in this network of citation relations can clearly be distinguished. Figure 1 provides a mapping of the relevant journal environment in 1996. As expected, "biotechnology" is located between the fields of "applied microbiology," "biochemistry/molecular biology," and "chemical engineering." Furthermore, the journals *Water Research* and the *Journal of Chromatography A* are present as isolates in this environment.

The smaller subset relating to *Biotechnology and Bioengineering* at only the 2% level of total citations is indicated by using boldfaced characters both within the map and in the legend. We will work in the remainder of this study with this core set. *Biotechnology and Bioengineering* not only has the highest impact factor of the set, it also has the highest factor loading on the factor indicating biotechnology. It can, therefore, be considered as a "central tendency journal" as defined by Cozzens and Leydesdorff (1993).[3] With hindsight, this legitimates our initial selection of *Biotechnology and Bioengineering* as the starting point for the analysis.

Delineation of the Domain in Terms of Documents

The five journals that had an aggregated citations relations above a 2% threshold to the core journal *Biotechnology and Bioengineering*,[4] contained 1,023 documents in 1996.[5] Of these documents, 711 contain a European, Japanese, or American address.[6] Because of international coauthorship relations in these documents, there is a total of 787 country addresses. We shall use these 787 records as the basis for the institutional analysis.



(The five core journals included in the analysis (threshold = 2%) are boldfaced in the list and indicated in the picture.)

	journal name abbreviation	factor designation
A.	**APPL BIOCHEM BIOTECH**	**I. biotechnology**
B.	APPL ENVIRON MICROB	II. applied microbiology
C.	APPL MICROBIOL BIOT	II. applied microbiology
D.	BIO-TECHNOL	III. biochemistry
E.	BIOCHEMISTRY-US	III. biochemistry
F.	BIOCHIM BIOPHYS ACTA	III. biochemistry
G.	BIOPROCESS ENG	I. biotechnology
H.	BIORESOURCE TECHNOL	I. biotechnology
I.	**BIOTECHNOL BIOENG**	**I. biotechnology**
J.	BIOTECHNOL LETT	I. biotechnology
K.	**BIOTECHNOL PROGR**	**I. biotechnology**
L.	BIOTECHNOL TECH	I. biotechnology
M.	CHEM ENG J	IV. chemical engineering
N.	CHEM ENG SCI	IV. chemical engineering
O.	CYTOTECHNOLOGY	I. biotechnology
P.	**ENZYME MICROB TECH**	**I. biotechnology**
Q.	J BACTERIOL	II. applied microbiology
R.	J BIOL CHEM	III. biochemistry
S.	**J BIOTECHNOL**	**I. biotechnology**
T.	J CHEM TECHNOL BIOT	I. biotechnology
U.	J CHROMATOGR A	VII. chromatography
V.	J FERMENT BIOENG	II. applied microbiology
W.	J IND MICROBIOL	II. applied microbiology
X.	J SCI IND RES INDIA	II. applied microbiology
Y.	NATURE	III. mol. biology/ general
Z.	P NATL ACAD SCI USA	III. mol. biology/ general
a.	PROCESS BIOCHEM	I. biotechnology
b.	REV CHEM ENG	I. biotechnology
c.	SCIENCE	III. mol. biology/ general
d.	WATER RES	VI. water research

FIG. 1. Factor-analysis and MD-SCAL for the citation patterns of *Biotechnology and Bioengineering* (1996; threshold = 100%).

[3] "Citing" can be considered as the evolutionary operator in the present, while citedness refers to codification in the archive (Leydesdorff, 1993). In the "cited" dimension, the biotechnology grouping contains only six journals, of which *Biotechnology & Progress* is most representative. As indicated by its name, this journal reports results relevant to other disciplines. Given the applied character of the emerging field, however, the biotechnology journals draw ("citing") upon existing disciplines twice as much as the latter draw on this group in terms of citations.

[4] *Bioprocess Engineering* is not processed by the *Science Citation Index* in 1996, but the *Journal Citation Reports* indicate that it would also have qualified for the 2% threshold.

[5] The set is composed of 969 articles, 24 reviews, 21 editorials, 7 corrections, and 2 notes.

[6] This set contained 676 articles, 20 reviews, 12 editorials, two corrections, and one note.

The underlying document set (of 711 documents) contains 9,626 title words, of which 2,427 are unique. Of these words, 245 occur more than five times.[7] These 245 words (10% of 2,427) occur 3,531 times in these titles, that is, 36.7% of the total number of occurrences (9,626). However, nine documents in the set of 787 records with institutional addresses did not contain any of these 245 words, and were, therefore, not included in the analysis. Thus, a matrix of 778 cases and 245 variables was constructed in which each row can be read as a representation of the frequently used words and cooccurrences of words in the title of the corresponding document.

The words and their cooccurrences can be considered as the observable variation at the field level, whereas the domain delineation in terms of documents indicates the geographical origins of the articles. As the reader may easily see, a similar matrix can be constructed using, for example, the most frequently occurring citations in this document set. Citations, however, are codified as textual elements to a larger extent than words and co-words (Leydesdorff, 1989). By using words and their cooccurrences, one observes the intellectual space as represented in the respective domain in the widest of its ramifications (Leydesdorff & Wouters, 1999).

Multivariate Analysis of the Data Matrix

Using the idea of cooccurrences of words as indicators of intellectual organization, the matrix of words and documents can be factor analyzed with the words as variables. The factor analysis provides us with a representation of the structure of the network of words. Each word has a position in the multidimensional network of cooccurrences among the words. This position can be expressed in terms of factor loadings representing the coordinates of each point in this architecture.

Do the data points—representing words—occur in clouds, or more evenly scattered over the multidimensional space? This assessment is crucial for the interpretation. The corresponding problem at the operational level of the statistics is to determine how many groupings one can distinguish. What is the dimensionality of this space, and how can one provide it with an interpretation? In general: when the structure of the network is more pronounced, the choice of the dimensionality can increasingly be warranted. Because words and co-words can be provided with different meaning in different contexts, the expectation is that the structuration of networks of words remains relatively weak (Leydesdorff, 1997).

