ELSEVIER

Research Policy 30 (2001) 485–508

research policy

www.elsevier.nl/locate/econbase

Technological change and network dynamics Lessons from the pharmaceutical industry

L. Orsenigo [a], F. Pammolli [b,*], Massimo Riccaboni [c]

[a] *Department of Economics, Bocconi University, Milan, Italy*
[b] *DSAS, Faculty of Economics Richard M. Goodwin, University of Siena, Piazza San Francesco, 7, 53100 Sienna, Italy*
[c] *Sant'Anna School of Advanced Studies, Pisa, Italy*

Received 1 September 1999; received in revised form 19 November 1999; accepted 8 February 2000

Abstract

In this paper, we investigate how underlying relevant technological conditions induce distinguishable patterns of change in industry structure and evolution. A mapping is detected between the specific nature of problem decompositions and research techniques at the micro level of knowledge bases, and patterns of structural evolution at the macro level of the industry network. The graph-theoretic techniques we introduce map major technological discontinuities on changes observed at the level of dominant organization forms. They might have applications in other domains, whenever the identification of structural breaks and homological relationships between technological and industrial spaces are important issues.

Keywords: Pharmaceutical industry; Biotechnology industry; R&D; Technological change; Alliances

1. Introduction

Networks of collaborative relationships among firms and other institutions are widely recognized as an important organization form of innovative activities.

One can find in the literature widely different interpretations of the nature, motivations, structure and functions of networks, ranging from more sociologically oriented approaches to economic explanations based on (various mixes of) alternative theoretical backgrounds, e.g. transaction costs, contract theories, game theory, competence-based accounts of firms and organizations.

These interpretations generate widely different predictions about the evolution of collaborative relationships over time.

However, most of these approaches and explanations seem to agree in principle that, especially in high-growth, technology-intensive industries, networks of collaborative relationships should be considered and analyzed as organizational devices for the coordination of heterogeneous learning processes by agents characterized by different skills, competen-

* Corresponding author. Tel.: +39-0577-232-187; fax: +39-0577-232-188.
E-mail address: pammolli@unisi.it (F. Pammolli).

0048-7333/01/$ - see front matter
PII: S0048-7333(00)00094-9

cies, access to information and assets (see Pisano, 1991; Barley et al., 1992; Arora and Gambardella, 1994; Powell et al., 1996; Walker et al., 1997).

Beyond a rather generic agreement, though, the existing literature on networks does not address in detail the nature and specific properties of relevant knowledge bases and search activities to be used as explanatory constructs (see Dosi, 1982, 1988).

Against this background, this paper aims to establish a closer connection between the structure and evolution of scientific/technological knowledge and the structure and evolution of organization forms in innovative activities. More precisely, we deal with the relationships between some fundamental attributes of the evolution of relevant knowledge bases in pharmaceutical R&D and relevant properties of the structure and evolution of the industry network.

Our findings strongly suggest that a mapping is in place between the specific nature of problem decompositions and research techniques observed at the micro level of knowledge and technology dynamics and the patterns of structural evolution detected at the macro level of the industry network.

Our empirical analysis of network evolution relies on graph theoretical tools and measures to investigate an extensive data set that covers more than 5000 collaborative agreements among around 2000 firms/institutions from 1978 to 1997.

The mathematical language provided by the theory of directed graphs enables us to show how the nature and evolution of underlying relevant technological conditions induce distinguishable patterns of change at the level of industry structure and evolution. A set of indicators is developed, which turn out to be very useful to unravel the complex properties of empirical objects such diverse as technological and industrial structures.

The paper is organized as follows.

Section 2 briefly highlights the nature and goals of some fundamental research heuristics and techniques developed by firms and institutions in the last 20 years in their efforts to discover and develop new effective drugs. A fundamental distinction is captured, between co-specialized and transversal research technologies/strategies. That is, between research hypotheses and techniques that tend to be specific to particular domains and research techniques that are generic and, at the same time, complement co-specialized hypotheses and techniques in the course of research activities.

In Section 3, we highlight some implications of the nature of these heuristics and research strategies on the organization of innovative activities and on patterns of evolution of the network of R&D collaborative relationships.

In Section 4, we turn to the empirical analysis of the evolution of the network. Graph theory and numerical representations of networks are introduced, coming to show the existence of a striking homomorphic relationship with the structure and evolution of most recurrent research hypotheses and techniques used in problem solving activities. We refer to the notion of Canonical Decomposition of a graph in order to disentangle two major drivers/components of the structural evolution of the net, i.e., co-specialized and transversal actors that rely on co-specialized and transversal research techniques.

The presentation of the main findings and the discussion of some implications for the analysis of organization and industrial dynamics close the paper.

2. The growth of scientific and technological knowledge in pharmaceutical R&D

The last 25 years have witnessed a revolution in biological sciences, with significant basic advances in molecular biology, cell biology, biochemistry, protein and peptide chemistry, physiology, pharmacology, and other relevant scientific disciplines. The application of these new bodies of knowledge to pharmaceutical industry has had an enormous impact on the nature of R&D activities, on organizational capabilities required to introduce new drugs, and on patterns of industry evolution (see Galambos and Sturchio, 1996; Henderson et al., 1999).

In fact, the so-called "*molecularization*" of physiology, pathology and pharmacology, corresponds to a principle according to which for the development of new powerful and selective drugs search has to penetrate deeply into the human organism to unravel the biochemical interactions at the cellular, infra-cellular and, most importantly, molecular levels.

According to the molecular biology paradigm, the route to understanding of human organism (nature) is through the dissection of the system in its constituent parts, followed by the study of these parts. The properties of the whole — and hence its behavior — are the sum of the properties of the parts, while pathologies are analyzed in terms of specific alterations of the molecules that constitute the human organism. This philosophy has had profound effect on the methods of inquiry, leading scientists to pursue the pattern: "study: dissect, identify, classify, and dissect further" (Testa and Meyer, 1995, p. 6).

In this perspective, the development of new drugs rests on the ability to generate more fundamental (general) theories, which yield an increasingly "deeper" explanation of processes that take place at higher levels of organization of matter inside the human organism.

Notably, with reference to the range of possibilities for therapeutic intervention, the convergence at the level of scientific explanations generated by the progress of fundamental knowledge corresponds to the identification of longer and more complex chains of causal events. In fact, for almost all the more complex pathologies, the inner dynamics of knowledge has been leading to a proliferation of a priori hypotheses on plausible research trajectories. Whilst new scientific explanations and discoveries can lead to deeper knowledge and, moreover, more fundamental explanations of the nature of processes that happen in the human organism can focus search at a given level of analysis, the very same achievements generate new hierarchies of sub-hypotheses.

This dynamics creates a dilemma: by definition, theories that are more fundamental explain more; simultaneously, they multiply the number of points of entry for the discovery and the development of new therapeutic treatments.

In other words, the very process of convergence at the level of scientific explanations can lead to a process of divergence in research strategies generated along the hierarchy of increasingly specific sub-hypotheses, with an increase in the number of alternative routes for intervening in the disease process.

To put it differently, scientific progress certainly "simplifies" the search space, eliminating certain alternatives that are proven to be wrong (Nelson, 1959; Arrow, 1962; David et al., 1992). However, at the same time, scientific discoveries generate a "deformation" and an expansion of the research space, by suggesting new competing hierarchies of sub-hypotheses, as well as previously unconceivable opportunities of discovery. Moreover, many research techniques and biological targets tend to be typically characterized by high degrees of co-specialization. That is to say, research techniques tend to be relatively specific to particular fields of application. Thus, a proliferation is induced in the number of trajectories, techniques, and ex ante conceivable exploration strategies.

Moreover, technologies such as genomics, gene sequencing, transgenic animals, have started to supply the industry with a huge number of novel biological targets thought to be relevant to a vast array of diseases defined at the molecular level, and developing highly sensitive assays incorporating these targets.

The substantial growth of biological knowledge on the human organism at the cellular, molecular and genetic levels notwithstanding, the discovery and development of drugs has continued to be a lengthy, expensive and often unsuccessful process. Within this context, the increasing number of plausible targets has generated severe bottlenecks in the drug discovery process, due to the difficulty of quickly and cheaply analyzing function and disease relevance of newly discovered targets and matching related compounds (see Vos, 1991).

Against this background, one of increasing costs and bottlenecks, during the 1980s and 1990s new developments in solution phase and solid phase chemistries, high throughput screening technologies (HTS), information technologies, and combinatorial chemistry have led to the development of a set of research technologies that allow to achieve a higher breadth of applications, measured in terms of the number of disease areas and biological targets to which the technology may be applied.

In particular, while several thousand genetic targets could not have been addressed with the methods of conventional medicinal chemistry, the development of combinatorial chemistry libraries, together with new techniques for high-throughput screening and ever-improving bio-informatics tools, has gradually made it possible to test a large number of

potential drug targets against an even larger number of chemical entities.[1]

More generally, during the 1990s, a set of generic research technologies has been developed: from PCR, to protein structure modeling, rapid computer based drug assay and testing, recombinant chemistry techniques, drug delivery systems, chemical separation and purification techniques, that allow researchers to screen thousands of potentially promising compounds.

In short, the recent evolution of research strategies and heuristics in pharmaceutical R&D can be characterized by discerning between two main search regimes, that have started to coexist within the industry. The first regime is essentially based on biological hypotheses and molecules that tend to be specific to given fields of application (co-specialized technologies), while the second regime is characterized by the emergence of new generic tools (transversal technologies).

In the case of co-specialized research hypotheses and molecules, the characterization of biological targets and the corresponding design/experimentation of each new drug tend to require individual analysis. Lessons learned from the design and experimentation of one biological hypothesis/molecule cannot be immediately transferred to other biological domains, in order to develop other classes of drugs. Conversely, transversal technologies are in principle applicable to multiple biological targets and diseases.

