2013-15/September 2013

Document de travail

THE RISE AND FALL OF R&D NETWORKS

Mario V. Tomasello
(ETH Zurich, Switzerland)

Mauro Napoletano
(OFCE-Sciences Po, SKEMA Business School and Campus de Sophia Antipolis, France)

Antonios Garas
(ETH Zurich, Switzerland)

Frank Schweitzer
(ETH Zurich, Switzerland)

ofce

The Rise and Fall of R&D Networks

Mario V. Tomasello[*], Mauro Napoletano[†], Antonios Garas[‡] and Frank Schweitzer[§]

August 30, 2013

Abstract

Drawing on a large database of publicly announced R&D alliances, we track the evolution of R&D networks in a large number of economic sectors over a long time period (1986-2009). Our main goal is to evaluate temporal and sectoral robustness of the main statistical properties of empirical R&D networks. By studying a large set of indicators, we provide a more complete description of these networks with respect to the existing literature. We find that most network properties are invariant across sectors. In addition, they do not change when alliances are considered independently of the sectors to which partners belong. Moreover, we find that many properties of R&D networks are characterized by a rise-and-fall dynamics with a peak in the mid-nineties. Finally, we show that such properties of empirical R&D networks support predictions of the recent theoretical literature on R&D network formation.

1 Introduction

This work investigates the properties of empirical R&D networks across many sectors and over time. In several industries, and especially in those with rapid technological growth, innovation relies on general and abstract knowledge often built on scientific research (Powell et al., 1996). This has allowed for a division of innovative labor and fostered collaboration across firms (Arora and Gambardella, 1994a,b). Accordingly, the last three decades have witnessed a significant growth in the number of formal and informal R&D collaborations (e.g. Hagedoorn, 2002; Powell et al., 2005).

Several works have tried to shed light on the structural properties of R&D networks. These empirical studies have shown that R&D networks are typically sparse and characterized by

[*]Corresponding author. E-mail: mtomasello@ethz.ch. Chair of Systems Design, Department of Management, Technology and Economics, ETH Zurich, Weinbergstrasse 56/58, 8092 Zurich, Switzerland.

[†]OFCE and SKEMA Business School, Sophia-Antipolis, France, and Sant'Anna School of Advanced Studies, Pisa, Italy. Mail Address: OFCE, SKEMA Business School, Campus de Sophia Antipolis 60, rue Dostoevski BP 85 06902 Sophia Antipolis Cedex, France. E-mail: mauro.napoletano@sciencespo.fr.

[‡]Chair of Systems Design, Department of Management, Technology and Economics, ETH Zurich, Weinbergstrasse 56/58, 8092 Zurich, Switzerland. E-mail: agaras@ethz.ch.

[§]Chair of Systems Design, Department of Management, Technology and Economics, ETH Zurich, Weinbergstrasse 56/58, 8092 Zurich, Switzerland. E-mail: fschweitzer@ethz.ch.

heavily asymmetric degree distributions (e.g. Hanaki et al., 2010; Powell et al., 2005; Rosenkopf and Schilling, 2007). Furthermore, R&D networks display the "small world" property (e.g. Fleming et al., 2007; Fleming and Marx, 2006). In other words, they are characterized by short average path length and high clustering (Watts and Strogatz, 1998). At the same time, it has been observed that some R&D network properties may change over time. One prominent example is the rise and fall of small world properties that Gulati et al. (2012) found in the R&D network of the global computer industry.

The above empirical studies have greatly contributed to the understanding of empirically observed R&D networks. However, they have often focused only on a small number of industries or have rarely considered the properties of the networks at different time periods. Finally, they have focused on a limited set of network measures (e.g. size, degree heterogeneity, small world property).

The increasing importance of R&D collaborations for industrial innovation has also originated theoretical research on R&D networks. In these theoretical works, R&D collaboration allows innovation either via resource sharing (Goyal and Joshi, 2003; Goyal and Moraga-Gonzalez, 2001; Westbrock, 2010) or via the recombination of firm's knowledge stock with those of its partners (Cowan and Jonard, 2004; König et al., 2011, 2012). One key prediction of these models is that – under non-negligible costs of collaboration – R&D networks should be organized as core-periphery architectures, i.e. they should display a core of densely connected firms, in turn linked with a periphery of firms having few alliances among them. Nevertheless, to the best of our knowledge no empirical study has tried so far to confirm or deny the presence of core-periphery architectures in R&D networks.

Our work contributes to the foregoing empirical and theoretical literature along several dimensions. *First*, we analyze the R&D networks in a large number of manufacturing and service sectors. After analyzing the *pooled R&D network*, i.e. the network containing all alliances independently of the sectors to which the partners belong, we study a series of R&D networks for several industrial sectors at a 3-digit SIC level. Via this disaggregated analysis, we are able to check whether the network properties that have been analyzed by the current literature for sectors like computers (e.g. Hanaki et al., 2010) or pharmaceuticals (e.g. Powell et al., 2005) are robust across different sectors of activity. In addition, by comparing the properties at the pooled and at the sectoral levels, we are able to check for the presence of *universal* properties of R&D networks that hold irrespectively of the scale of aggregation at which they are observed. *Second*, we perform a longitudinal analysis of empirical R&D networks. In particular, we consider the network dynamics in the period from 1986 to 2009. This procedure allows us to check whether network properties are robust over time, or if instead they exhibit different trends in different time-periods. *Third*, we investigate a broad set of network properties. We start our analysis by studying the basic network measures that have so far been considered in the empirical literature (size, degree heterogeneity, small world property). In addition, we study measures related to more complex features of the network, such as assortativity (i.e. the presence of positive correlation in the number of alliances among firms, see also Newman, 2003), the existence of communities (e.g. Newman and Girvan, 2004) as well as the presence of core-periphery architectures. This way, we extend the existing knowledge on R&D networks by adding new stylized facts to the already existing ones. Furthermore, the analysis of core-periphery architectures allows a fresh test of the predictions of the recent theoretical models on R&D networks.