[7] Single character words, abbreviations, numbers, and the following words were removed from this list: an, and, at, by, for, from, in, it, of, on, onto, the, to. “Between,” “under,” and “with” were included in the analysis, because the use of these prepositions was sometimes specific. Furthermore, singulars and plurals of nouns and verbs were equated by removing the trailing “s” (Leydesdorff, 1995).

Along the other dimension of this data matrix, that is, the rows, one can analyze the cases in terms of the underlying structures using the institutional addresses. Because the institutional addresses are known *ex ante,* one is allowed to use discriminant analysis for testing the hypothesis that there is a correspondence between the repertoires and the geographically (or more generally, institutionally) delineated sets. The title words are then used as independent variables for predicting group membership, and one is able to assess the quality of this prediction. Which percentage of papers is correctly predicted to belong to the set(s) under study on the basis of specific combinations of words in their respective titles? Discriminant analysis is akin to factor analysis because both techniques are based on eigenvector-analysis and parametric statistics.

We shall pursue discriminant analysis below, first, for the three industrial blocks of the E.U., the United States, and Japan, and then for the 14 European member states (14, because there were no papers with an address from Luxembourg in the sample). In a next step, we combine the results of the factor analysis and the discriminant analysis to specify the coupling of the two distinguished operations in terms of self-organization theories.

Second-Order Analysis of the First-Order Results

The results of the factor analysis remain theoretical constructs. The so-called “eigen-structure” is an analytical expectation with respect to the structure in the data. However, the specification of these expectations enables us to proceed from the results of first-order data analysis to second-order theorizing. Although the multivariate analysis is based on mapping and the *induction* of inferences about structures on the basis of observable variations, the operation of different (hypothetical) structures upon each other cannot be observed directly. However, one is able to specify the ways in which the results of the data analysis are significant in distinguishing among the various *possible* configurations and then the results of the various first-order (static) analyses can be used for informing these second-order (dynamic) expectations.

For example, if the three blocks (E.U., United States, and Japan) proved to be *completely discrete* in terms of the combinations of words used, one would expect a match between a three-factor solution and the lists of words that are specific to each of these three regional blocks. However, we also anticipate interaction among the regional blocks. If the three blocks were *completely coupled* at the global level, one would expect, in addition to the specific word sets, a common (global) word set to be represented by a fourth factor. Thus, in this case a four factor solution would provide the best fit.

In practice, one expects a situation *between* complete coupling and complete discreteness, and then more than four factors can be expected, because the coupling may be different between the United States and Japan from the coupling between the United States and Europe. Having not

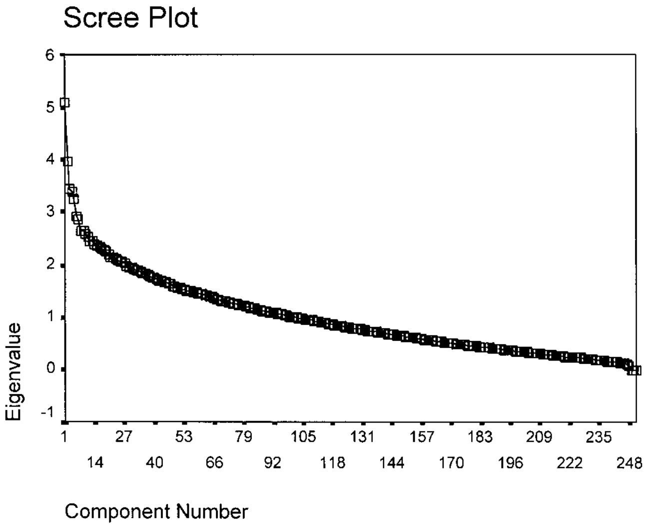


FIG. 2. Scree plot of the eigenvalues in the network of 245 words and their co-words (99 eigenvectors have an eigenvalue larger than 1).

yet provided the reader with the results of the factor and the discriminant analysis, the further specification may become rather abstract. Thus, we return to the methodological aspects involved in a later section after the discussion of the results of scientometric analysis of the data.

Results of the Scientometric Analysis

Results of the Factor Analysis

Exploratory factor analysis of the variables informed us that the cooccurrences of words were not specific enough to provide us with useful information about the eigenstructure of this matrix. Figure 2 exhibits the scree plot of the eigenvalues against the number of factors. This test indicates that 99 factors have an eigenvalue larger than unity and thus explain more of the variation than the average variable. This should be considered as a poor result. It means that the (245) words and their cooccurrences do not exhibit sufficient structure to legitimate any choice of dimensionality of a factor solution. In other words, the number of groupings of words remains an arbitrary choice of the analyst.

This problem of the arbitrariness of choosing a cluster level is well known in scientometric analysis of citation and cocitation data. Small et al. (Small & Sweeney, 1985; Small, Sweeney, & Greenlee, 1985) have proposed "variable level clustering" in the case of cocitation data. As noted, citations are an order of magnitude more specific than words. Citations are codified, whereas words can be provided with different meanings even within restricted document sets (Leydesdorff, 1989; cf. Amsterdamska & Leydesdorff, 1989). Thus, the delineations among clusters of words can be expected to be even more problematic than choosing the level of cocitation clustering (Leydesdorff, 1992b).

In summary, we used a highly restricted domain of a core set of five journals in the field of biotechnology. Nevertheless, the multivariate analysis of the (co-)word network did not provide us with reliable information about the intellectual organization of the (biotechnology) network. The failure to specify inductively the number of relevant dimensions further legitimates a deductive approach to this uncertainty based on second-order theorizing (below). We will then focus on the first 10 factors that have eigenvalues larger than 2.5.

Results of the Discriminant Analysis

Using the attribution of the 778 American, Japanese, and European cases as a grouping variable, a fit of 77.6% correctly classified papers is obtained when using discriminant analysis (Table 1). Using simulation techniques, Van den Besselaar and Heimeriks (2000, pp. 89–93) found that this level of discrimination itself cannot be considered statistically significant. The uncertainty in the overlap (23.4% or 174 papers) indicates interaction at the global level. However, this classification technique enables us to distinguish among three sets of papers for which the institutional addresses (in the E.U., the United States, or Japan) were correctly predicted ($N = 604$; see Table 1).