However, pharmaceutical R&D "deals with a system — the human body — far more complicated than any mechanical or electronic system" (Gambardella, 1995, p. 16). For this reason, co-specialized hypotheses and transversal techniques stay coupled to each other, in the context of research projects and development activities carried out under conditions of strong uncertainty.[2]

3. From growth of knowledge to network dynamics

So far, we have identified some properties of the processes of scientific discovery underpinning research activities in the pharmaceutical industry. An extensive literature has documented some of the consequences that the advent of molecular biology has produced on the organization of innovative activities, both at the firm level and at the industry level (Orsenigo, 1989; Henderson, 1994; Gambardella, 1995; Mc Kelvey, 1995; Galambos and Sturchio, 1996). In particular, it has been emphasized that the emergence of a dense network of collaborative relationships among firms of different types and other research institutions has been a major feature of the recent evolution of the pharmaceutical industry (see Powell, 1996; Powell et al., 1996).

In this section, we examine in more detail if and how the specific properties of the processes of scientific discovery in molecular biology have been influencing patterns of evolution of the network of collaborative relationships. Our main claim is that these basic properties ought to be preserved in the dynamics of the network, if such a form of organization of innovative activities has (at least partly) to be understood as an adaptive response to the structural cognitive features of the dynamics of research activities. That is, if the specific properties of learning processes influence and constrain the possible forms of organization of innovative activities.

[1] Combinatorial chemistry enables rapid and systematic assembling of a variety of molecular entities, or building blocks, in many different combinations to create tens of thousands of diverse compounds that can be tested in drug discovery screening assays to identify potential lead compounds. Large libraries are available to be tested against both established and novel targets to yield potential lead compounds for new medicines. Such vast numbers of compounds have been introducing a substantial challenge to the drug discovery process and have created a need for faster and more efficient screening. High throughput screening (HTS) methods make it possible to screen vast populations of compounds via automated instrumentation: that is, complex workstations capable of performing several functions with the help of mechanical arms or simpler automated dilution devices.

[2] For example, new technologies including high throughput methods for sequencing genes, for monitoring and comparing their expression in different situations, and following their inheritance in families prone to particular diseases, depend crucially on the integration of molecular biology with robotics, and analytical instrumentation. The integration of these disciplines has started to provide powerful capabilities for generating and analyzing large volumes of data about genes and their expression, making it possible for the first time to mount a systematic search effort to discover and characterize the genes and biochemical pathways which underlie human diseases.

Let us briefly summarize the basic properties of the dynamics of knowledge discussed in the previous section. First, a process of fast expansion of biological knowledge in the fields of biochemistry, physiology and pathology has been surging within the industry. Secondly, such growth of knowledge has taken the form of a specification process, in which a general hypothesis gives origin to a variety of sub-hypotheses that, in turn, develop other sub-hypotheses at lower levels of generality, and so on. Third, as a consequence, the structure of knowledge comes to have a distinct hierarchical nature (see Simon, 1969). Fourth, the overall process is highly cumulative, since it is based on a dynamics that introduces progressive specifications of biological hypotheses at each level of the hierarchy. Fifth, this dynamics of knowledge imposes a specific structure on the degree of stability of hypotheses. At higher levels of the hierarchy, hypotheses tend to stay relatively stable, since their falsification occurs over a relatively long time scale, being based on the falsification/selection of hypotheses at lower levels of generality. Sixth, during the 1990s the appearance of general purpose technologies for the production and screening of new molecular structures has introduced a new dimension in the evolution of the relevant knowledge bases.

According to our conjectures, these basic properties ought to be reflected in the network of collaborative relationships. We address only indirectly the question why collaborative agreements have become such an important form of organization of innovative activities. This would imply the specification of a fully fledged model of how cognitive structures influence organization forms. We advance some rather specific hypotheses on how the structure of the network should look like and treat the empirical evidence as a sort of reduced form of a well-specified structural model.

It is important to notice that the task of specifying the linkages between the properties of the dynamics of knowledge and the structural evolution of the network is somewhat facilitated by the very special nature of the pharmaceutical–biotechnology industry, as a strongly science-based sector. Differently from other industries or technologies, in this case scientific research has had (and continues to have) a direct and immediate relevance for innovative activities. The proliferation of new companies specialized in the production of new techniques and products directly derived by cutting edge academic scientific research and the development of a dense network of collaborative relations among firms are — as it is well known — prominent features of the industry.

In the following empirical analysis, a research hypothesis is associated to a specific R&D project embedded in a firm/institution. Every firm/institution is defined by the collection of its research projects, while agreements are conceived as organizational devices through which hypotheses/techniques are combined and in which an *Originator* can be distinguished from a *Developer* (see Appendix A).

On these bases, we can advance the following testable "predictions".

First, as projects correspond to research hypotheses/techniques, and provided that the latter proliferate over time, originated by an increasing number of firms, we would expect an expansion of the network over time. This growth may take place both through the entry of new firms and by means of an increase in the number of agreements between existing agents. Secondly, the hierarchical structure of growth of knowledge should result in a process of hierarchization of the network, with the emergence of a core of firms/institutions who are able to manage general hypotheses/projects. Third, given the cumulative nature of the growth of knowledge, earlier (later) entrants in the network should embody more general and stable (specific and unstable) hypotheses. Thus, we would expect to observe the development of a stable core in the network — composed mainly by earlier entrants — linking with an expanding turbulent fringe of later, more co-specialized, entrants. Fourth, this structure would be perturbed by the entry of new agents embodying either new "general" hypothesis, a wide portfolio of specialized techniques, or "transversal" techniques. In such a circumstance, one would observe a reduction of the degree of hierarchization of the network. In fact, these agents should be in principle able to link with many other actors and — in the case of transversal techniques — they would induce a shift in the profile of relationships between earlier and later entrants.

Please note that we are not making any assumptions about the role of firm size, degree of diversifi-

cation and propensity to enter into collaborative relationships. These are clearly important firms characteristics that ought to be controlled for and that might induce dynamic patterns in the network similar to those described above. We shall discuss these issues in the concluding section.

The importance of the technological determinants of the structural evolution of the network of collaborative agreements can be appreciated, at a first glance, by looking at Fig. 1. Fig. 1 is based on a 3D graphical representation of the network by means of level curves. Columns correspond to the x-axis (*Originators*), while rows to the y-axis (*Developers*). Levels $z(x, y) = b_{ij}$ indicate the cumulated number of agreements between firms i and j, classified according to year of entry into the network, with darker regions representing areas of higher relational intensity.

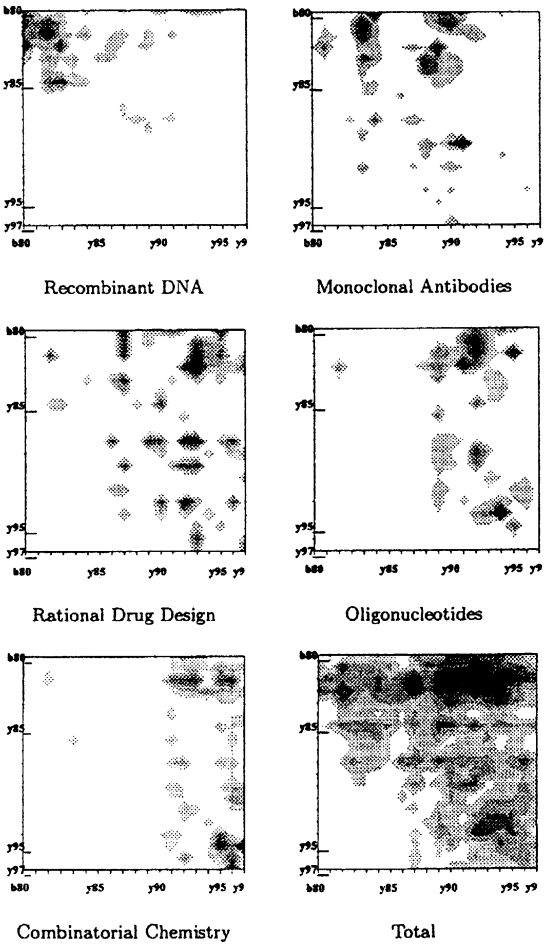


Fig. 1. Technological waves within the network.

Fig. 1 shows that: (i) *Originators* have entered the network by introducing successive waves of new research technologies, which shape the overall evolution of the network; (ii) Firms already active within the network have not played a major role as *Originators* in the new technological trajectories that have emerged after their entry; (iii) Rather, earlier entrants have gained access to the new technological trajectories mainly as *Developers*. (iv) As times goes by, the rate of entry in any given technological trajectory has been slowing down. That is to say, entrants have been closely linked to the generation of new technological trajectories.

All in all, evidences on patterns of entry, on relational roles of earlier and later entrants (*Originators/Developers*) and, finally, on new technological waves, suggest the existence of a dynamic process with the following properties. Major new technological breakthroughs initially induce the entry of new Firms/Institutions, which act as specialized technology *Originators*. As times goes by, *Developers* succeed in developing internal capabilities in the new fields. Correspondingly, relational intensity, as well as flows of entry, shift forward to new technologies and firms.

(v) After 1992, the emergence of *transversal technologies* like combinatorial chemistry has been perturbing the structure of the network. New entrants based on the new technological platforms and acting as *Originators* have been establishing relations with a large variety of Firms/Institutions, irrespective of age.

4. The evolution of the industry network

This section analyzes in detail the transformations occurred in the organization of innovative activities within the international pharmaceutical industry from 1978 to 1997.

Several graph theoretical measures are applied to investigate the evolution of the inter-organizational R&D activity that has characterized the pharmaceutical industry after the emergence of molecular biology.

The analysis is based on a unique data set built at the University of Siene by integrating several fonts. In particular, we merged a *proprietary database* on more than 14,000 pharmaceutical R&D projects with information about collaborative agreements drawn from a handful of well-known sector-specific databases (*Bioscan*, *Recombinant Capital*, *IBI*). Finally, we updated the resulting database by referring to annual reports (*SEC files*), and specialized press (*Scrip*, *Spectrum*).

The *collaborative agreement data set* considered for this paper covers 5056 agreements and 9785 research projects carried out by 2297 firms and institutions (*F/Is* from now on). Among them, 651 units have been classified as "Incumbent Firms" (*INC*: firms founded before 1973); 1372 units have been classified as "New Biotechnology Firms" (*NBF*: firms founded after 1973) and 274 units have been considered to be "Institutions" (*INST*: Universities, Hospitals, Public/Private Research Institutions). Mergers and Acquisitions have been taken into account by collapsing information relative to firms engaged in consolidation deals starting from the date of subscription.[3] As for collaborative agreements, the data set provides detailed information on typology, technological content, and date of signing.[4] Table 1 synthesizes the broad characteristics of the overall dataset.