We find that both the pooled and the sectoral R&D networks exhibit the same patterns for a wide set of network properties, like the fraction of firms in the largest connected component, the shape of degree distributions or the presence of small world properties. In addition, both the pooled and the sectoral R&D networks are organized into core-periphery architectures of the type predicted by the theoretical literature on R&D networks (e.g. König et al., 2012). In contrast, pooled and sectoral networks differ with respect to the presence of assortativity. In the pooled R&D network firms with many alliances tend to collaborate with partners involved in many alliances as well, whilst the opposite is found at the sectoral level, where firms with many alliances tend to collaborate with firms having fewer alliances. Furthermore, we find that most of the above properties display a rise-and-fall dynamics over time. For instance, both network size and connectivity of the pooled network have first increased over time, reaching a peak in the period 1994-1997, and then they have significantly decreased until the end of our observation period. Interestingly, the above non-monotonic dynamics is very pervasive, as we observe it in most of the sectoral R&D networks we study. More precisely, two distinct phases can be found both in the pooled and in the sectoral R&D network dynamics. The first phase (from 1990 to 2001) is characterized by a significant growth in the size of the network components wherein firms are directly or indirectly connected. As we show in Section 3, this dynamics was driven by a growth in the number of firms participating in R&D alliances rather than by the change in the number of alliances among firms already involved in previous collaborations. In the second phase (from 2001 to 2009) this trend reverses. The network breaks down into several small connected components, involving firms with few alliances.

The above results have several implications for the literature. First, the existence of network properties that are invariant with respect to the scale of aggregation is analogous to previous findings in the industrial dynamics literature (e.g. Bottazzi and Secchi, 2003; Lee et al., 1998) and favors the idea that some of the laws governing the evolution of R&D networks can be analyzed independently of the characteristics of the sector to which the firms belong. Second, the finding that both the pooled and the sectoral R&D networks are characterized by core-periphery structures confirms the predictions of the recent theoretical literature on R&D networks. Finally, our results show that the rise-and-fall dynamics, which has so far been emphasized only in relation to network size and small worlds (see Gulati et al., 2012), is also displayed by more sophisticated topological network properties (e.g. assortativity, core-periphery and nested architectures). In turn, the rise-and-fall dynamics could be explained by the dynamics of knowledge recombination associated with the R&D networks (see Section 8 for more discussion).

The paper is organized as follows. Section 2 describes the data and the methodology used to build the networks of R&D alliances. In Section 3 we discuss results about the basic properties of R&D networks, such as network size, network density and the emergence of a giant component. Section 4 analyzes the characteristics of the degree distributions. In Sections 5 and 6 we study more sophisticated network properties, such as assortativity, and the presence of small worlds and communities in the network. Section 7 studies the presence of core-periphery architectures. In Section 8 we discuss the implications of our results in light of the existing theoretical and empirical literature on R&D networks. Finally, Section 9 concludes.

2 Data and Methodology

A *R&D network* is a representation of the research and development alliances occurring between firms in one or more industrial sectors in a given period of time. Every network consists of a set of *nodes* and *links* connecting pairs of nodes. In our representation, each node of the network is a *firm* and every link represents a *R&D alliance* between two firms. By R&D alliance, we refer to an event of partnership between two firms, that can span from formal joint ventures to more informal research agreements, specifically aimed at research and development purposes. To detect such events, we use the *SDC Platinum* database, provided by Thomson Reuters, that reports all publicly announced alliances, from 1984 to 2009, between several kinds of economic actors (including manufacturing firms, investors, banks and universities). We then select all the alliances characterized by the "R&D" flag; after applying this filter, a total of 14829 alliances are listed in the dataset.

Information in the SDC dataset is gathered only from announcements in public sources, such as press releases or journal articles. Nevertheless, despite the bias that could be introduced by such a collection procedure, Schilling (2009) shows that the SDC Thomson dataset provides a consistent picture with respect to alternative alliance databases (e.g. CORE and MERIT-CATI) in terms of alliance activity over time, geographical location of companies and industry composition.

Because the SDC Platinum dataset does not have a unique identifier for each firm, all the associations between alliances and firms (i.e. the construction of the network itself) are based only on the firm names reported in the dataset. Thus, it could happen that two or more entries are listed with different names, because they appear in two distinct alliance events, even though they correspond to the same firm. For this reason, we check all firm names and control for all legal extensions (e.g. "ltd", "inc", etc.) and other recurrent keywords (e.g. "bio", "tech", "pharma", "lab", etc.) that could affect the matching between entries referring to the same firm. We decide to keep as separated entities the subsidiaries of the same firm located in different countries. The raw dataset contains a total of 16313 firms, which are reduced to 14561 after running such an extensive standardization procedure.

In our network representation, we draw a link connecting two nodes every time an alliance between the two corresponding firms is announced in the dataset. An alliance is associated with an *undirected* link, as we do not have any information about the initiator of the alliance. When an alliance involves more than two firms (*consortium*), all the involved firms are connected in pairs, resulting into a fully connected clique. Following this procedure, the 14829 alliance events listed in the dataset result in a total of 21572 links. Similarly to Rosenkopf and Schilling (2007), the R&D network we consider in our study is *unipartite*, as we only have one set of actors ("the firms"), whose elements may be connected – or not – by publicly announced alliances.[1]

Multiple links between the same nodes are in principle allowed (two firms can have more than one alliance on different projects). Nevertheless, as we aim at studying the connections between firms, and not the number of alliances a firm is involved in, we discard this information and use *unweighted* links in our network representation. For this reason, we define the *degree* of a node as the number of other nodes to which it is linked, i.e. the number of partners that a firm has – not the number of alliances. Furthermore, a firm appears in the R&D network only if it is involved in at least one alliance. Our study is focused exclusively on the embeddedness of firms into an alliance network. For this reason, isolated nodes are not part of our network representation.