In other words, whereas the grouping among the words could not be considered statistically significant (above), the weak structure in the network of words does also not significantly correlate with the geographical division. Thus, one has no *a priori* grounds to assume any correlation

TABLE 1. Classification results of discriminant analysis of European (1), American (2), and Japanese (3) addresses.[a]

			Predicted group membership			
		GV	1	2	3	Total
Original	Count	1	305	76	13	394
		2	58	247	19	324
		3	4	4	52	60
	%	1	77,4	19,3	3,3	100,0
		2	17,9	76,2	5,9	100,0
		3	6,7	6,7	86,7	100,0

[a] 77,6% of original grouped cases correctly classified.

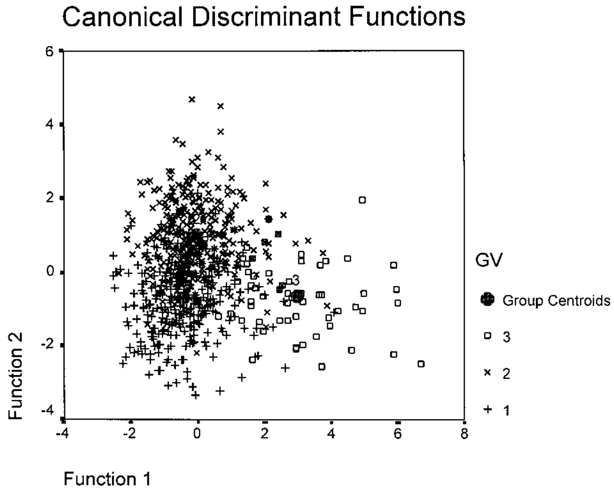


FIG. 3. Scatter plot of groups of papers with European (+), American (×) and Japanese (□) address.

between the word distributions in these three sets of papers and the word structure in the global set. The statistical analyses mainly suggest the lack of clear structures. The three sets of correctly classified papers, however, provide us with specific representations in terms of occurrences of words and co-words in the titles of these sets, respectively.

Using a geometrical metaphor one may envisage partly overlapping projections of the multivariate space of words and co-words on the respective geographies. The all-group scatter plot exhibits the relative differentiation of the three groupings (Fig. 3). In summary, the words and co-words in the ("biotechnology") document set did not reveal an inner logic that structures them (in terms of eigenvectors of their network of relations), but the geographical conditions provide an *ex ante* criterion enabling us to remove the misplaced records (as a remaining uncertainty).

A Second-Order Test for Self-Organization

Let us resume the analytical line of our argument. The words and co-words have been considered as the variation. On the one hand, this variation is conditional upon the initial selection of a core set of "biotechnology" journals. On the other hand, the variation is structured in terms of geographical addresses. Thus, two structures were hypothesized: a journal structure and a geographical structure provide us with specific selections. These two selections operate upon each other and the result is an observable variation. (One may also wish to express this as a Boolean AND in the retrieval.)

The variation is observable as the values in the cells of the matrix described above. The rows and columns of the matrix represent the selecting structures: the intellectual organization in terms of words versus the institutional organization in terms of document addresses. The covariation between the two axes can then be considered as a window through which these structures communicate. As far as the two distributions do not interact, they—by definition—condition each other.

How can one indicate the relation between the observable variation and the two hypothesized structures? As noted, the interaction between the two selections, that is, the observable variation was insufficiently structured for an inductive inference using multivariate analysis techniques. Let us turn to theorizing for a hypothetical specification of the interaction between the multivariate space of the intellectual organization and the geographical spaces of each of these three world regions.

If the three geographical regions were completely different, one would expect (as noted above) a representation of the intellectual domain in three specific groupings of words to provide an optimal fit with the representations in terms of words and co-words of these three regions. In Kauffman's (1993) terminology, one assumes three units ($N = 3$) and no links between them ($K = 0$) in this case. A three-factor solution should then provide a best match with the three separate word frequency lists using only the correctly classified cases.

At the other extreme, if the three regions ($N = 3$) were completely coupled, that is, maintain saturated relations with the other two units ($K = 2$), then there would be an international dimension in addition to these three geographical dimensions. A four-factor solution would be expected to provide us with a best fit in this case. The question then becomes the position of the global factor in relation to the geographically identifiable factors. Additionally, one can ask the question as to whether a single factor solution would also provide a good fit? In this latter case, the geographical dimensions would no longer play a role.

Finally, if the three regions are partially and differently coupled and/or in different dimensions ($N = 3$ and $0 < K < 2$), one can expect more factors than four. In addition to the geographical organization, the intellectual organization is then supposedly differentiated. Consequently, one would expect more complex patterns, perhaps enabling us to distinguish between a factor representing the American-Japanese relation and another indicating a dimension in the repertoire shared mainly among American and European scholars.

In operational terms, we can make a precise representation of each of the three regions in terms of the word-frequency distribution in the specific sets of papers which were correctly classified (above). The factor analysis of the whole set provides us with loadings for all the words on the various factors, and we can compare these factor loadings with the word frequency tables while using the option in SPSS to force different numbers of factors. Thus, one is able to develop a test of the above reasoning. The test is against the odds, because the first-order instruments of multivariate analysis were only weakly successful in making the relevant distinctions.

In summary, three situations can be distinguished:

(1) if there were a one-to-one correspondence between each specific region and the factor solution, the three-factor solution would provide us with the highest (and significant) correlation;
(2) if there were additionally an international dimension, the four-factor solution would exhibit the highest correlation;
(3) a number of factors larger than four would show the highest correlation if a more complex arrangement were expected.

The European Dimension in the Global Comparison

Figure 4 exhibits the results of the analysis using Pearson correlations between word frequency listings for the regional sets of documents and factor loadings for the respective number of factors being forced upon the factor analysis from 1 to 10.[8] First, we observe that the three regions share a pattern of a better fit in the case of a growing number of factors, that is, in relation to a more complex semantic structures of words and co-words. Remember, however, that we are reducing a structure of 245 words into a limited number (≤10) of latent dimensions.

The American word distribution in particular fits better if one assumes a more complex word pattern in the global distribution. The Japanese word distribution does not correlate significantly ($p < 0.01$) with any of the factors until more then eight factors are forced. (As expected, because of the above noted overlap, the Japanese, American, and European word distributions are significantly correlated among themselves.)

[8] We abstract from the sign of the correlation, because the factor loading may differ in sign, depending on previously extracted factors in the same analysis. Pearson correlation were used because factor analysis and discriminant analysis are both based on parametric statistics.