Table 1
The collaborative agreement data set

Type of contracts		Technology	
License	3039	Miscellanea	958
Research	1359	Drug Delivery	650
Development	1641	Monoclonals	489
Equity	860	Screening	463
Collaboration	818	Recombinant DNA	405
Supply	453	Synthetics	364
Option	445	Oligonucleotides	348
Distribution	388	Combinatorial Chem.	217
Marketing/Promotion	326	Gene Sequencing	207
M and A	321	Gene Expression	193
Joint Venture	226	Rational Drug Design	127
Asset Purchase	186	Transcription Factors	107
Manufacturing	169	Cell Therapy S.C.F.	103
Warrant	108	Phototherapy	36
Loan	93	No Information	389
n.a.	26	Total	5056

Starting from the complete database, the subset consisting only of the R&D agreements has been selected. A total of 3973 agreements signed by 1709 F/Is have been extracted. The *R&D agreement data* set contains information on 349 INCs, 1112 NBFs and 248 INSTs.

Table 2 classifies agreements according to stage of signing. Interestingly, more than 88% out of the total number of collaborations were subscribed before the starting of the development stage. Furthermore, more than 76% of the total number of R&D agreements include a licensing contract.

In Appendix A the network of R&D collaborative agreements is rigorously defined in graph theoretical terms, and the formal apparatus required for the analysis of its structural evolution is highlighted. In particular, the overall network is referred to as a digraph (Harary et al., 1975). More specifically, the digraph is identified according to a time *orientation*. That is to say, for any given R&D project, we distinguish the F/I that acts as the *Originator* (*o*) from the one that acts as the *Developer* (*d*). In addition, the digraph has been *ordered* on the basis of time of F/I entry within the network. To put it differently, each node (F/I) of the graph has been labeled by the date of signing of its first agreement.

Two distinct time dimensions have been identified: the first one is defined at a micro level (the distinction between project *Originator and* project *Developer*); the second is singled out at a macro level (the emergence of the overall industry network as a product of F/Is entry and new agreements).

In what follows, the digraph is analyzed, in order to explain its main structural properties in terms of determinants of structural inertia and persistence, and drivers of structural instability and change. To accomplish this goal, the following four major steps will be undertaken in the following sections.

(1) Some generic properties of the evolution of the graph are analyzed. In particular, we observe that the graph expands almost exponentially over time and that such growth is essentially driven by the

[3] It is worth nothing that M&A activities strongly contributed to the process of hierarchization of the net.

[4] Every agreement may include different contract typologies at the same time. The information on the technological content is available for every agreement.

Table 2
Classification of R&D agreements according to stage of signing

Phase	%
Discovery	47.08
Lead molecule	17.09
Formulation	15.89
Preclinical	8.49
Clinical I	3.72
Clinical II	4.74
Clinical III	2.99

entry of new firms/institutions, while the density of the graph slightly decreases.

(2) Some permanent structural properties of the digraph are identified. Despite the steady rate of growth of the overall network, we find high levels of structural stability, both in terms of degree of asymmetry, intransitiveness, and hierarchization. Moreover, the digraph is shown to be "time reverse", as time *order* and time *orientation* are inversely related.

(3) The degree and sources of structural instability within the graph are investigated. As a reference point (a sort of null hypothesis), we start hypothesizing a conservative process being in place. At any point in time, such an inertial process would reproduce the same invariant structural properties. If such a process captured the dynamics of the network, one would have observed a smooth structural change, despite the intense growth of the network. In particular, given that the growth of the network is driven by flows of entry, structural inertia would be the effect of a cumulative, incremental technological dynamics. Moreover, given the time reversal phenomenon we mentioned above, it would be possible to locate the source of structural stability at the level of the process driving the entry of new *Originators*.

However, the empirical analysis carried out in order to test the structural inertia hypothesis has revealed two major sources of departure from such a conservative process. On the one hand, *a strong first mover advantage* is observed for firms that entered the network before 1981. On the other hand, some important *destructuring patterns* are identified for the years following the peak of entry of 1992.

(4) The departures from the structural inertia hypothesis are examined using the notion of Canonical Decomposition of a bipartite graph (Dulmage and Mendelsohn, 1958, 1959), which allows us to categorize F/Is according to the role they play in the dynamics of the network. Two groups of subjects are identified; a group of F/Is, which interact locally with given types of partners, and another group whose interactions are de-localized, i.e. are not restricted to a particular category of partners. What is even more interesting, is that F/Is belonging to any one of these two categories are immediately identifiable by the nature of the competencies they embody. The formers are active in those technological subfields that are recognized to be co-specialized, while the others are active in transversal technologies.

In synthesis, our empirical analysis reveals that major changes in the network structure take place in correspondence with major shifts occurring at the level of the underlying scientific and technological bases.

In order to identify such a relation, we have built an original formal apparatus for the representation of the structural evolution of a network of interacting agents.

4.1. Growth of the network and patterns of entry

For the period from 1985 to 1997, Fig. 2 shows the number of firms founded per year, the number of R&D projects started/ended per month and, finally, the 1-year moving average of monthly-subscribed R&D agreements. Over time, the number of ties grows approximately in proportion to the number of firms within the network. As a consequence, we observe a steady decrease in the density of the net, which moves from about 1% at the beginning of the 1980s to less than 0.15% in 1997. The analysis of

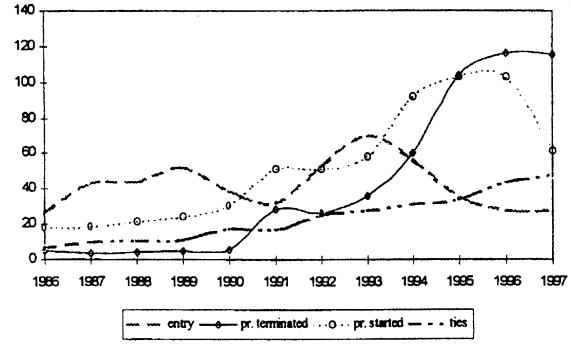


Fig. 2. Entrants, R&D projects, and R&D agreements.

patterns of firms' entry in pharmaceutical industry reveals the existence of two peaks in 1988 and 1992. *Both R&D projects and collaborative agreements are driven by flows of entry*, with an average time lag of, respectively, 2 and 3 years. It is worth noticing that the number of collaborative agreements parallels the number of R&D projects over the whole period until 1992. Starting from 1992, two different patterns are detectable. From 1992 to 1994, it is possible to observe a higher growth in the number of R&D projects as compared to that of agreements on the contrary, since 1994 an opposite pattern has started to be in place.

4.2. Structural properties of the graph

The network of agreements at time t is represented as a digraph $G_t(E,V)$, whose vertices V and edges E consist, respectively, of F/Is active in pharmaceutical research and development (V) and of R&D formal collaborations among them (E), drawn up by time t .The digraph G_t can be represented by an *adjacency matrix* $G_t \Leftrightarrow A(G)_t = [a_{do}]_t$. Matrix entry a_{do} is equal to 1 if an edge $e(d,o)$ does exist at time t, while a_{do} is equal to 0 otherwise. Matrix rows consist of all the vertices V_d (*Developers*), while matrix columns consist of all the vertices V_o (*Originators*). Thus, rows and columns vectors define, respectively, the sets of projects for which each F/I has acted respectively as an Originator and a Developer until time t.

dDegree(i,t) and oDegree(i,t) of vertex i at time t are given by the sums of matrix entries over row and column i. The total Degree(i,t) equals the sum of dDegree and oDegree.

As we have already pointed out, the set of vertices can be ordered according to time of entry into the network. Consequently, it is possible to permute the adjacency matrix in order to obtain a matrix $A(G)_{\leq t} = [a_{do}]_{\leq t}$, where $d \in \{1, \ldots, n\}$, $o \in \{1, \ldots, m\}$, with $\{\varepsilon(1) \leq \ldots \leq \varepsilon(d) \leq \ldots \leq \varepsilon(n) \leq t\}$, and $\{\varepsilon(1) \leq \ldots \leq \varepsilon(o) \leq \ldots \leq \varepsilon(m) \leq t\}$, where ε is the month of entry into the network.

Afterwards, it is possible to pass from $A(G)_{\leq t}$ to $A(G)_{\leq (t+1)}$ by adding rows and columns corresponding to F/Is entering the network at time $(t+1)$ and updating the entries of the new matrix according to latest agreements.

In the context of the present paper, we shall also use a more concise representation of the digraph structure at time t, by considering the block matrix $B(G)_{\leq t(\theta)}$ obtained by collapsing rows and columns of matrix $A(G)_{\leq t}$ that correspond to F/Is belonging to a common cohort of entrants defined by the time period $\theta = [t,\ t+\theta)$ (Generation). Entries b_{ij} of $B(G)_{\leq t(\theta)}$ indicate the total number of agreements between *Generations i* (*Developers*) and j (*Originators*) at time t (see Table 3).

The analysis of the structural properties of the digraph has led to the following results.

4.2.1. The digraph is asymmetric

For almost all relationships $e(d,o)$, $\varepsilon(d) < \varepsilon(o)$ i.e., the *Originator* usually entered the network after the *Developer* did. Early entrants have been acting mostly as *Developers*. Moreover, earlier generations of *Developers* have been establishing a large number of agreements with a large number of later entrants, which have been acting as *Originators*.

Data presented in Table 3 show that a large fraction of R&D agreements are associated with projects started by younger firms and/or research institutions, and then developed by older firms. In other words, the graph is characterized by a strong prevalence of inter-generation agreements over intra-generation agreements.

This result is confirmed by two tests carried out on block matrix $B(G)_{\leq t(\theta)}$, according to different values of θ.