Both the links and the nodes of the R&D network are characterized by an entry/exit dynamics. Alliances between firms have a finite duration (see Deeds and Hill, 1999; Phelps, 2003). This causes some firms to disappear from the network, after they no longer participate in any alliance. Likewise, many new firms that were not listed in any previous alliance may enter the network at the beginning of a new year. Our longitudinal study clearly requires precise temporal information about the formation and the deletion of alliances. The SDC Platinum dataset contains the beginning date of every alliance, but there is no information about any

[1] Our work differs from previous empirical studies (e.g. Cantner and Graf, 2006; Hanaki et al., 2010; Lissoni et al., 2013) which construct the network through the association of firms with patents and/or inventors. Those studies use patent data to build the network and associate elements in the set "firms" to the elements in the set "patents". This way, the network they obtain is *bipartite*.

of the ending dates (firms do usually not organize press releases to announce the end of an alliance). We are thus forced to make some assumptions about the alliance durations. We start by drawing the duration of every alliance from a normal distribution with mean value from 1 to 5 years and standard deviation from 1 to 5 years, and we find that all our results remain qualitatively unchanged by changing the mean value and the standard deviation within these ranges. More precisely, the variation of the standard deviation has nearly no influence on the patterns exhibited by of all measures we compute on the networks. The variation of the mean alliance duration changes the absolute values of the network indicators, but it does not affect their time-evolution and peak positions. Given the strong robustness of the R&D network to the variation of alliance lengths, we take a conservative approach and assume a fixed 3-year length for every partnership, consistently with previous empirical work (e.g. Deeds and Hill, 1999; Phelps, 2003; Rosenkopf and Schilling, 2007). More precisely, we link two nodes when an alliance between the corresponding firms occurs and we delete this link 3 years after its formation. In this way, we are able to build 26 snapshots of the R&D network – one for every year – from 1986 to 2009. From now on we call the network containing all companies, irrespective of their industrial sector, the *pooled R&D network*.

Every firm listed in the SDC Platinum dataset is associated with its SIC (Standard Industrial Classification), a US-government code system for classifying industrial sectors. This allows us to build the *sectoral R&D networks* for the several sectors that we identify in the dataset. A sectoral R&D network centered around a given sector contains only alliances in which at least *one* of the partners has a three-digit SIC code matching the selected sector (see also Rosenkopf and Schilling, 2007, for a similar approach). The rules for link deletion are the same as in the pooled R&D network. More precisely, we select for our study the 30 largest industrial sectors, in terms of number of firms engaged in alliances in 1995 (the year in which the pooled R&D network reaches its maximum size). This list includes manufacturing and service sectors. It has to be noticed that the latter includes also sectors like "laboratories and testing companies" and "universities". Table 1 provides the list of the different sectors we consider in our study.

We study both the pooled R&D network and the sectoral R&D networks by computing a set of network indicators along the whole observation period. All the results are presented below. We group our analysis into five sections: basic network statistics, heterogeneity in alliance behavior, assortativity, small world and communities, core-periphery structures.

3 Basic Network Statistics

Fig. 1 shows six snapshots of the pooled R&D network. The plots are produced using the library *igraph* for the R package, and the networks are displayed using the Fruchterman-Reingold algorithm (cf. Fruchterman and Reingold, 1991). This is a force-based algorithm for network visualization which positions the nodes of a graph in a two-dimensional space so that all the edges are of similar length and there are as few crossing edges as possible. The result is that the most interconnected nodes are displayed close to each other in the two-dimensional plot. The ten largest industrial sectors are depicted with different colors. The figure shows that two clusters always dominate the pooled R&D network: a cluster centered on pharmaceutical companies and a cluster centered on *ICT*-related companies.

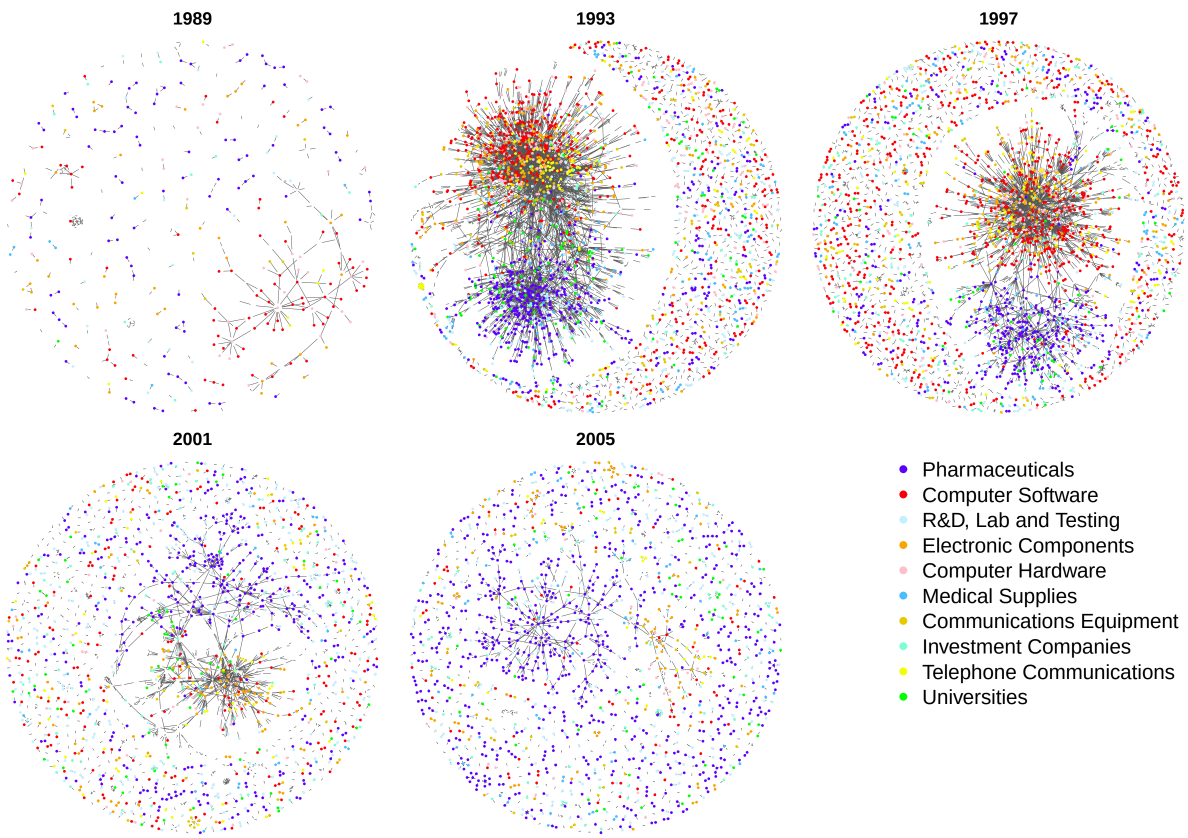

Figure 1: Pooled R&D network snapshots in 1989, 1993, 1997, 2001 and 2005. We plotted – in different colours – only the ten largest sectors, in order to ease visualisation.