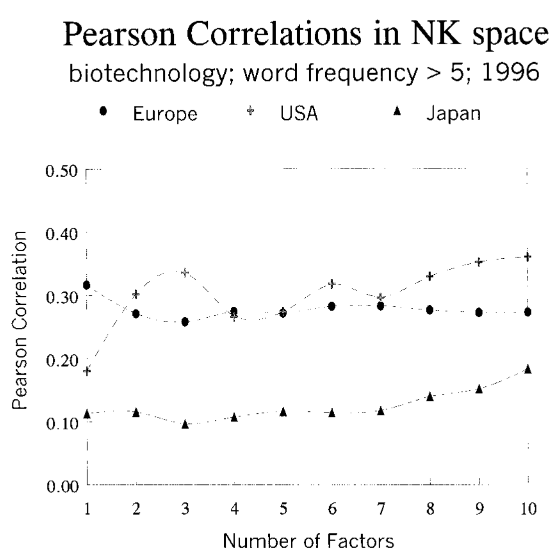


FIG. 4. Pearson correlations between Word Frequency List and Factor Loadings for a variable number of factors.

TABLE 2. Pearson correlations for two, three, and four factor solutions with the word distributions in sets of correctly classified documents.

	EU	USA	Japan
factor 1	0.2715**	0.1040	0.0856
factor 2	0.2149**	0.3022**	0.1157
factor 1	0.2585**	0.0352	0.0922
factor 2	0.2155**	0.3367**	0.0792
factor 3	0.1081	0.0789	0.0967
factor 1	0.2750**	0.0536	0.1079
factor 2	0.1888*	0.2668**	0.0567
factor 3	0.0265	−0.0243	0.0496
factor 4	0.1282	0.2659**	0.0913

Number of cases: 245; one-tailed signif: *0.01; **−0.001.

Both the American and the European set correlate significantly to the principal component in the case of a single factor solution. When more factors are forced, the solutions can be rotated, and therefore, the interpretation can be more meaningful. In the case of two factors, the European set correlates significantly with both factors one and two, but to a higher extent with factor one (see Table 2). Factor two correlates highest with the American repertoire.

This configuration of two main factors dominates the solution until more than six factors are allowed. The internal differentiation of the sets, which becomes significant when forcing four factors, gradually becomes more important than the differentiation between the European and the American sets. In summary, the European vocabulary correlates significantly with the vocabulary shared with the Americans, but this global vocabulary correlates more highly with the American one. The European vocabulary correlates additionally with a factor that has no significant correlation to the American repertoire.

In terms of our hypothesis, the results in Figure 4 suggest that the three-factor solution makes the American set somewhat more correlated, while this same configuration has a depressing effect on the European correlation. Remember that this three-factor solution corresponds with the hypothesis of a geographically contained repertoire. Thus, this closure is strongest in the American case. In the case of a four-factor solution, the correlations for the American and European sets are almost equal. As noted, these correlations are with two different factors (that is, factors one and two, respectively). Factor two can be considered as representation of the American repertoire, which also functions as the global one, while factor one is specifically European. The specificity of the latter repertoire of words and co-words turns it into a strong structural element at the global level.

In summary, the European vocabulary seems to be a bit more dominant in this set at the global level (because loading on the first factor), while the national embeddedness of the American vocabulary, which is also significantly shared among the two groups, becomes visible in the case of forcing three factors. The Japanese set couples only when the American set becomes further differentiated, that is, in

TABLE 3. Classification Results of document sets for 14 member states of the European Union (Luxembourg excluded).

			Predicted group membership														
		GVEU	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Original	Count	1	11	0	0	0	0	0	3	0	0	0	0	0	0	0	14
		2	0	8	0	2	0	0	0	1	0	0	0	0	0	0	11
		3	0	0	12	0	0	0	1	0	0	0	1	0	0	0	14
		4	0	1	2	74	0	0	4	0	0	0	1	0	0	0	82
		5	0	0	0	0	9	0	0	0	0	0	0	0	0	0	9
		6	0	1	0	0	0	44	3	0	0	0	0	0	0	1	49
		7	2	0	0	5	0	0	63	0	0	0	3	0	0	0	73
		8	0	0	0	0	0	0	0	6	0	0	0	0	0	0	6
		9	0	0	0	0	0	0	0	0	5	0	0	0	1	0	6
		10	0	0	0	0	0	0	0	0	0	13	0	1	0	0	14
		11	1	0	0	0	0	0	3	0	0	0	32	1	0	2	39
		12	0	0	0	0	0	0	0	0	0	0	0	15	0	0	15
		13	0	0	0	4	0	1	2	0	0	0	3	0	34	0	44
		14	0	0	1	0	0	0	1	0	0	0	0	0	1	15	18
		Ungrouped cases	22	11	10	48	39	44	47	13	18	30	22	33	29	18	384
	%	1	78,6	,0	,0	,0	,0	,0	21,4	,0	,0	,0	,0	,0	,0	,0	100,0
		2	,0	72,7	,0	18,2	,0	,0	,0	9,1	,0	,0	,0	,0	,0	,0	100,0
		3	,0	,0	85,7	,0	,0	,0	7,1	,0	,0	,0	7,1	,0	,0	,0	100,0
		4	,0	1,2	2,4	90,2	,0	,0	4,9	,0	,0	,0	1,2	,0	,0	,0	100,0
		5	,0	,0	,0	,0	100,0	,0	,0	,0	,0	,0	,0	,0	,0	,0	100,0
		6	,0	2,0	,0	,0	,0	89,8	6,1	,0	,0	,0	,0	,0	,0	2,0	100,0
		7	2,7	,0	,0	6,8	,0	,0	86,3	,0	,0	,0	4,1	,0	,0	,0	100,0
		8	,0	,0	,0	,0	,0	,0	,0	100,0	,0	,0	,0	,0	,0	,0	100,0
		9	,0	,0	,0	,0	,0	,0	,0	,0	83,3	,0	,0	,0	16,7	,0	100,0
		10	,0	,0	,0	,0	,0	,0	,0	,0	,0	92,9	,0	7,1	,0	,0	100,0
		11	2,6	,0	,0	,0	,0	,0	7,7	,0	,0	,0	82,1	2,6	,0	5,1	100,0
		12	,0	,0	,0	,0	,0	,0	,0	,0	,0	,0	,0	100,0	,0	,0	100,0
		13	,0	,0	,0	9,1	,0	2,3	4,5	,0	,0	,0	6,8	,0	77,3	,0	100,0
		14	,0	,0	5,6	,0	,0	,0	5,6	,0	,0	,0	,0	,0	5,6	83,3	100,0
		Ungrouped cases	5,7	2,9	2,6	12,5	10,2	11,5	12,2	3,4	4,7	7,8	5,7	8,6	7,6	4,7	100,0

[a] 86,5% of original grouped cases correctly classified.