The first test is the Conditional Symmetry Model (McCullagh, 1978; Everitt, 1977), applied to the

Table 3
Intergeneration and intrageneration R&D agreements

		O					
		INC	NBF_1	NBF_2	NBF_3	NBF_4	INST
D	INC	203	387	722	434	218	32
	NBF_1	12	23	55	25	14	17
	NBF_2	18	22	77	40	42	309
	NBF_3	13	10	41	38	35	246
	NBF_4	8	6	27	22	32	94
	INST	1	4	7	3	5	8

INC = Firms founded before 1973; NBF_1 = Firms founded between 1973 and 1981; NBF_2 = Firms founded between 1982 and 1986; NBF_3 = Firms founded between 1987 and 1991; NBF_4 = Firms founded after 1992; INST = Research Institutions.

ordered data matrices. According to the model, the null hypothesis is:

$$H_0: P(b_{ij}) = P(b_{ji}) \text{ for } i < j;$$

That is to say, $P(b_{ij})$, the probability of observing a given number of agreements between a generation i of *Developers* and j of *Originators* is equal to $P(b_{ji})$, i.e. there is no structural bias leading younger/older F/I act more frequently as Originators or as Developers.

According to the model, the ratio between the frequency of values above and below the main diagonal is set constant and equal to δ:

$$\delta = F_{ij}/F_{ji},\ i < j.$$

According to Agresti (1984), the estimators for the constant δ and the frequencies F_{ij}, F_{ji} are given by:

$$\hat{\delta} = \frac{\sum_{i \le j} b_{ij}}{\sum_{j \le i} b_{ji}}$$

$$E_{ij} = \frac{(b_{ij} + b_{ji})}{(\hat{\delta} + 1)}, \quad \text{if } i \ge j$$

$$E_{ij} = \frac{\hat{\delta}(b_{ij} + b_{ji})}{(\hat{\delta} + 1)}, \quad \text{if } i \le j$$

After running the model over our data for $B(G)_{\le t(12)}$, we found that $\delta = 1.8163$, the X^2 test being highly significant (p value < 0.01). This result confirms the insight gained by inspection of Table 3. In particular, one observes many more agreements between earlier generations of *Developers* and all the subsequent generations of *Originators*.

In addition to the Conditional Symmetry Model, a series of *Permutation Tests* (Tsuji, 1997) have been carried out on matrix $B(G)_{\le t(\theta)}$. According to the *Permutation Test*, the mean degree of asymmetry is measured by the expression:

$$D = \frac{\left(\sum_i \sum_j |\, b_{ij} - b_{ji}\,|\right)}{n} \quad \text{for } i \le j$$

where n is the number of blocks (generations) of the matrix. The original matrix $B(G)_{\le t(\theta)}$ undergoes a large number of random permutations and, each time, the mean degree of asymmetry, $D(p)$, is computed again. The fraction of permutations with $D(p) > D$ is always minor than 0.01. That is to say, the probability that the observed degree of asymmetry is purely random is very low.

In sum, the network of agreements is shown to be highly asymmetric. Moreover, the results of the *Permutation Test* reveal that the degree of asymmetry measured by the value of δ in the Conditional Symmetry Model is actually the outcome of the time order of the matrix and not of other possible ways of ordering the matrix itself.

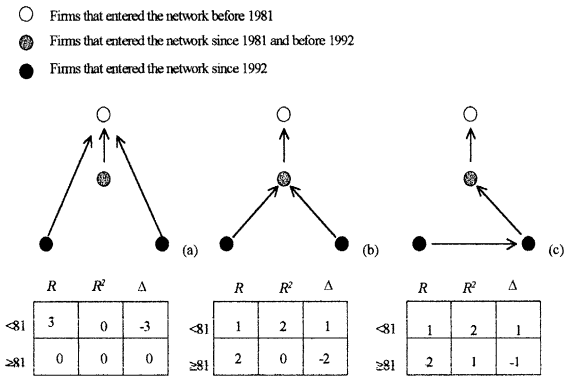
In a nutshell, the digraph can be said to be *time reverse*, as on average, time order and time orientation are inversely related.

4.2.2. The digraph is intransitive

A graph is transitive if it contains a relation $e(u,w)$ for every couple of edges $e(u,v)$ and $e(v,w)$. That is to say, the more each node can link indifferently with any other node in the network, the more a graph is transitive. Transitivity is essential in order to distinguish among alternative structural hypotheses, and various indices have been proposed for measuring it (Frank and Harary, 1982). In fact, intransitiveness implies some form of hierarchisation of the structure of the agreements over multiple levels (Hummon and Fararo, 1995).

In order to check for the existence of an high degree of intransitiveness, we first calculated the number of paths of length two (8666) present in our network. Paths of length one correspond to simple edges $u \to w$. Paths of length two (R^2) correspond to sequences of two agreements $u \to w \to v$. Then, we calculated the percentage of transitive triads upon the total number of paths of length two within the digraph (see Harary et al., 1975). In our data this percentage comes up to be very low, since it equals to 0.00018. This result is highly significant even after taking into account the low graph density ($d = 0.00136$),[5] unambiguously confirming that the digraph is significantly asymmetric.

[5] The ratio $\{F(e(u, w), e(u, v), e(v, w))\}/\{F(e(u, v), e(v, w))\ F(e(u, w))\}$ computed over every triad of vertices u,v,w, equals to about 0.13.



	R	R^2	Δ
<81	3	0	-3
≥81	0	0	0

	R	R^2	Δ
<81	1	2	1
≥81	2	0	-2

	R	R^2	Δ
<81	1	2	1
≥81	2	1	-1

Fig. 3. Alternative structures of the network.

4.2.3. The digraph has a hierarchical structure

We are now able to show that the observed degree of intransitiveness has to be interpreted as a result of the temporal structure of the network. To do that, we analyze the distributions of paths of length one (R) and two (R^2) according to the time of entry into the network. Specifically, we calculate the difference between the share of paths of length one and paths of length two ($\Delta = R^2\ (\%) - R\ (\%)$), respectively, for *Developers* that entered the network before and after 1981, and for *Originators* that entered the network before and after 1992.[6] Paths of length two identify a sequential structure where intermediate nodes exist, which have an agreement as *Developers* with one agent and an agreement as *Originators* with another ''terminal'' agent. Computation of the values of Δ allows us to identify the relevance of these intermediate nodes. To fix ideas, compare the structures shown in Fig. 3.

The first structure (a) is maximally hierarchical, with firms that entered the network before 1981 attracting all the agreements originated by younger generations. In the second structure (b), the upsurge of an intermediate layer is accounted for. As it is evident, the intermediate layer, is composed by firms that act as *Developers* in their linkages with younger generations and, at the same time, play as *Originators* with respect to the previous generation. Finally, the third elementary structure (c) represents a reduction of the overall degree of hierarchization of the net, driven by the emergence of intra-generation agreements.

Data presented in Table 4 show that the overall network appears to be very similar to the second benchmark structure until 1992. On the contrary, after 1992 it appears to be the result of the coexistence of structures of type (b) and (c).

In synthesis, not only the graph is intransitive, but it also has a distinct hierarchical structure, which is associated with the presence of different generations of firms, which play different roles within the network. Firms that entered the network before 1981 play a fundamental role in structuring it by linking as *Developers* to subsequent entrants. Later entrants perform a different role: they link both with older and younger generations, respectively as *Originators* and *Developers*.

Finally, however, it has to be noted that firms that have entered the net after 1992 have established a higher number of intragenerational agreements than firms of previous generations. As a consequence, a lower value of Δ is observed for agreements between orginators entered after 1992 and *Developers* entered after 1981.

4.3. The structural inertia hypothesis

We now move to unravel the nature of the generative processes underlying the evolution of the net over time. In order to test our null hypothesis of a conservative process going on, let's suppose that the degree $\mathrm{Deg}(i,t)$, that is the total number of agreements of $\mathrm{F/I}_i$ at time t, depends upon how long it has been present within the network, and on the number of potential partners that are active during the same period of time. In this case, $\mathrm{Deg}(i,t)$ may

Table 4
The value of Δ according to date of entry into the network (t_ε)

	O		
	Δ	$t_\varepsilon \le 1992$	$T_\varepsilon > 1992$
D	$t_\varepsilon \le 1981$	4.21	5.79
	$t_\varepsilon > 1981$	−8.28	−1.73

[6] We have done the same exercise using different dates. The years 1981 and 1992, however, seem to have a very important role in the the patterns of hierachization of the network.

be expressed as a function of a value t^*, that is a measure of time weighted by taking into account flows of entry. In practice, we clean the observed values of $\mathrm{Deg}(i,t)$ from the effects associated with differences in periods of presence within the network and number of potential partners at any given time. Since the digraph is time reverse, $d\mathrm{Deg}(i,t)$, the number of agreements as a Developer of $\mathrm{F/I}_i$ at time t, is distinguished from $o\mathrm{Deg}(i,t)$, the number of agreements as an Originator of $\mathrm{F/I}_i$ at time t. Then, for each F/I belonging to the same generation, two different t^* values, namely t_d^* and t_o^* have been calculated. More precisely:

$$t_d^* = \sum_{t=\bar{\varepsilon}}^{t} n(\varepsilon_o = t);$$

$$t_o^* = \sum_{t=1}^{\bar{\varepsilon}} n(\varepsilon_d = t).$$

where: $n(\varepsilon_o = t)$, $n(\varepsilon_d = t)$, are the number of firms entering the network as *Originators* and *Developers* at time t and τ is the last period of observation. In Fig. 4, $d\mathrm{Deg}(\bar{\varepsilon}, t_d^*)$. That is, degrees of *Developers* that entered the network in the same month $\bar{\varepsilon}$ are plotted as crosses, while the degrees of *Originators* $o\mathrm{Deg}(\bar{\varepsilon}, t_o^*)$ are plotted as triangles.

The analysis of $d\mathrm{Deg}(\bar{\varepsilon}, t_d^*)$ and $o\mathrm{Deg}(\bar{\varepsilon}, t_o^*)$ reveals two major deviations from a structural inertia hypothesis:

1. Since $\mathrm{Deg}(\bar{\varepsilon}, t^*) > \mathrm{Deg}(\tilde{\varepsilon}, t^*)$ for $\bar{\varepsilon} < 1981 \leq \tilde{\varepsilon}$, a persistent first mover advantage effect is present ($t = 1981$);
2. Since $d\mathrm{Deg}(\bar{\varepsilon}, t_d^*) > o\mathrm{Deg}(\bar{\varepsilon}, t_o)$ for $\bar{\varepsilon} > 1992$, *an inversion of the Developer/Originator* profile is detected after 1992 ($t_2 = 1992$).