Fig. 1 denotes the presence of different phases in the evolution of the R&D network. More precisely, the plots suggest the presence of a significant network growth until 1997, and a reversal of this trend in the last periods of our sample. To shed more light on this phenomenon, we report in Table 1 the network size, in terms of number of firms taking part in the R&D network – i.e. companies involved in at least one alliance. The observation period 1986-2009 is divided into six sub-periods of 4 years each and we average the network size within each sub-period. Table

1 confirms the presence of a rise-and-fall dynamics in the pooled network. More precisely, the number of companies involved in R&D alliances increases to a peak in the mid-nineties and then shrinks again, both at the pooled and the sectoral level (see Table 1). In each sector, the number of firms involved in R&D alliances has a peak in the years 1994-1997. Interestingly, only the Pharmaceutical sector, besides the peak in the period 1994-1997, has an additional peak of slightly larger size in the period 2006-2009. The presence of a peak in the period 1994-1997 is a characteristic of many further network measures considered in this study and leads us to define that period as the "golden age" of R&D networks.

	1986-1989	**1990-1993**	**1994-1997**	**1998-2001**	**2002-2005**	**2006-2009**
Pooled Network	280	2515	4918	2626	2219	1829
Manufacturing Sectors						
Pharmaceuticals (283)	77	645	935	682	825	949
Computer Hardware (357)	51	385	744	202	92	29
Electronic Components (367)	54	328	581	253	222	165
Communications Equipment (366)	17	207	475	181	113	60
Medical Supplies (384)	10	164	280	122	119	123
Laboratory Apparatus (382)	10	139	243	116	94	87
Motor Vehicles (371)	6	108	190	97	85	78
Aircrafts and parts (372)	8	83	136	60	40	26
Inorganic Chemicals (281)	15	108	152	50	45	31
Household Audio-Video (365)	9	110	164	90	65	30
Plastics (282)	11	97	121	44	36	18
Electrical Machinery NEC (369)	2	54	96	26	24	37
Special Machinery (355)	2	33	82	34	17	11
Crude Oil and Gas (131)	3	42	72	62	35	27
Naut./Aeronaut. Navigation (381)	1	49	82	21	16	12
Organic Chemicals (286)	5	44	60	18	23	18
Service Sectors						
Computer Software (737)	69	560	1488	549	284	122
R&D, Lab and Testing (873)	26	477	848	534	596	500
Universities (822)	3	192	374	166	152	83
Telephone Communications (481)	12	184	350	132	82	22
Investment Companies (679)	14	138	298	232	207	125
Professional Equipment Wholesale (504)	4	64	142	26	8	8
Engineer.,Architec.,Survey (871)	2	74	129	62	26	16
Radio and TV Broadcasting (483)	2	26	88	22	7	4
Electric Services (491)	NaN	50	78	38	26	15
Electrical Goods Wholesale (506)	NaN	26	84	19	10	8
Cable and TV Services (484)	NaN	18	78	8	6	3
Motion Picture Production (781)	NaN	15	91	14	4	1
Business Services (738)	1	15	66	37	30	5
Management,Consulting,PR (874)	1	28	96	61	64	28

Table 1: Network size for the pooled and the sectoral R&D networks (SIC codes are in brackets). The values are averages within each sub-period. *Note*: missing values refer to sectors with not enough observations.

A deeper investigation shows that the growth in size of the R&D network in the mid-nineties corresponds to a decrease in its density (defined as the number of existing links divided by the number of all possible links in the network). This is shown in Fig. 2, where the density of the pooled R&D network, (and its mid-nineties decline), is compared to the network size (and its mid-nineties peak). This means that the expansion of the R&D network was not generated by an increase of the alliances among the firms that were already part of the network. Instead, it was mainly the result of new alliances created by entrant firms. After the "golden age", the shrinking of the network is associated with a decrease in the number of nodes. This fall in the number of firms participating into alliances has however no effect on the density of the network, which remains constant until the end of the observation period (cf. Fig. 2).

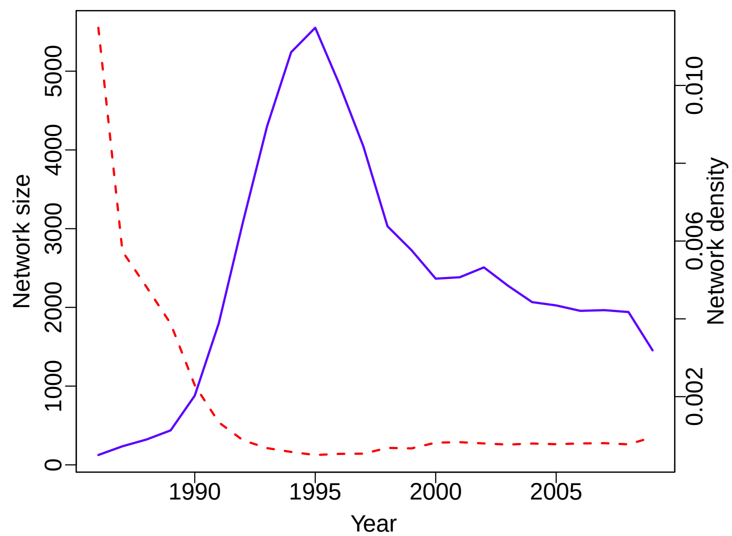

Figure 2: Time-evolution of size (solid line, left axis) and density (dashed line, right axis) of the pooled R&D network.

Next, we compute the fraction of nodes belonging to the largest *connected component* of the network. A connected component is defined as a set of nodes which are connected to each other by at least one path (i.e. a sequence of links). We refer to the largest connected component as the *giant component* of the network. The giant component size to the overall network size ratio (or *giant component fraction*) is a rough indicator of the network connectedness. Our results are reported in Table 2. This measure has been computed for every year from 1986 to 2009 and then averaged within six sub-periods of 4 years each. Similarly to the network size, the giant component fraction displays a non-monotonic trend at the pooled level, reaching a peak in the mid-nineties and then shrinking again. The emergence of a giant component in the network is of particular interest, as different theoretical works (e.g. Goyal and Joshi, 2003; König et al., 2012) have stressed the importance of the relation between high network connectedness and efficiency in terms of aggregate profits. We also find that the emergence of such non-monotonic dynamics in the giant component is very robust to sectoral disaggregation. Indeed, we observe it in almost all the sub-networks representing the different industrial sectors (see Table 2). More precisely, 19 out of the 30 sectoral R&D networks show a giant component peak either in the 1990-1993 or in the 1994-1997 period. The sectors that do not have a peak show a more volatile evolution of their giant component. Among these, only 4 are manufacturing industries (Inorganic Chemicals, Household Audio-Video, Special Machinery, Organic Chemicals), while the other sectors are related to services or sales.