1—Austria; 2—Belgium; 3—Denmark; 4—England, Scotland, Wales and Northern Ireland (UK); 5—Finland; 6—France; 7—Germany; 8—Greece; 9—Ireland; 10—Italy; 11—Netherlands; 12—Portugal; 13—Spain; 14—Sweden.

relation to more specific dimensions of the global repertoire.[9] In the case of solutions with a low number of dimensions, the representation of the Japanese repertoire remains submersed within the American one.

The European Decomposition

The European Union can also be considered as a collection of national systems which, as an aggregate, is able to have an impact at the international level more than as a network. For the purpose of this analysis, the subset of documents with a European address was decomposed in terms of the nation states that constitute the European Union. Because there were no papers with a Luxembourg address in this set, the $N = 14$.

[9] Whether these specific dimensions correspond to specialties in biotechnology can be investigated by decomposing the significant correlations using information theory (Theil, 1972; Leydesdorff, 1995).

The discrimination of papers using the national addresses as discriminating variables is correct in 86.5% of the cases. The classification table is provided in Table 3. Note that, for example, 21.3% (that is, three of the 14) papers with an Austrian address are predicted as of German origin. Belgian papers, on the other hand, are most difficult to predict in terms of the words used in their titles. The prediction is 100% correct for papers with a Finnish, Greek, or Portuguese address (cf. Van den Besselaar & Heimeriks, 2000).

The fit between the word distributions in the correctly classified national sets and the factor solutions can be made visible by repeating the analysis of the previous section, *mutatis mutandis*. The curves for the 14 European countries are shown in Figure 5a and b (for seven member states each). Upon visual inspection, no general patterns are discernable. The curves are mainly different. Some countries (e.g., Austria) exhibit a better match when a larger number of factors is taken into account, while others (like the U.K.) do better in the range of lower numbers of factors. Denmark has a high profile over the whole range, followed by the

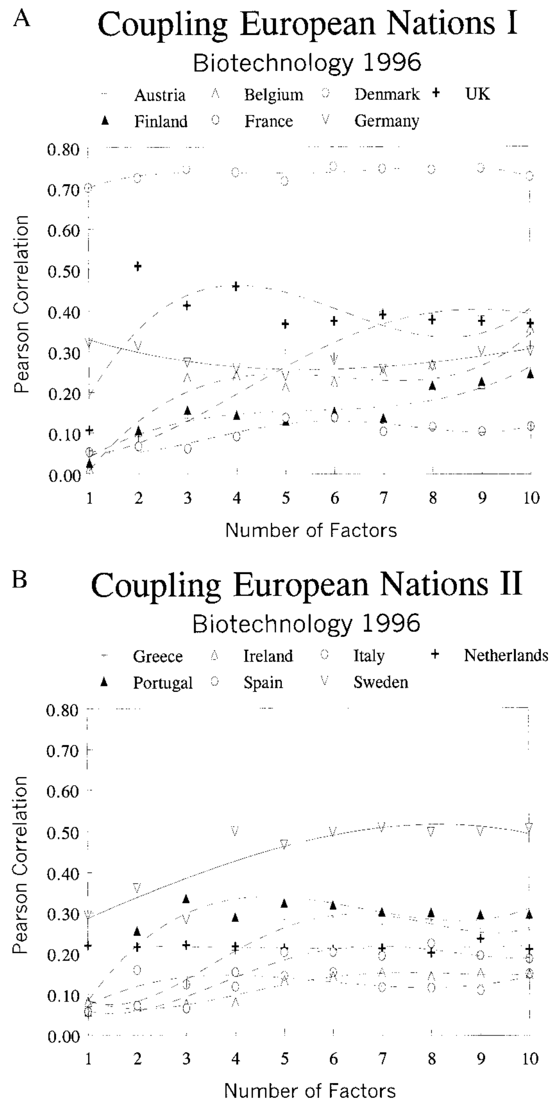


FIG. 5. Pearson correlations between Word Frequency Lists and Factor Loadings for seven European nations given different numbers of factors.

U.K. and Sweden. While all couplings with an $r > 0.2$ are statistically significant, the French set is never significantly coupled to the global repertoire.

There are strong arguments from the literature to assume that the European nation states are stable configurations (Skolnikoff, 1993; cf. Leydesdorff, 1992a).[10] In terms of the new evolutionary theorizing, one would formulate that the national systems occupy specific local optima (e.g., Kauffman, 1993). These local optima have been constructed historically. However, given the number of relations among these 14 European member states, the landscape may be rather flat at the network level despite differences in national patterns. The various countries are sometimes *positioned* similarly in relation to the rest of the world (notably, the United States).

For example, one can reasonably assume that from an American perspective, collaboration with a Danish or an Austrian partner does not make that much of a difference. Coauthorship relations with American colleagues may outnumber international coauthorship relations within Europe (Glänzel, 2001). This communality in position may help to reinforce the unification of Europe in addition to the development of internal relations (Burt, 1982). The external relations, however, are shaped at the level of individual authors and institutes, that is, not necessarily nation specific.

Our argument has been above that European development is entrained by the eigendynamics of the transient between two optima, that is, a global communication system shared with the Americans, and a differentiation within this system between the American and European repertoires. Although the European repertoire is significantly correlated to the factor solutions as an independent factor, its disaggregation into national repertoires exhibits mainly differences.[11] The self-organization of the European information society should therefore not be considered as a local, but as part of a global phenomenon. At this global level, the number of local optima for the system is limited,[12] while within the EU system various configurations can be stabilized.