In other words, after controlling for differences in time horizons and in the number of F/Is active whitin the network at any given point in time, earlier entrants tend to establish a larger number of agreements than later ones. Notably, the first mover advantage effect is stronger than it would have been under the conservative process hypothesis. Besides, firms which entered the net after 1992 have established more agreements as *Developers* than expected according to the hypothesis of a conservative growth process being in place.

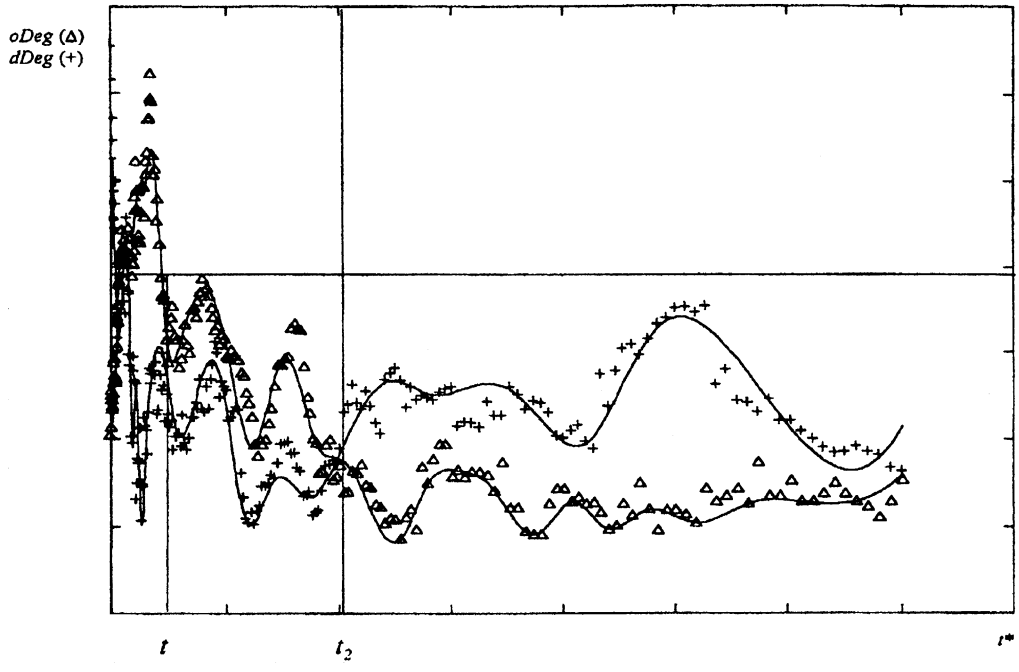


Fig. 4. *Originators* and *Developers* profiles.

On the whole, the results discussed so far on the structural properties of the graph in terms of patterns of growth, and degrees of hierarchization, asymmetry and intransitiveness in t, can be summarized by means of Fig. 5.

Thin arrows give a stylized representation of the structural inertia hypothesis. Bold arrows are meant to capture the violations to that hypothesis. Let summarize them. First, we observe a first mover advantage effect (vertical bold arrow). Second, a change is detected in the Developer/Originator profile after 1992 (horizontal bold arrows).

In Fig. 5, the orientation of the arrows reflects the time reversal phenomenon, i.e. the prevalence of inter-generation agreements over intra-generation agreements. *I* indicates firms that entered the network before 1981 and that benefit from a significant first mover advantage. *C* indicates firms that behave following the structural inertia hypothesis. *T* indicates firms that induce deviations from that pattern after 1992.

4.4. Departures from the structural inertia hypothesis

In this section, we analyze the nature and determinants of the relational roles played by firms/institutions that are Cospecialized and firms/institutions that are transversal within the graph (*C* and *T* in Fig. 5). First, we examine if F/Is of type *T* constitute an homogeneous group in terms of their relational profile. Secondly we advance and test the hypothesis that the major deviations in the structure of the network are related to the appearance of a new type of firms. Third, we show that these new firms embody what we called transversal technologies, which generate entirely different relational patterns than before. In synthesis, the observed structural changes in the graph are shown to be related to the emergence of a new class of research technologies.

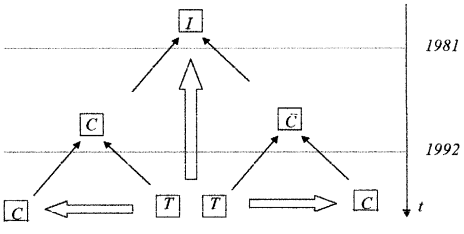


Fig. 5. Main structural properties of the network.

As we already know, after 1992 a new dynamic process starts to interact with the conservative process discussed earlier to generate the structure of the network.

To test this conjecture, we now try to identify the relational role that different generations of firms and different firms within the same generation play in the network at different points in time (every year). In other words, we ask whether the graph can be meaningfully decomposed in specific subgraphs containing firms and institutions which play unambiguous relational roles. To do that, we analyze the nature and origins of deviations from a matching condition at different points in time. More precisely, we try to couple unambiguously individual *Originators* to individual *Developers*. If each specific Developer were coupled to a specific Originator we would obtain a perfect matching. However, we may find some *Developers* that are not linked only to a specific set of *Originators*, but attract a large number of different *Developers* and lead to a hierarchization of the network. We call them Transversal *Developers* (*TransDev*). Similarly, we might observe *Originators* that make agreements with different agents. This would be the case of what we may call Transversal *Originators* (*TransOr*).

In order to identify firms that play different relational roles within the network, a Canonical Dulmage–Mendelsohn decomposition has been performed (see Appendix B). The digraph has been transformed into a bipartite graph and each node has been classified only either as a *Developer* or as an *Originator*.[7] Fig. 6 synthesizes the logic and results

[7] In a bipartite graph, the vertex set $V(G)$ is partitioned into two sets V_1 and V_2 in such a way that no two vertices in the same subset are adjacent. In particular, to represent the pharmaceutical R&D network as a bipartite graph, the vertex set V has been partitioned into two subsets D and O. As a vertex is forbidden to be included at the same time in partitions D and O, vertices v_d and v_o (F/Is that act respectively as *Developers* and *Originators*) have to be treated independently. As for F/Is which operate at the same time as *Developers* and as *Originators*, we consider for each of them two different vertices in set D and O, respectively. As a result, we are allowed to consider the bipartite graph $bG_{\Delta\tau}((O, D), E)$, which represents the agreements drawn up during a given period $\Delta\tau$ among *Developers* on the one side and *Originators* on the other.

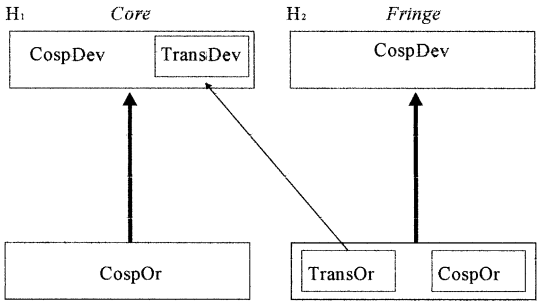


Fig. 6. Transversal and co-specialized nodes within the graph.

of the Dulmage–Mendelsohn decomposition. Boxes H_1, H_2 represent the two non-trivial subgraphs for which a matching can be found. In each box, we observe two subsets of *Developers* and *Originators*. Box H_1 contains the relational core of the network (approximately, the persistent relational component of the network: i.e. firms which have a large number of agreements and/or have entered the network early on), while box H_2 includes the relational fringe of the network. The matching in box H_1 captures the main structuring process of the network that we termed as the conservative process. However, we also identify a subset of *Developers* in box H_1 that link with a subset of *Originators* in box H_2. They correspond to what we defined above as Transversal *Developers* (*TransDev*) and Transversal *Originators* (*TransOr*).

The two sets of firms/institutions denoted as *TransDev* and *TransOr* can be thought of as the structural attractors of the network, i.e. they attract most of the agreements in each period of time (technically, they are present in all the intersections among minimum coverage vertex sets: see Appendix B).

TransDev and *TransOr* firms play a transversal role within the network, i.e. they cannot be assigned an unambiguous relational role. Transversal *Developers* (*TransDev*) establish several relationships with a wide variety of firms. On the other side, within the *Originators* group, a clear distinction can be drawn between a set of firms that are co-specialized in their relational behavior (*CospOr*), i.e. they are matched, and a set of firms that play a transversal role within the network (*TransOr*).

These results confirm that different kinds of relationships are present into the graph and hence that a

Table 5
First 20 Firms/Institutions by number of agreements according to: number and ranking of R&D projects, worldwide sales ranking, 1997

Network ranking	No. of ties	Firms and institutions	R&D projects	Sales rank
1	145	Novartis	$224_{(2)}$	3
2	141	Hoffmann-LaRoche	$112_{(12)}$	6
3	88	Smith Kline	$152_{(7)}$	9
4	81	Merck and Co	$207_{(4)}$	2
5	77	Bristol-Myers Squibb	$209_{(3)}$	4
6	74	American Home Products	$124_{(10)}$	8
7	69	Lilly	$138_{(8)}$	12
8	62	Abbott	$93_{(13)}$	18
9	60	Pfizer	$77_{(19)}$	7
10	52	Schering- Plough	$113_{(11)}$	15
11	51	Pharmacia and UpJohn	$174_{(6)}$	11
12	46	Glaxo Wellcome	$204_{(5)}$	1
13	45	Centocor	$22_{(101)}$	–
14	43	Genentech	$45_{(33)}$	–
15	41	Incyte	$10_{(257)}$	–
16	40	Bayer	$44_{(35)}$	16
17	39	Parke- Davis	$88_{(16)}$	–
18	37	Genetics Institute	$19_{(123)}$	–
19	36	NIH	$131_{(9)}$	–
20	34	Chiron	$64_{(24)}$	–

conservative process cannot represent its whole structural evolution.

It is now possible to demonstrate that the relational roles that have been identified correspond to firms embodying different types of technologies and that the changes over time in such roles correspond to the emergence of a new set of technologies, i.e. transversal technologies.

On the Developer side, the core of the network is persistently composed by a relatively small group of firms. Table 5 classifies firms according to date of foundation and presents information on the cumulative number of R&D ties, on number of ongoing R&D projects, and on ranking in terms of worldwide pharmaceutical sales in December 1997. For the group of actors that compose the core of the network a strong positive correlation between the number of R&D agreements, R&D projects and market sales is clearly observable.