Furthermore, Fig. 3 shows the time-evolution of the number of all connected components of the network and of their average size.[2] Both indicators have a peak in the years around 1995 (i.e. the ones corresponding to the 1994-1997 sub-period). This is indicative of the tendency of firms to form more (and larger) connected components until 1995. Afterwards, a fragmentation process takes place. The average size of network components starts to decrease; the number of the components remains stable for two more years, but eventually declines as well (cf. Fig. 3). As a result, the large R&D network of the "golden age" period 1994-1997, dominated by a giant component, is replaced by a network with less (and smaller) components. The same results hold for sectoral R&D networks.[3] Fig. 1 visualizes this dynamics: the pooled R&D network is characterized by the presence of a giant component that expands until 1997 and subsequently leaves space to a growing periphery of disconnected dyads (pairs of allied firms).

The above analysis reveals the existence of patterns that are invariant to the scale of aggregation or the sector where they are observed. Namely, both the pooled and sectoral R&D networks experience a robust growth in both size and connectedness until 1997. In particular, the years between 1994 and 1997 (the "golden age" of R&D networks), witness not only a higher number of alliances, but also the emergence of a significantly large giant component. This robust growth is then replaced by a decline phase, characterized by both a reduction in the number of alliances and the breaking-up of the network into smaller components. In the next section, we will go into more detail on how these alliances are organized, by studying the degree distributions of the pooled and sectoral R&D networks.

[2]The distribution of the size components is extremely right skewed and fat-tailed. This is due to the fact that one or few large components co-exist with many disconnected pairs of connected firms. Even though the aritmetic mean is not entirely meaningful or predictive for heavy-tailed distributions, we still report it not only because it is fully computable (we have finite size networks), but also because it gives an idea about the evolution of the component sizes over the period we study. Same remarks apply to the analysis of the average degree that we discuss in Section 4.

[3]Because of space constraints the sectoral plots are not shown. However, they are available from the authors upon request.

	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Pooled Network	0.10	0.53	0.53	0.33	0.26	0.20
Manufacturing Sectors						
Pharmaceuticals (283)	0.08	0.58	0.68	0.49	0.36	0.32
Computer Hardware (357)	0.27	0.59	0.67	0.51	0.28	0.13
Electronic Components (367)	0.15	0.53	0.61	0.49	0.38	0.13
Communications Equipment (366)	0.18	0.42	0.55	0.25	0.25	0.15
Medical Supplies (384)	0.21	0.04	0.05	0.05	0.06	0.05
Laboratory Apparatus (382)	0.26	0.15	0.13	0.08	0.08	0.07
Motor Vehicles (371)	0.79	0.52	0.39	0.15	0.21	0.10
Aircrafts and parts (372)	0.65	0.47	0.38	0.23	0.20	0.16
Inorganic Chemicals (281)	0.30	0.26	0.17	0.15	0.12	0.29
Household Audio-Video (365)	0.61	0.57	0.61	0.63	0.60	0.28
Plastics (282)	0.23	0.25	0.20	0.23	0.15	0.19
Electrical Machinery NEC (369)	1.00	0.36	0.22	0.20	0.15	0.11
Special Machinery (355)	0.88	0.25	0.13	0.19	0.27	0.26
Crude Oil and Gas (131)	0.67	0.15	0.14	0.10	0.11	0.15
Naut./Aeronaut. Navigation (381)	1.00	0.38	0.26	0.21	0.22	0.24
Organic Chemicals (286)	0.73	0.13	0.17	0.25	0.13	0.22
Service sectors						
Computer Software (737)	0.33	0.54	0.54	0.23	0.11	0.06
R&D, Lab and Testing (873)	0.13	0.19	0.27	0.11	0.10	0.07
Telephone Communications (481)	0.43	0.61	0.58	0.25	0.26	0.28
Universities (822)	0.90	0.17	0.25	0.10	0.08	0.05
Investment Companies (679)	0.21	0.36	0.27	0.23	0.28	0.10
Professional Equipment Wholesale (504)	0.69	0.13	0.16	0.23	0.37	0.28
Engineer.,Architec.,Survey (871)	1.00	0.12	0.15	0.11	0.12	0.20
Motion Picture Production (781)	NaN	0.39	0.24	0.22	0.62	0.50
Management,Consulting,PR (874)	1.00	0.23	0.07	0.09	0.09	0.11
Radio and TV Broadcasting (483)	1.00	0.40	0.17	0.16	0.42	0.61
Cable and TV Services (484)	NaN	0.35	0.16	0.31	0.53	0.75
Business Services (738)	1.00	0.48	0.08	0.11	0.14	0.65
Electrical Goods Wholesale (506)	NaN	0.29	0.12	0.15	0.25	0.34
Electric Services (491)	NaN	0.35	0.11	0.15	0.24	0.21

Table 2: Fraction of the giant component for the pooled and the sectoral R&D networks (SIC codes are in brackets). The values are averages within each sub-periods. *Note*: missing values refer to sectors with not enough observations.

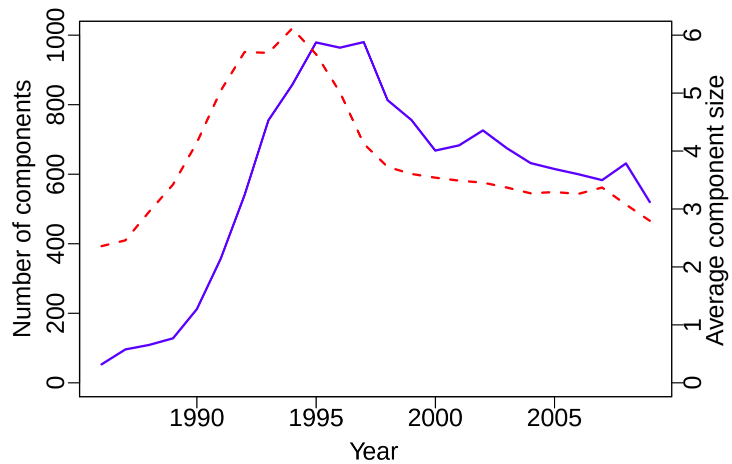

Figure 3: Time-evolution of number of connected components (solid line, left axis) and average size distribution of connected components (dashed line, right axis) in the pooled R&D network.