In summary, there is systemness in the European repertoire, but the main component of the variation stems from the aggregation of and not from the interaction among the national repertoires. Europeanization can from this perspective be considered as a process of self-organization: an intermediate level is constructed that can be retrieved as relevant at the global level, but which itself is dominated by the different dynamics of its national subsystems.

Further Extensions

Without further testing, these interpretations remain based on a single measurement point, that is, "biotechnology" in 1996. Given this limitation, it seemed useful to extend the analysis using 1997 data. In a second extension, we will play with the threshold in the word frequency as one of the other parameters.

[10] Using Kauffman's (1993) general formula, the number of local optima is $\{2^N/(N+1)\}$ for $K = N - 1$. A system with $N = 14$ and $K = 13$ would then have more than 1,000 local optima. In this domain, however, K is probably smaller. For example, in Table 2 the number of off-diagonal cells with a value larger than zero is 28 out of 182 (that is, 15.4%), indicating a K of the order of two ($N = 14$, $K < 3$).

[11] The lower triangle of the correlation matrix among word occurrences in the national subsets contains (14 * 13)/2 = 91 cells. Only 42 of these correlations are significant at 0.01 level. Thus, the 14 word distributions for the national groupings are significantly correlated in less than 50% of the cases.

[12] Because $N = 3$, the number of local optima would be $\{2^3/(3 + 1)\} = 2$ if $K = 2$ (see footnote 10). As argued, K is between 2 and zero, and probably above 1. For theoretical reasons, the dimensionality of a system can be expected to remain fractional.

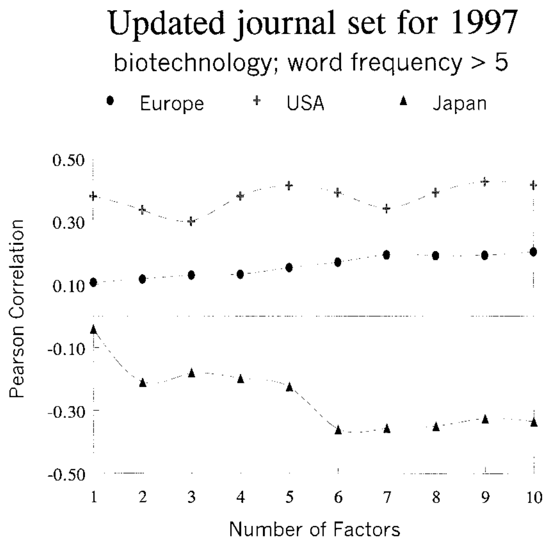


FIG. 6. Pearson correlations between Word Frequency Lists and Factor Loadings for different numbers of factors in 1997.

Comparison with the 1997 Data

The publication of the *Journal Citation Reports* for 1997 enabled us to update the analysis. We pursued the analysis both for the same journal set as used in 1996 and for the revised journal set, that is, based on a new journal–journal citation analysis. Compared with 1996, the *Journal of Chemical Technology and Biotechnology* maintained in 1997 an aggregated citation relation with the core journal (that is, *Biotechnology and Bioengeneering*) at the 2% threshold level, while the citation pattern of the *Journal of Biotechnology* was now sorted within the "applied microbiology" group using factor analysis. Thus, there is considerable development from year to year at this level of aggregated journal–journal citations.

Figure 6 exhibits a picture analogous to Figure 4 for the updated journal set. The American distribution is much more dominant in this case than in the previous one, and the main structure is the opposition between the American and Japanese sets. These two word sets correlate on the same factor, but with an opposite sign. The European set is significantly correlated only when six or more factors are forced, that is, after the internal differentiation of the American correlation pattern (which is reached at the five-factor solution).

When the unchanged journal set of 1996 is used, the differences are less pronounced. The European set remains significantly coupled to the solutions with three or more factors, but at a lower level than the Japanese set. The decomposition of the European set into the nation states shows less variation when using this unchanged journal set, while the variation remains *more* similar to the results for 1996 when using the updated journal set. In summary, the results suggest that the updated journal sets provide us with more information about the dynamics of the system than the analysis using a fixed journal set. This is in accordance with the evolutionary expectation, but it further complicates the analysis.

The remarkably more pronounced and different visibility of the Japanese factor in 1997 is not an artifact of the relative number of papers in the 1997 set in comparion to the 1996 set. Based on the updated 1997 journal set, 719 records containing addresses in the E.U., the United States, and Japan could be retrieved, of which 700 were included in the analysis (using 228 words that occured more than five times). Of these 700 records, 47 or 6.7% contained a Japanese address as against 60 out of 778 or 7.7% in the previous case. However, it would lead us away from the research question of this study to pursue the analysis of the variation in the coupling of these different Japanese document sets.

Variation in the Word Frequency Threshold

Another parameter that we are able to vary is the number of words included in the analysis. By setting a higher threshold for the selection in this dimension, the substantive differences are expected to become more pronounced, while the European level would hardly be affected if this network is mainly institutional. This expectation is confirmed by the results exhibited in Figure 7.

In this case, the analysis is based on the 149 words that occur more than nine times in the set (as opposed to 245 words occurring more than five times above). The curve for the European document set is only marginally affected by this change in threshold, while the American document set couples more strongly to this more restricted word set. This result suggest that the European level provides an institutional effect that feeds back on the underlying variation among the European national systems, but that it is less important in the cognitive dimension, because the coupling is *not* sensitive to a stronger selection of the (substantive) title words.

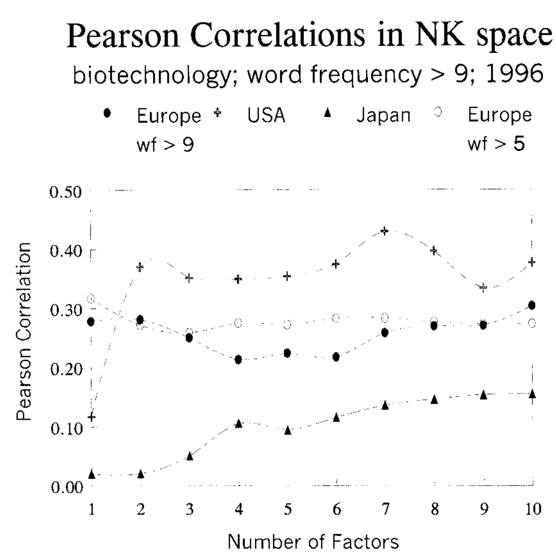


FIG. 7. Effect of word frequency threshold on the coupling of the American document set.