As shown in Table 6, the set of firms playing a TransDev role is composed by the very same highly stable group of large R&D intensive pharmaceutical firms that entered the network early on and that have been playing a role of structural attractors during the whole history of bio-pharmaceutical industry. Moreover, those firms that started to act as TransDev since the beginning of the 1990s were already part of the core of the network in the previous years.

Table 6
First 15 firms active as TransDev, 1981–1997

TransDev Firms	Number of years
Hoffmann-La Roche[a]	7
Glaxo Wellcome[a]	6
Smith Kline[a]	6
Abbott[a]	5
Bayer[a]	4
Bristol-Myers Squibb[a]	4
Merck and Co.[a]	4
Pfizer[a]	4
Schering-Plough[a]	4
Ciba-Geigy/Novartis[a]	4
DuPont[a]	3
Hoechst Marion Roussel	3
Lilly[a]	3
Sandoz/Novartis[a]	3
Wyeth-Ayerst[a]	3

[a] Firms that were *Cosp Dev* already before 1992

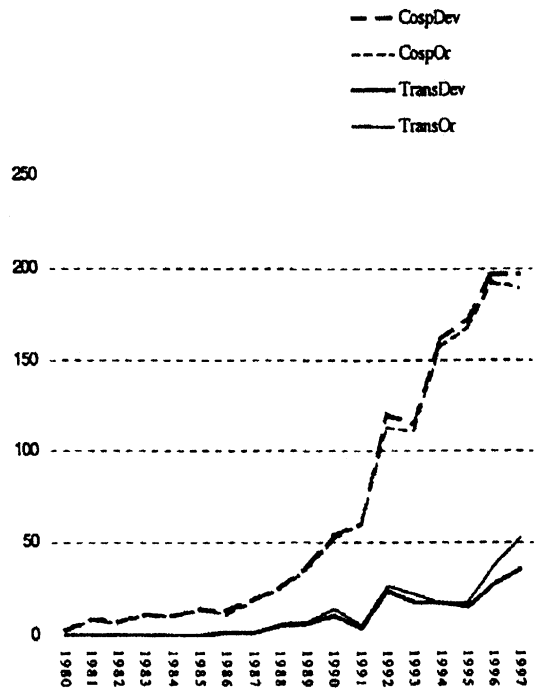


Fig. 7. Number of firms by relational category.

Fig. 7 plots the moving average of the number of firms classified according to relational categories in terms of co-specialization/transversality. It shows that a set of firms playing a transversal role within the network has taken off after 1992. At the same time, throughout the whole time period under observation, the number of firms that have been acting within the network as *CospOr* steadily increase. Correspondingly, from 1992 to 1997 the network has been characterized by the coexistence of both *CospOr* and *TransOr* firms.

On the Originator side, we already showed a correspondence existing between the emergence of transversal technologies and patterns of entry of new generations of *Originators*. We are now able to prove that technological transversality is a major determinant of relational transversality within the industry network.

More precisely, firms that have been identified as Transversal *Originators* into the graph by means of our analytical procedures embody Transversal Technologies.

Further information on the technological bases of relational transversality has been gained through a detailed analysis of the technological background of Transversal *Originators* based on personal interviews, information provided by 10K and 10Q SEC files reports, specialized press, and our proprietary data set at the University of Siena on R&D projects within the industry.

Transversal *Originators* are actually active in fields characterized by the presence of transversal research technologies, such as new drug delivery systems, combinatorial chemistry, genomics, genomic libraries, proteomics, highthroughput screening, and bioinformatics. In particular, Appendix C focuses on all most important firms that have been classified by the Algorithm as Transversal *Originators*. As it is possible to notice, these firms are active in fields characterized by the existence of general purpose platform technologies such as genomics, genomic libraries, proteomics and combinatorial chemistry, reporting R&D projects and agreements in the selected technological areas. Almost all the firms which were included in our R&D agreements data base and that are active on the basis of these platforms have been categorized as Transversal *Originators* by the Dulmage–Mendelsohn procedure.

The insight we have been gaining on the technological determinants of the changes in the structure of the network sheds additional light on the nature and determinants of the alrteady mentioned persistence by a core of established firms on the Developer side. On this, we have been reconstructing the dynamics of the TransDev component after 1992. As our analysis was able to detect, the core of the network initially expands, driven by flows of entry of new co-specialized firms and structured by the hierarchization of the network associated with the dominance of the molecular biology regime of cospecialized hypotheses and molecules. Until about 1992 the relational core of the net was populated mostly by early entrants. After 1992, the underlying technological discontinuities induced by the emergence of the new transversal technologies induce a significant turnover in the core of the network on the Developer side.

In other words, new transversal entrants have started to act as *Originators* not only in their relationships with early entrants, but also with young entrants lacking capabilities and knowledge bases in the fields of chemical diversity generation and screening. However, in the following years, established firms active as *Developers* have regained very quickly their structural role in the evolution of the industry network. In a nutshell, the entry of new Transversal *Originators* and the correspondent shift at the level of relational behaviors did not deeply modify the overall core–periphery profile of the industry network.

5. Concluding discussion

In this paper we have analyzed the structural evolution of the network of collaborative agreements in pharmaceutical R&D in the last 20 years. Our results reveal that some fundamental properties of the processes of growth of relevant knowledge bases are preserved in the structural evolution of the net.

Specifically, both the growth of knowledge and the structural evolution of the network have been characterized by fast expansion, proliferation of research trajectories and techniques, and hierarchization. The cumulative nature of such processes has been imposing different degrees of structural stability at different levels of the hierarchy. Finally, major changes in the network structure have occurred in correspondence with the emergence of a new set of transversal technologies.

We think that our results, while specific to the pharmaceutical industry, might bear interesting implications for a variety of both empirical and conceptual issues.

First, our findings may contribute to the broad debate on the nature and motivations of the network of alliances. Secondly, they can contribute to the analysis of the relationships between science and technology, public research and industrial R&D and the like. More generally, they may have some implications for theories which aim at explaining the forms of organization of innovative activities, patterns of division of labour and industrial dynamics, particularly those which emphasize the relevance of the notions of competencies, and dynamics capabilities of firms.

In synthesis, the main conclusion of this paper might be that the specific nature of technology and related learning processes matters in shaping (or, at least, in defining some boundaries to the possible) organization forms of R&D, patterns of division of labour and industrial dynamics.

In our view, the formation and subsequent evolution of the network of R&D alliances can be interpreted primarily as an adaptive response to the emergence of a radically new knowledge base within the industry, that is molecular biology. Scientific progress, however, did not only simplify the search space by providing more general theories. It also led to an expansion of the relevant search space, significantly deforming it. Firms — both large established companies and NBFs — could master at best only fragments of the relevant knowledge. The high rate of growth of knowledge, its evolution into increasingly specific and uncertain directions and — especially after 1992 — the appearance of transversal technologies, have led to the generation of a wide variety of approaches and lines of research.

These properties of relevant knowledge bases and related learning processes have induced particular patterns of division of labour between different types of firms. In general, our results indicate that two different logics of exploration and technological advance have been coexisting and complementing each other in the process of network evolution. The first avenue has been following a trajectory of increasing specification of biological hypotheses. The second has been progressing towards the development of transversal techniques to generate and screen compounds and molecules. The first trajectory has been generating patterns of division of labour in which older generations of firms have been working at higher levels of generality linking with successive generations of new entrants, who typically embodied increasingly specific hypotheses and techniques. The second trajectory has tended to alter this inter-generation structure. In synthesis, several mechanisms have influenced the patterns of division of labour dynamically interacting to produce quite complex structures.

In both cases, established R&D-intensive pharmaceutical firms have been able to absorb the new knowledge by interacting with new entrants. In fact, the expansion of the network has been driven mainly by the entry of new agents embodying new techniques. The network has taken a distinct hierarchical structure, with different firms operating at different levels of generality, which was perturbed but not broken by transversal techniques.

The above evidences support, in our view, two hypotheses already advanced in the literature, namely: (a) the cumulativeness of learning and competence building processes (see Henderson et al., 1999); (b) the significant capabilities by established multi-technology R&D intensive corporations to absorb new knowledge and techniques generated outside firms boundaries, despite major technological discontinuities and breakthroughs initially resulting in the growth of specialized technology producers (Cohen and Levinthal, 1989; Henderson, 1994; Granstrand et al., 1997).

The evidence presented in this paper suggests also that firms have found serious difficulties in modifying their structural position within the network. Put it in another way, specialist firms have tended to remain specialists, while early entrants have enjoyed significant first mover advantages, precisely because they have been able to embody knowledge at a high level of generality. Thus, a major asymmetry seems to have characterized the evolution of the network: while in many cases "generalist" firms have been able to (gradually) absorb increasingly specific knowledge (at least along particular trajectories of research), specialist firms found it much harder to move into the opposite direction.

First mover advantages, the asymmetry between "generalists" and specialists and — more broadly — the observed process of hierarchization of the network, may well be related to other "more traditional" variables, such as firms size, degrees of diversification, available resources, etc. In more general terms, one can legitimately wonder if the observed dynamics of the network is an "unconditional object", which might have been generated by processes and influenced by different variables than those emphasized in this paper.

Indeed, controlling for variables like firm size, diversification, propensity to make agreements, etc., constitutes an important part of our current research agenda. It is worth noting, however, that an explanation based on conventional firms features is not in contrast with our interpretation. Moreover, the results we get support the potential value of an ap-

proach that emphasizes the relevance of the specific properties of relevant knowledge bases, learning, and technologies.

Finally, this paper might have further implications from a more technical perspective. The graph-theoretic techniques we have used proved useful in mapping major technological discontinuities on changes observed at the level of dominant organization forms. They might have applications in other domains, whenever the identification of structural breaks and homological relationships between technological and industrial spaces are important issues.

Acknowledgements

The authors wish to thank Giovanni Dosi, Alfonsa Sambandella, Stanley Metcalfe, Keith Pavitt, Pier Paolo Saviotti, and paricipants to seminars held at Universidad Pompeu Fabra, Barcelona; Universidad Complutense, Madrid; SPRU, University of Sussex; INRA, Grenoble; Columbia University, New York; BETA, Strasbourg, for useful comments. Support from the Merck Foundation (EPRIS project), the European Commission, DG XII, and from the Italian Ministry of University and Scientific Research (MURST) is gratefully acknowledged.