4 Heterogeneity in alliance behavior

A large part of literature has analyzed the properties of the degree distributions in R&D networks. Empirical studies have shown that degree distributions in R&D networks tend to be highly skewed. Moreover, some studies find exponential distributions (Riccaboni and Pammolli, 2002), while others find power-law distributions (Powell et al., 2005). The presence of a power-law distribution would indicate the existence of an underlying multiplicative growth process (Reed, 2001; Simon, 1955). In the context of R&D networks this means that firms which have many collaborations already attract more new partners than firms with only few collaborations. This idea underlies the "preferential attachment" model by Barabasi and Albert (1999), which predicts the emergence of a power-law degree distribution. However, this model assumes that all firms (even the new entrants) know how many collaborations every other firm in the network has. This may become unrealistic, especially in large networks or situations in which this information is not publicly available. More realistic models assume that firms have only local information about the network. The network formation model introduced by König et al. (2013) assumes that firms search for the most central partner in their local neighborhood. Their model generates exponential degree distributions with power-law tails. In the model of Jackson and Rogers (2007), agents also form links locally, which can result in power-law degree distributions as well as exponential degree distributions, depending on various parameters. We extend the existing discussion about the degree distributions in R&D networks by studying their evolution over time and comparing the results between different sectors. Given the small size of many of our networks, we did not test or validate any functional form, but we rather measured the statistical properties of the degree distributions, in order to assess their main features and get insights into the underlying network formation process.

As already mentioned in Section 2, we define the degree as the number of partners of a firm, and not the number of alliances. For this reason, we count multiple alliances between the same two firms as one, and we count all the firms participating in the same consortia as distinct partners. Furthermore, like in Section 3, the whole observation period is divided into six sub-periods lasting 4 years. All the measures we present are computed by aggregating firm degree data relative to the same sub-period. Fig. 4 shows the degree distributions of the pooled R&D network in the six analyzed sub-periods. More precisely, given each degree distribution, we report its *complementary cumulative distribution function* $P(x)$, defined as the fraction of nodes having degree greater than or equal to x:

$$P(x) = \int_x^\infty p(x')\mathrm{d}x'. \tag{1}$$

where $p(x')$ is the *probability density function*, defining the fraction of nodes in the network with degree x. The complementary cumulative distribution function is more robust than the probability density function against fluctuations due to finite sample sizes, particularly in the tail. We find that the degree distribution of the pooled R&D network is very broad and skewed, in all periods. Moreover, the shape of the degree distribution is independent of the network size. For instance, the degree distributions of the pooled R&D network in the "golden age" 1994-1997 (maximum degree ~ 200) has a very similar shape to that of the early period 1986-1989 (maximum degree ~ 20). In addition, most of the sectoral R&D networks (not shown here) exhibit this kind of degree distribution, during the whole observation period.

Table 3 shows the first four moments of the degree distribution of the pooled network in each sub-period. In all periods, the degree distribution displays high variance associated with high right-skewness and excess kurtosis. In addition, the p-values of the Kolmogorov-Smirnov test show that the degree distributions of the pooled network are extremely far from the Normal benchmark. Moreover, Table 3 shows that all the four moments of the degree distribution increase in the first years of the sample, reaching a peak either in the 1990-1993 or in the 1994-1997 period, and then decrease again. The mean degree has a value of 1.51 partners per firm in the early period 1986-1989; it then exhibits a peak value in 1990-1993 (2.52 partners per firm), which remains almost unchanged in 1994-1997 (2.51 partners per firm), showing that firms have on average more alliance partners in the "golden age" of alliance formation. The average number of partners per firm eventually decreases again, reaching a value of 1.49 in the late period 2006-2009.

As we discussed above, the degree distribution in the pooled R&D network is highly