Conclusions

Scientific communications are extremely well archived, and therefore, we have a wealth of data at our disposal when we study the dynamics of the sciences and science-based technologies. Information retrieval has been developed particularly for studying scientific literatures, and elaborate methods were standardized by scientometricians during the 1970s and 1980s. In this study, we have tried to show how these data and methods can be used to address questions from a second-order systems perspective about the emergence and self-organization of the European information society.

The main thrust of our argument has been methodological. We could have used patent data instead of scientific publications (cf. Narin & Noma, 1985; Narin & Olivastro, 1992). However, licensing and patent practices are expected to differ among national innovation systems (Nelson, 1993). Furthermore, a European Patent Office is in place. Thus, one expects more regional bias in the organization of patent data than in the scientific literature. The development of a European repertoire significantly different from the American and Japanese ones at the level of global scientific communications, that is, within internationally established journals, seems to us a stronger argument for a process of self-organization at the European level as a (partially) unintended consequence of ongoing developments.

We have argued that Europe is different from the United States and Japan in that it is going through transformation processes at two levels at the same time, but without a necessary coupling between them. The contingency of the coupling allows for the process of autopoiesis, in principle. First, there is the European unification process that is stimulated particularly in the area of new sciences and technologies through the large European programs. These programs use a political definition of a theme. The theme of a program may refer to an emerging area of science and technology as in the case of biotechnology. The substantive definition of the field, however, is beyond the control of the European Commission.

Developments take place in a global market, and definitions are codified within the relevant communities of scientists and engineers who carry the system of publications as authors and coauthors. Although one would expect ongoing translations between the domains of policy making and scientific production, we did not find the observable variation to be significantly codified in terms of words and co-words. Yet, the two systems are supposed to disturb each other through these translations.

In the case of Europe, the translations are further complicated because of the different relations between national and supranational levels. From an evolutionary perspective, one would expect the two transformation processes to reinforce each other only after they have reached the basin of an attractor or, in other words, when the various dynamics have begun to resonate. Our results indicate that the various European national contributions mainly differ in terms of the couplings of their respective repertoires to the global developments. Nevertheless, a significant European interaction level could be retrieved.

This European repertoire seems to be fragile, because it is less pronounced in the 1997 data set than in the 1996 one. Our results suggest that the European integration is more institutional than cognitive, because it is less sensitive to raising the threshold of the words included than in the American case. These network dynamics, however, may still be a relevant context in the emergence of biotechnology as a Mode-2 science in the case of Europe, while in the case of the United States and Japan these developments are more firmly integrated within their respective national systems of innovation (Giesecke, 2000). Without further testing, these interpretations of the results remain speculative.

Let us finally return to the methodological contribution to theorizing about complex dynamics contained in this article. We have argued that self-organization theory is a second-order type of theorizing: first, we analyzed the observable data as variation in relation to (hypothetical) structures responsible for the selections. On the basis of the results of the first-order analyses of the data, we were able to test our ideas at the second-order level. However, the first-order explorations were carefully designed for this purpose.

The two types of analysis are of a different nature. Whereas no statistical significance could be retrieved inductively in the initial analyses of word patterns, these negative results did not prevent us from using word patterns in testing the second-order hypothesis. On the one hand, second-order theorizing has a more elusive character, because one proceeds on the basis of hypothetical "what if"-type of questions. Note that the uncertainties in using various instruments at two layers of analysis have to be multiplied ($P \cdot Q$), that is, not added ($P + Q$).

On the other hand, we have put our hypotheses to a strong test by using standard statistics based on parametric assumptions. However, scientometric distributions are highly skewed. Had we used nonparametric statistics (like probabilistic entropy; cf. Leydesdorff, 1995), we might have found more precise results enabling us to go further in the interpretation. The test of a European vocabulary in these title words, however, was significant, and the disappearance of this commonality when one pursues the analysis at the level of the collectivity of nation states was also significant.

Normative Implications

When a multilayered network structure is self-organizing its further development in terms of layers of recursive selections, the prediction of its further development may be counterintuitive. The emerging network system can be considered as moving on the transient towards an attractor. The point of control of the reorganization itself can then be at shifting places, because the development is expected to be controled in terms of the functionality for the momentum of this unintended development at a next-order systems (e.g., global) level.

The variable language of steering is then to be replaced with a language of fluxes and the geometrical representation of a control center is correspondingly replaced with an algorithmic simulation of different levels of control that may operate in parallel. The European dilemma of organizing European networks in the mode of "subsidiarity" to national initiatives or in a more federal mode provides policy makers with options to change their mode of operation over time and in different areas of policy making (Leydesdorff & Oomes, 1999).

If one assumes that "biotechnology" as a Mode-2 research system has incorporated its applicational contexts, the European system can be considered a fortiori as an amalgam of industrial, political, and scientific propulsions. Thus, one expects neither the science system nor the political system itself to be propelling, but the amalgamation with an industrial system in a Triple Helix or Mode-2 configuration (Leydesdorff & Etzkowitz, 1998; McKelvey, 1996, 1997). The internal differentiation of the political dimension in a national and a supranational level creates room for further recombinations without sacrificing existing advantages. Thus, Europeanization enriches the system with options for niche creation.

When the system of control itself can no longer be fixed, it can be allowed to fluctuate reflexively to stimulate the self-organization of the European information society. In such cases, a complex mixture of policy initiatives coordinated only loosely at various levels of aggregation may unintentionally enable research groups to make the adaptations required by their continuously changing environments.

Acknowledgments

The authors acknowledge TSER project SOE1-CT97-1060 entitled "The Self-Organization of the European Information Society." The project was coordinated with a project about "The Self-Organization of the Japanese Information Society (SOJIS)." We are grateful to colleagues within these projects for stimulating discussions.

References

Amsterdamska, O., & Leydesdorff, L. (1989). Citations: Indicators of significance. Scientometrics, 15, 449–471.

Anderson, J., Collins, P.M.D., Irvine, J., Isard, P.A., Martin, B.R., Narin, F., & Stevens, K. (1988). On-line approaches to measuring national scientific output—A cautionary tale. Science and Public Policy, 15, 153–161.