Appendix A. Graphs and digraphs

In order to define the notion of a digraph, we have to introduce a definition of what a network (net) is. A net is generally defined (see Harary et al., 1975; Slepian, 1968; Diersel, 1997) by the following axiom system:

1. A finite and non-empty set V of elements v called "vertices";
2. A finite set E of elements e called "edges";
3. A function f whose domain is E and whose range is contained in V;
4. A function s whose domain is E and whose range is contained in V;

A digraph (oriented graph) is a net which does not include neither loops ($f(e) \neq s(e)\ \forall e \in E$) nor parallel edges ($f(e_i) = f(e_j)$ and $s(e_i) = s(e_j) \Rightarrow e_i = e_j\ \forall e_i,\ e_j \in E$).

Within the context of this paper, the structural proprieties of the network of R&D agreements are investigated by interpreting sets V, E and functions f, s in the following way:

1. V: The set of Firms/Institutions (F/Is) that have at least one R&D project in their pipelines. In our case each firm is associated with a set of projects. In other words, v should be thought as the set of projects of F/I, while V should be thought as the collection of the project sets corresponding to each F/I;
2. E: The pharmaceutical R&D projects included in the data set;
3. $o(e)$: F/I that started an R&D project e. In addition, v_o denotes the subset of v projects originated by each F/I;
4. $d(e)$: F/I that develop an R&D project e. In this case, v_d denotes the subset of projects v developed by each F/I.

As a consequence of the above definitions, every edge e within the graph is an oriented edge defined by a couple (o,d). As far as our empirical analysis of the network structure is concerned, we take into account only the subset of the R&D projects for which $o \neq d$. That is to say, only projects associated to two or more F/Is are considered (no self loops). Moreover, we treat multiple and repeated relationships among the same actors as a single edge (no parallel edges).

In order to study the dynamics of the digraph we define both a time orientation and a time order of the graph. As the development phase follows by definition the starting date of a project, maps o and d substantiate a *time orientation* of the graph either. What we need now is a time order defined on the vertices according to the year of entering into the network. Formally:

– $\tau(e(o))$: month in which project e is started by o;
– $\tau(e(o,d))$: month in which F/I d start to cooperate with firm o to develop project e. By definition $\tau(e(o,d)) \geq \tau(e(o))$;

– $\varepsilon_0(v)$:$\min_{v_o}\tau(e(o,d))$ the month in which v signs its first agreement as an *Originator*;
– $\varepsilon_d(v)$:$\min_{v_d}$:$\tau(e(o,d))$: the month in which v signs its first agreement as a *Developer*;
– $\varepsilon(v) = \min_v \tau(e(o,d)) = \min_v(\tau_o,\tau_d)$: date of entry within the network (the month in which v signs its first agreement).

In other words, with reference to the structural evolution of the network, a time ordering has been established according to both the year of foundation and the year of entry of any given F/I within the network. It is important to notice that both orderings are *complete*. On the contrary, the time-oriented graph generated by the distinction between *Originators* and *Developers* correspond, to a *partial* order set (see Asratian et al., 1998, Chap. 10) and in particular, to a time partially ordered set $\Gamma = (T, \vdash)$. According to ordered set theory, a non-empty subset $C = \{t_1, t_2, \ldots, t_k\} \subseteq T$ such that $t_1 \vdash t_2 \ldots \vdash t_k$ is called a chain. If $C = T$, the time order is *complete*. Moreover, two elements of T are said to be comparable if they appear together in the same chain C. Conversely, non-empty set of pairwise incomparable elements is called an antichain. Finally, the partition of Γ into disjoint time chains corresponds to a time decomposition of the network.

Appendix B. Dulmage–Mendelsohn decomposition

In order to identify the structure of the bipartite graph at different points in time ($bG_{\Delta t}$), a condensation procedure has been applied to the bipartite graph. This procedure generates a graph minor $bG_{\Delta t}[M]$ obtained by shrinking every strongly connected subgraph, replacing it with a vertex, and then substituting each set of parallel lines with single lines.

In the case of a bipartite graph, the concept of a strongly connected component is equivalent to that of a strong Hall component. Vertices in a Hall component are perfectly matchable, that is, there is a matching (a set of edges in which no two edges have a common end vertex) which covers every vertex within it (for further details, see Diersel, 1997).

The lines belonging to a matching are said to be admissible, while the remaining ones are called inadmissible. [8] Fig. 6 represents graphically the outcome of the analytical procedure described so far.

The application of a Canonical Dulmage–Mendelsohn decomposition algorithm (see Dulmage and Mendelsohn, 1958, 1959; Lovasz and Plummer, 1986, Ch. 4, p. 137) to the bipartite graph $bG_{\Delta t}$ produces the following results:

1. Two subgraphs (non-trivial) H_1, H_2, which are the connected components of the induced subgraph $bG_{\Delta t}[M]$;
2. H_1, H_2 are two elementary bipartite graphs;
3. Since the number of connected components of $bG_{\Delta t}[M]$ is greater than one, by permuting rows and columns the corresponding bi-adjacency matrix $A(bG)_{\Delta t}$can be put into the form:

$$\begin{bmatrix} A_1 & * \\ 0 & A_2 \end{bmatrix}$$

where matrices A_1, A_2 are the bi-adjacency matrices corresponding to the subgraphs H_1, H_2 while * represents the transversal ties between the two sub-matrices.

In the Canonical Dulmage–Mendelsohn decomposition we have applied, a major role is assigned to the interplay between maximum matching and minimum vertex cover. As transversal vertices are included in every minimum vertex covering of the graph, a greater proportion of such kind of vertices over the total number of *Originators* implies a higher number of time chains in which the graph can be decomposed.

[8] An edge e is inadmissible if and only if there exists a non_null minimum vertex covering — i.e., a covering consisting of as few elements as possible — $C_o \subseteq V_o$ of vertices in V_d (and vice versa) such that e belongs to that cover ($e \in E(G[C])$) (see Lovasz and Plummer, 1986; Asratian et al., 1998).

Appendix C. Transversal Originator firms (‘) out of firms active in combinatorial chemistry (cc), genomics (g), genomic libraries (gl), proteomics (p), and target based screening (tbs)

Originators	Found Year	Developers	R&D projects	R&Db Tech.
3 D Pharma (‘)	1993	BioCryst; Merck KGaA	DirectedDiversity ThermoFluor	cc,tbs;
Affymetrix (‘)	1991	Amersham Pharmacia Biotech; Beckman Coulter; Eos; Gemini Research; Gene Logic; Genetics Institute; Howard Hughes Medical Institute; Lilly; Merck and Co.; Novartis; Pfizer; Roche, Gene Logic, Human Genome Sciences	GeneChip, genomics library, gene expression, gene discovery, bacterial GeneChip probe arrays, inflammation, breast cancer, gene discovery, G-protein coupled receptor pathways, prostate cancer	g; gl
Alanex	1991	Aurora	Pharmacophore Directed Parallel Syn., ALANET	cc
AlphaGene	1993	Genetics Institute	FLEX, genetics libraries	g; gl
Ariad (‘)	1991	Genovo; Hoechst Marion Roussel; Harvard University; Stanford University	gene expression regulation technology	g
ArQule (‘)	1993	Abbott; ACADIA; Aurora; ICAgen; Monsanto; Ontogeny; R W Johnson Research Institute; Roche Bioscience; Sankyo; Scriptgen; Signal; Solvay; Wyeth Ayerst	Directed array, Mapping array	cc
Aurora (‘)	1995	Axys; Becton Dickinson; Warner Lambert, Allelix; Bristol-Myers Squibb; Cytovia; Lilly; Merck and Co; Roche Bioscience; SIDDCO	Genomics technology, Aurora screening technology, fluorescent	g; tbs
Axys (‘)	1997	Parke Davis; Pharmacia and Upjohn; Protein Design, Luminex; ZymoGenetics	New targets database, combinatorial chemistry, RAMMP, Liquid arrays, paracrine/endocrine signalling molecules gene database	cc, g; gl
Cadus	1992	Genome Therapeutics; SmithKline Beecham Massachusetts Institute of Technology	APEX; LIVING CHIP	tbs; g
Camb. Antib. (‘)	1990	ICOS; Progenitor; Wyeth Ayerst	ProxiMol, ProAb	g; gl
Cognetix	1993	Merck and Co; SIBIA	Combinatorial chemistry, conopeptide libraries	cc;
CombiChem (‘)	1994	ICOS; Ono; Roche Bioscience	DISCOVERY ENGINE	cc
CuraGen (‘)	1993	Biogen; Genentech; ArQule	Quantitative Expression Analysis; PathCalling database; Multiplex Interaction Assay GeneCalling database; SeqCalling; CombiGen	g, gl; p; tbs
Discovery Technologies	1993		Combinatorial chemistry, cyclin dependent kinases, receptor tyrosine kinases,	cc., tbs;
Dyax (‘)	1995	Affymax; Bristol-Myers Squibb; Burnham	phage display technology	cc