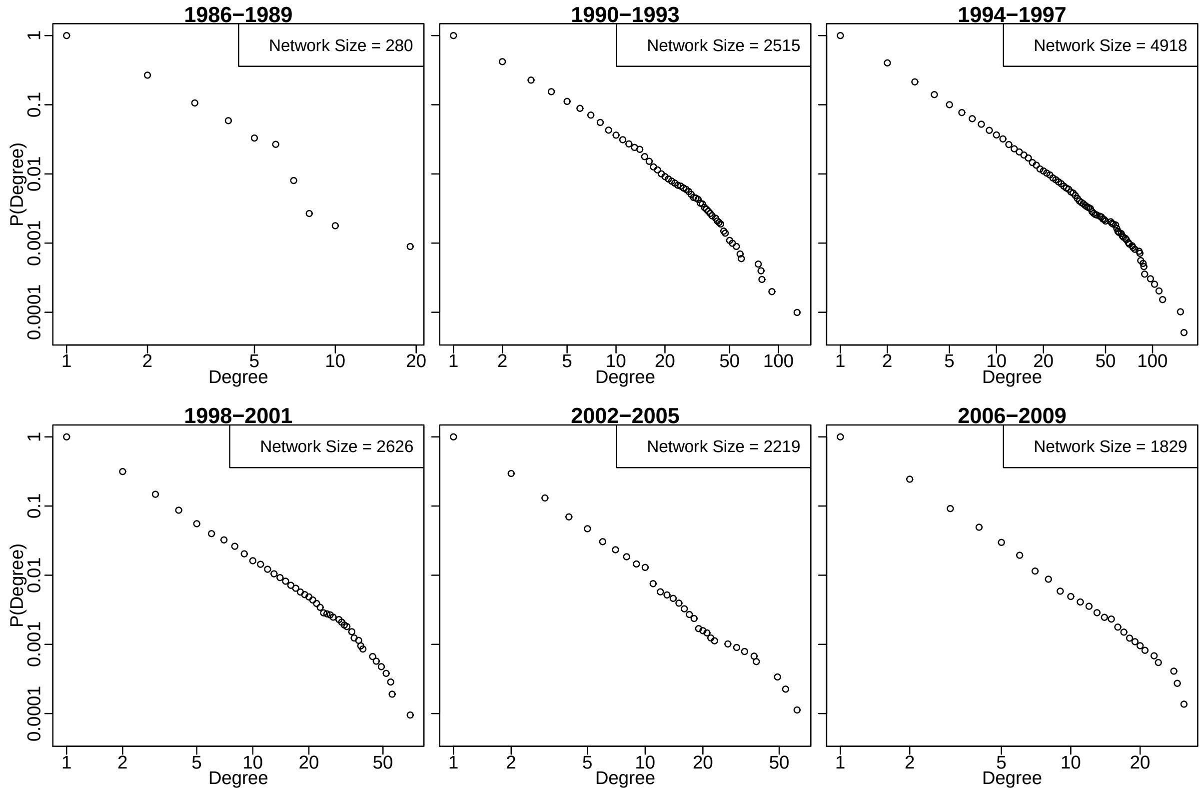

Figure 4: Complementary cumulative degree distributions of the pooled R&D network.

dispersed, as shown by standard deviation values that are always comparable or even larger than the mean values. This holds especially for the 1994-1997 period, when the standard deviation has a peak at 4.98, while the mean value is 2.51 partners per company. Same considerations apply to the evolution of the skewness and kurtosis coefficients over time. In particular, the very high values of the kurtosis coefficient (especially in the period 1994-1997) are indicative of heavy tails in the R&D networks degree distributions, which in turn imply the presence in the networks of "hubs" concentrating a high number of alliances.

	1986-1989	**1990-1993**	**1994-1997**	**1998-2001**	**2002-2005**	**2006-2009**
Mean	1.51	2.52	2.51	1.87	1.70	1.49
SD	1.22	4.30	4.98	2.77	2.11	1.45
Skewness	4.90	9.35	11.28	9.26	10.56	7.92
Kurtosis	47.30	158.40	206.69	133.70	200.25	104.84
KS test p-Value	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$

Table 3: Degree distribution statistics and p-values of Kolmogorov-Smirnov (KS) test for the pooled R&D network.

The degree distributions of the sectoral R&D networks display patterns that are similar to those of the pooled R&D network.[4] In particular, all sectoral degree distributions are characterized by high variance associated with significant skewness and kurtosis in all sub-periods. We report in Table 4 the values of the average degree for the pooled and the sectoral R&D networks in the six sub-periods, clearly confirming such a cross-sector similarity. In all sectoral networks, firms have on average more collaborators during the "golden age" of alliance activity (1994-1997). The only two exceptions are represented by two manufacturing industries, motor vehicles (having a peak in 1986-1989) and organic chemicals (that has a first peak in 1986-1989 and a second one in 1994-1997).

	1986-1989	**1990-1993**	**1994-1997**	**1998-2001**	**2002-2005**	**2006-2009**
Pooled Network	1.51	2.52	2.51	1.87	1.70	1.49
Manufacturing Sectors						
Pharmaceuticals (283)	1.22	2.09	2.22	1.82	1.57	1.55
Computer Hardware (357)	1.50	2.10	2.45	2.30	1.55	1.09
Electronic Components (367)	1.32	2.18	2.38	2.15	1.81	1.44
Communications Equipment (366)	1.10	1.82	2.03	1.57	1.48	1.34
Medical Supplies (384)	1.00	1.26	1.31	1.21	1.20	1.16
Laboratory Apparatus (382)	1.00	1.41	1.36	1.24	1.20	1.19
Motor Vehicles (371)	2.31	1.89	1.78	1.40	1.49	1.29
Aircrafts and parts (372)	2.00	2.25	2.00	1.68	1.41	1.40
Inorganic Chemicals (281)	1.28	1.48	1.53	1.23	1.17	1.27
Household Audio-Video (365)	1.44	2.11	2.61	2.32	2.20	1.58
Plastics (282)	1.07	1.54	1.55	1.46	1.29	1.11
Electrical Machinery NEC (369)	1.00	1.45	1.52	1.26	1.11	1.10
Special Machinery (355)	1.00	1.34	1.37	1.24	1.21	1.07
Crude Oil and Gas (131)	1.09	1.70	1.68	1.51	1.28	1.11
Naut./Aeronaut. Navigation (381)	1.33	1.49	1.49	1.23	1.13	1.09
Organic Chemicals (286)	1.26	1.17	1.26	1.14	1.09	1.12
Service Sectors						
Computer Software (737)	1.70	2.16	2.21	1.52	1.27	1.13
R&D, Lab and Testing (873)	1.08	1.68	1.81	1.40	1.43	1.27
Telephone Communications (481)	1.19	2.84	2.53	1.42	1.57	1.28
Universities (822)	1.27	1.66	1.76	1.51	1.35	1.11
Investment Companies (679)	1.04	1.74	1.62	1.53	1.63	1.35
Professional Equipment Wholesale (504)	1.22	1.24	1.42	1.22	1.09	1.00
Engineer.,Architec.,Survey (871)	1.00	1.36	1.40	1.17	1.07	1.09
Motion Picture Production (781)	NaN	1.38	1.36	1.02	1.00	1.00
Management,Consulting,PR (874)	1.00	1.20	1.20	1.19	1.16	1.06
Radio and TV Broadcasting (483)	1.33	1.69	1.31	1.15	1.11	1.11
Cable and TV Services (484)	NaN	1.34	1.51	1.03	1.17	1.00
Business Services (738)	1.00	1.17	1.22	1.15	1.16	1.05
Electrical Goods Wholesale (506)	NaN	1.35	1.34	1.06	1.05	1.07
Electric Services (491)	NaN	1.57	1.38	1.22	1.22	1.25

Table 4: Average degree (number of partners) for the pooled and the sectoral R&D networks (SIC codes are in brackets). *Note*: missing values refer to sectors with not enough observations.