Burt, R.S. (1982). Toward a structural theory of action. New York: Academic Press.

Cozzens, S.E., & Leydesdorff, L. (1993). Journal systems as macro-indicators of structural change in the sciences. In: A.F.J. Van Raan, R.E. de Bruin, H.F. Moed, A.J. Nederhof, & R.W.J. Tijssen (Eds.), Science and technology in a policy context (pp. 219–233). Leiden: DSWO/Leiden University Press.

Dosi, G. (1982). Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. Research Policy, 11, 147–162.

Etzkowitz, H., & Leydesdorff, L. (Eds.). (1997). Universities in the global economy: A triple helix of university–industry–government relations. London: Cassell Academic.

Frenken, K. (2000). A complexity approach to innovation networks. The case of the aircraft industry (1909–1997). Research Policy, 29(2), 257–272.

Fujigaki, Y. (1998). Filling the gap between discussions on science and scientists' everyday activities: Applying the autopoiesis system theory to scientific knowledge. Social Science Information, 37(1), 5–22.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). The new production of knowledge: The dynamics of science and research in contemporary societies. London: Sage.

Giesecke, S. (2000). The contrasting roles of government in the development of the biotechnology industry in the U.S. and Germany. Research Policy, 29(2), 205–223.

Glänzel, W. (2001). National characteristics in international scientific co-authorship. Scientometrics, 51(1), 69–105.

Kauffman, S.A. (1993). Origins of order: Self-organization and selection in evolution. Oxford: Oxford University Press.

Larédo, P. (1997). Technological programs in the European union. In H. Etzkowitz & L. Leydesdorff (Eds.), Universities in the Global Economy: A Triple Helix of University-Industry-Government Relations (pp. 33–43). London: Cassell Academic.

Latour, B. (1987). Science in Action. Cambridge, MA: Harvard University Press.

Leydesdorff, L. (1989). Words and co-words as indicators of intellectual organization. Research Policym 18, 209–223.

Leydesdorff, L. (1992a). The impact of EC science policies on the transnational publication system. Technology Analysis and Strategic Management, 4, 279–298.

Leydesdorff, L. (1992b). A validation study of LEXIMAPPE. Scientometrics, 25, 295–312.

Leydesdorff, L. (1993). The impact of citation behaviour on citation structure. In A.F.J. Van Raan, R.E. de Bruin, H.F. Moed, A.J. Nederhof, & R.W.J. Tijssen (Eds.), Science and Technology in a Policy Context (pp. 289–300). Leiden: DSWO/Leiden University Press.

Leydesdorff, L. (1995). The challenge of scientometrics: The development, measurement, and self-organization of scientific communications. Leiden: DSWO Press, Leiden University; at http://www.upublish.com/books/leydesdorff-sci.htm.

Leydesdorff, L. (1997). Why words and co-words cannot map the development of the sciences. Journal of the American Society for Information Science, 48, 418–427.

Leydesdorff, L. (2001). A sociological theory of communication: The self-organization of the knowledge-based society. Parkland, FL: Universal Publishers, at http://www.upublish.com/books/leydesdorff.htm.

Leydesdorff, L., & Cozzens, S. (1993). The delineation of specialties in terms of journals using the dynamic journal set of the SCI. Scientometrics, 26, 133–154.

Leydesdorff, L., & Gauthier, É. (1996). The evaluation of national performance in selected priority areas using scientometric methods. Research Policy, 25, 431–450.

Leydesdorff, L., & Etzkowitz, H. (1998). The triple helix as a model for innovation studies. Science and Public Policy, 25(3), 195–203.

Leydesdorff, L., & Oomes, N. (1999). Is the European monetary system converging to integration? Social Science Information, 38(1), 57–86.

Leydesdorff, L., & Wouters, P. (1999). Between texts and contexts: Advances in theories of citation. Scientometrics, 44, 169–182.

Lundvall, B.-Å. (Ed.). (1992). National systems of innovation. London: Pinter.

Maturana, H.R., & Varela, F.J. (1980). Autopoiesis and cognition: The realization of the living. Dordrecht: Reidel.

Maturana, H.R., & Varela, F.J. (1984). The tree of knowledge. Boston: New Science Library.

McKelvey, M.D. (1996). Evolutionary innovations: The business of biotechnology. Oxford: Oxford University Press.

McKelvey, M.D. (1997). Emerging environments in biotechnology. In H. Etzkowitz & L. Leydesdorff (Eds.), Universities in the Global Economy: A Triple Helix of University-Industry-Government Relations (pp. 60–70). London: Cassell Academic.
Narin, F. (1976). Evaluative Bibliometrics. Cherry Hill, NJ: Computer Horizons, Inc.
Narin, F., & Noma, E. (1985). Is technology becoming science? Scientometrics, 7, 369–381.
Narin, F., & Olivastro, D. (1992). Status report: Linkages between technology and science. Research Policy, 21, 237–249.
Nederhof, A.J. (1988). Changes in publication patterns of biotechnologists: An evaluation of the impact of government stimulation programs in six industrial nations. Scientometrics, 14, 475–485.
Nelson, R.R. (Ed.). (1993). National innovation systems: A comparative study. New York: Oxford University Press.
OECD. (1988). Biotechnology and the changing role of government. Paris: OECD.
Skolnikoff, E.B. (1993). The elusive transformation: Science, technology and the evolution of international politics. Princeton, NJ.: Princeton University Press.
Small, H., & Sweeney, E. (1985). Clustering the science citation index using co-citations I. A comparison of methods. Scientometrics, 7, 391–409.
Small, H., Sweeney, E., & Greenlee, E. (1985). Clustering the science citation index using co-citations II. Mapping science. Scientometrics, 8, 321–340.
Theil, H. (1972). Statistical decomposition analysis. Amsterdam: North-Holland.
Van den Besselaar, P., & Heimeriks, G. (2000). Codification and self-organization in the European STI system. Final report to the European Commission. Amsterdam: University of Amsterdam.
Van den Besselaar, P., & Leydesdorff, L. (1996). Mapping change in scientific specialties: A scientometric reconstruction of the development of artificial intelligence. Journal of the American Society for Information Science, 47, 415–436.