Originators	Found Year	Developers	R&D projects	R&Db Tech.
		Institute; Cambridge Antibody Technology; Chiron; Chugai; Corvas; Cytogen; DuPont; Genzyme; Merck and Co; MorphoSys; Pharmacia and Upjohn; Scios		
EnzyMed (‘)	1994	Merck and Co	BIOACTIV	cc.
Gene Logic (‘)	1995	Japan Tobacco; Organon	rEST database, Flow-Thru Chip, GENE EXPRESS, Multiplex Selection of Transcription Factors, READS, VIRIA	g;gl;tbs
Genetics Institute (‘)	1980	Bayer; Chiron; Chugai; Genentech; Immunex; Kirin Brewery; Ontogeny; Rhone-Poulenc Rorer; Sankyo; Scios	DiscoverEase	g
Genome Pharmaceuticals Pharmaceuticals	1992		Genomics technology, gene expression profiling, protein interaction mapping	g; p
Genome Therapeutics (‘)	1991	Bayer; Bristol-Myers Squibb; Hoechst Marion Roussel; Schering Plough	PathoGenome	gl; g
Genometrix	1993	GeneMedicine	DNA microarray technology	g
Genzyme (‘)	1981	Bayer, PaineWebber R&D Merck and Co Partners	Combinatorial chemistry, COMPILE, screening technology, p53/MDM2 interaction inhibitors, Solid Phase Epitope Recovery SPHERE	cc., tbs
Human Genome Sciences (‘)	1992	Genetic Therapy; Isis; Transgene; SmithKline Beecham; Merck KgaA; Synthelabo; Takeda	Human cDNA database, genomics technology, microbial genome database	g; gl
Hyseq (‘)	1992	Kirin Brewery; Perkin Elmer, University of California at San Francisco	HyChip, SEQUENCING BY HYBRIDIZATION, gene discovery, cardiovascular disease	g; gl
IGEN (‘)	1982	Abbott; Agouron; Amgen; Bristol-Myers Squibb; Peptide Therapeutics; Pfizer; Schering Plough; ZymoGenetics	ORIGEN	tbs
Incyte (‘)	1991	Lilly; Monsanto; Roche, Abbott; Pfizer; Schering; Schering Plough, Abbott; Ariad; AstraZeneca; BASF; Bristol-Myers Squibb; Genentech; Glaxo Wellcome; Hoechst Marion Roussel; Johnson and Johnson; Novartis; Novo Nordisk; Organon;	AureusGEM, CandidaGEM, LifeSeq Atlas, LifeSeq FL, mGSD-library, GSD-screen, PathoSeq.	g, gl tbs

Originators	Found Year	Developers	R&D projects	R&Db Tech.
		Pharmacia and Upjohn; Rhone-Poulenc Rorer; Roche; SmithKline Beecham, Scriptgen		
Irori	1995	Bristol-M. Squibb; Rhone-Poulenc Rorer	Combinatorial chemistry	cc.
Lexicon Genetics	1995	DuPont Merck; Genome Research Institute; ZymoGenetics	DOWNSTREAM GENE TRAPPING, LexGene, OmniBank	g
Lynx (‘)	1992	BASF; DuPont; Hoechst Marion Roussel	Massively parallel signature sequencing	g
Molecumetics	1992	Bristol-Myers Squibb	MolecuSet, SMART Library Technology	cc.
MorphoSys	1992	Chiron; DuPont	HuCAL, self-assembling multimeric and multivalent structures, trinucleotide-directed mutagenesis, selectively infective phage	cc; g; tbs
Myriad	1991	Bayer; Schering	Genomics tech., island hopping gene seq., ProNet	g, gl
NeXstar (‘)	1995	Fujisawa; Glaxo Wellcome	systematic evolution of ligands by exp. enrichment	cc
Novalon	1996	Boeh. Ing.; Genzyme; Millennium; SARCO	BIOMOLECULAR RECOGNITION SYSTEM	Tbs
Oxford Asymmetry (‘) Asymmetry (‘)	1992	Ares Serono; Bayer; BioChem Pharma; Pfizer; Vertex	combinatorial chemistry OmniBank	cc.
Panlabs (‘)	1970	Arena; Berlex; Bristol-M. Squibb; Genelabs; Geron; Karo Bio; Synthelabo; UCB, Tripos	OPTIVERSE	Cc
Peptide Th. (‘)	1986	Lilly; Novartis, Medeva	RAPiD, MolVaD	cc.; tbs
Pharmacopeia (‘)	1993	AstraZeneca; Bayer; Daiichi; Novartis; Organon; Schering	ECLiPS	cc
Progenitor (‘)	1992	Cambridge Antibody Technology	ATLAS, genomics technology, embryonic stem cell, growth factor receptor gene, yolk sac stem cell	g
Ribozyme (‘)	1992	Chiron; Glaxo Wellcome; Parke Davis; Roche Bioscience; Schering	Target Validation and Discovery, RNA editing	g; tbs
Sangamo BioSciences (‘)	1995	AstraZeneca; Bayer; DuPont; Japan Tob.; Millennium; Pfizer; S.K.B.; Targeted Gen.	Universal GeneTools	g
Scotia	1994	SuperGen	Combinatorial Lipids	cc
Scriptgen (‘)	1993	Boehringer Ingelheim; Lilly; Roche	ATLAS, SCAN	tbs
Sphinx (‘)	1987	Chugai; Kyowa Hakko; Taisho	combinatorial chemistry	cc
Synteni (‘)	1994	Geron; Monsanto; Schering Plough	GEM	g

Originators	Found Year	Developers	R&D projects	R&Db Tech.
Telik (‘)	1986	Sankyo; Scios; Sosei	TRAP	tbs
Trega (‘)	1990	Bristol-M. Squibb, Biogen; Isis; Parke Davis, Northwest Neurologic; Procter and Gamble	combinatorial biology, ChemFolio, Tea-Bag	cc
Tripos (‘)	1979	Bristol-M. Squibb, Arena Ph.; MDS Panlabs, Karo Bio, Menarini; Hoechst M. R.	ChemSpace, LeadQuest	cc
Xenometrix (‘)	1991	Affymetrix; Aurora; Cerep; Gene Logic; GeneTrace Systems; Genzyme; Incyte; PHASE-1 Molecular Technology; SmithKline Beecham	Genomic library, gene response profiles, human cell line-based	g; gl
Xenova (‘)	1986	Bristol-M. Squibb; Parke Davis; AstraZeneca	NatChem, QTC, ASSET	tbs, cc

References

Agresti, A., 1984. Analysis of Ordinal Categorical Data. Wiley, New York.

Arora, A., Gambardella, A., 1994. The changing technology of technical change: general and abstract knowledge and the division of innovative labor. Research Policy 23 (5), 523–532.

Arrow, K., 1962. Economic welfare and the allocation of resources for investment. In: Nelson, R.R. (Ed.), The Rate and Direction of Inventive Activities. Economic and Social Factors, NBER n. 13. Princeton Univ. Press, Princeton, pp. 609–625.

Asratian, A.S., Denley, T.M.J., Hoggkvist, R., 1998. Bipartite Graphs and Their Applications. Cambridge Univ. Press, Cambridge, MA.

Barley, S.R., Freeman, J., Hybels, R.C., 1992. Strategic alliances in commercial biotechnology. In: Nohria, N., Eccles, R.G. (Eds.), Networks and Organizations. Harvard Business School Press, Cambridge, MA, pp. 311–347.

Cohen, W., Levinthal, D., 1989. Innovation and learning: the two faces of R&D. The Economic Journal 99, 569–596.

David, P.A., Mowery, D., Steinmuller, W.E., 1992. Analyzing the economic payoffs from basic research. Economics of Innovation and New Technology 2, 73–90.

Diersel, R., 1997. Graph Theory. Springer-Verlag, Berlin.

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.

Dulmage, A.L., Mendelsohn, N.S., 1958. Some generalization of the problem of distinct representatives. Canadian Journal of Mathematics 10, 230–241.

Dulmage, A.L., Mendelsohn, N.S., 1959. A structure theory of bipartite graphs of finite exterior dimensions. Transactions of the Royal Society Canada Series 53, 1–13.

Everitt, B.S., 1977. The Analysis of Contingency Tables. Chapman & Hall, London.

Frank, O., Harary, F., 1982. Cluster inference by using transitivity indices in empirical graphs. Journal of the American Statistical Association 77 (380), 835–840.

Galambos, L., Sturchio, J., 1996. The pharmaceutical industry in the twentieth century: a reappraisal of the sources of innovation. History and Technology 13 (2), 83–100.

Gambardella, A., 1995. Science and Innovation. The U.S. Pharmaceutical Industry During the 1980s. Cambridge Univ. Press, Cambridge, UK.

Grandstrand, O., Patel, P., Pavitt, K., 1997. Multi-technology corporations: why they have “distributed” rather than “distinctive core” competencies. California Management Review 39 (4), 8–25.

Harary, F., Norman, R.Z., Cartwright, D., 1975. Structural Models: An Introduction to the Theory of Directed Graphs. John Wiley and Sons, New York.

Henderson, R., 1994. The evolution of integrative capabilities: innovation in cardiovascular drug discovery. Industrial and Corporate Change 3 (3), 607–630.

Henderson, R., Orsenigo, L., Pisano, G.P., 1999. The pharmaceutical industry and the revolution in molecular biology: exploring the interactions between scientific institutional and organizational change, In: Mowery, D.C., Nelson, R.R. (Eds.), The Sources of Industrial Leadership Cambridge Univ. Press, Cambridge, MA.

Hummon, N.P., Fararo, T.J., 1995. Assessing hierarchy and balance in dynamic network models. Journal of Mathematical Sociology 20, 145–159.

Lovasz, L., Plummer, M.D., 1986. Matching theory, Annals of Discrete Mathematics 29. North Holland, Amsterdam.

Mc Cullagh, P., 1978. A logistic model for paired comparisons with ordered categorical data. Biometrika 64, 449–453.

Mc Kelvey, M., 1995. Evolutionary Innovation: The Business of Biotechnology. Oxford Univ. Press, Oxford.

Nelson, R.R., 1959. The simple economics of basic scientific research. Journal of Political Economy 67, 297–306.

Orsenigo, L., 1989. The Emergence of Biotechnology. St. Martin Press, New York.

Pisano, G.P., 1991. The governance of innovation. Vertical integration and collaborative arrangements in the biotechnology industry. Research Policy 20, 237–249.

Powell, W.W., 1996. Interorganizational collaboration in the biotechnology industry. Journal of Institutional and Theoretical Economics 152 (1), 197–215.

Powell, W.W., Doput, K.W., Smith-Doerr, L., 1996. Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology. Administrative Science Quarterly 41, 116–145.

Simon, H.A., 1969. The Architecture of Complexity, The Sciences of the Artificial. MIT Press, Cambridge, MA.

Slepian, P., 1968. Mathematical Foundations of Network Analysis. Springer-Verlag, Berlin.

Testa, B., Meyer, U.A., 1995. Advances in Drug Research. Academic Press, New York.

Tsuji, R., 1997. Permutation Tests for Symmetry and Transitivity in Real-Valued Data. Proceedings of JAMS 24th Conference, Ryutsu Keizai University, Ibaraki, Japan.

Vos, R., 1991. Drugs Looking for Diseases. Kluwer Academic Publishers, Dordrecht.

Walker, G., Kogut, B., Shan, W., 1997. Social capital, structural holes and the formation of an industry network. Organization Science 8 (2), 109–125.