The previous analysis indicates the presence of heavy tails in both the pooled and sectoral

[4]These results are not shown here, but are available from the authors upon request.

degree distributions. In order to get an estimate of the "heaviness" of those tails from a non-parametric point of view, we compute the Hill Estimator (Hill, 1975), a tool commonly used to study the tails of economic data. If n is the number of observations (in our case, the number of nodes in the R&D network) and k is the number of tail observations ($k \leq n$), the inverse of the Hill estimator (HE) is defined as:

$$\hat{h}^{-1} = k^{-1} \sum_{i=1}^{k} \left[\log(x_i) - \log(x_{min})\right], \tag{2}$$

where x_{min} represents the beginning of the tail and x_i, $i = 1 \ldots k$ are the tail observations, i.e. the degree values such that $x_i \geq x_{min}$. The smaller the HE value, the "heavier" the tail of the degree distribution is. In particular, the degree distributions of most biological, social and economic systems display values of the HE between 2 and 4 (see Clauset et al., 2009). A value of the HE lower than 2 indicates an extremely heavy-tailed distribution ("super heavy-tailedness"). At the other extreme, a value higher than 4 is indicative of degree distributions whose fat-tail property is not very pronounced ("sub heavy-tailedness"). Finally, the theoretical HE value predicted by the preferential-attachment model of Barabasi and Albert (1999) is 3.

Table 5 reports the values of the Hill estimator for both the pooled and the sectoral R&D networks in all the time periods. Let us start with the pooled network. The table shows that the HE first decreases, reaching a minimum in the golden-age period 1994-1997 and then increases again. This indicates that the degree of tail-heaviness undergoes a rise-and-fall dynamics similar to the other network measures discussed so far. Moreover, the table shows that in all sub-periods the HE ranges between 2 and 4. This rules out both super and sub heavy-tailedness. However, in all sub-periods but the first and the last one the values of the HE is significantly below 3, and the minimum is achieved in the golden age period 1994-1997 (2.34). This indicates that in those periods the degree distribution of the pooled R&D network cannot be predicted by the preferential-attachment model. In particular, our results show that the tails of the degree distribution of the pooled R&D network are fatter than what will be predicted by that model.

The values of the HE computed on the sectoral R&D networks reveal a rise-and-fall pattern similar to the one detected in the pooled network (see Table 5). In particular, most sectors display fatter tails in the periods of higher alliance activity. Moreover, HE values of most manufacturing sectors are comparable to those of the pooled network. In contrast, HE values are in general higher in service sectors. This indicates that the concentration of alliances among few hubs is less marked in this type of sectors.

	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Pooled Network	3.04	2.31	2.34	2.61	2.78	3.05
Manufacturing Sectors						
Pharmaceuticals (283)	5.19	2.91	2.45	2.58	2.89	3.02
Computer Hardware (357)	2.70	2.37	2.22	2.75	2.88	4.59
Electronic Components (367)	3.36	2.43	2.43	2.25	2.59	3.57
Communications Equipment (366)	NaN	2.66	2.50	2.43	2.71	2.65
Medical Supplies (384)	NaN	3.71	3.25	4.50	3.58	3.95
Laboratory Apparatus (382)	NaN	2.69	2.73	3.70	3.22	4.04
Motor Vehicles (371)	3.69	2.18	2.46	2.87	3.72	3.98
Aircrafts and parts (372)	5.07	2.24	2.47	3.77	3.43	3.06
Inorganic Chemicals (281)	3.07	2.31	2.50	3.23	3.71	2.35
Household Audio-Video (365)	3.49	2.48	2.10	2.04	2.09	2.89
Plastics (282)	3.48	3.79	2.34	2.22	3.50	4.36
Electrical Machinery NEC (369)	NaN	3.04	2.89	3.61	4.38	3.29
Special Machinery (355)	NaN	2.89	3.35	3.82	4.44	NaN
Crude Oil and Gas (131)	NaN	3.39	4.08	4.16	6.22	3.59
Naut./Aeronaut. Navigation (381)	NaN	2.53	2.45	4.19	4.10	NaN
Organic Chemicals (286)	3.08	3.86	4.88	4.58	NaN	4.00
Service Sectors						
Computer Software (737)	2.71	2.41	2.30	2.70	3.31	4.24
R&D, Lab and Testing (873)	NaN	2.77	2.69	3.65	3.23	3.60
Telephone Communications (481)	4.63	2.81	2.69	2.94	3.07	3.25
Universities (822)	NaN	2.96	2.72	3.14	3.10	6.01
Investment Companies (679)	NaN	2.86	2.85	2.79	2.85	3.09
Professional Equipment Wholesale (504)	NaN	4.09	3.05	2.60	NaN	NaN
Engineer.,Architec.,Survey (871)	NaN	2.74	2.58	3.14	5.33	NaN
Motion Picture Production (781)	NaN	3.24	3.24	NaN	NaN	NaN
Management,Consulting,PR (874)	NaN	2.91	3.11	3.38	4.43	NaN
Radio and TV Broadcasting (483)	NaN	3.49	3.48	5.17	NaN	NaN
Cable and TV Services (484)	NaN	4.08	3.23	NaN	4.10	NaN
Business Services (738)	NaN	4.59	3.59	4.01	3.59	NaN
Electrical Goods Wholesale (506)	NaN	2.50	2.44	NaN	NaN	NaN
Electric Services (491)	NaN	2.97	3.81	4.01	3.71	NaN

Table 5: Hill estimator (HE) for degree distributions in the pooled and the sectoral R&D networks (SIC codes are in brackets). *Note*: missing values refer to sectors with not enough observations.

5 Assortative and Disassortative R&D Networks

Assortativity is a network measure that identifies correlations between the centrality of a node and the centrality of its neighbors. Assortativity can be computed by using any measure of node centrality (see e.g. Borgatti, 2005, for a survey of centrality measures). However, in this study we use degree correlation, or *average nearest-neighbor connectivity* (Newman, 2002; Pastor-Satorras et al., 2001) as assortativity measure. A network is assortative if it is characterized by a positive correlation across the degrees of linked nodes. This implies that nodes tend to be connected to nodes with similar degree. At the other extreme, dissassortative networks have negative node degree correlation, i.e. nodes tend to be connected to nodes with dissimilar degree. Newman (2003) found that technological networks, such as the internet, are disassortative while social networks, such as the network of scientific co-authorships, are assortative. However, R&D networks can be assortative or disassortative, depending on the underlying topology of the network. For instance, Ramasco et al. (2004) develop models wherein agents establish links with most central actors in the network, and show that such a mechanism gives rise to disassortative networks. However, König et al. (2010) show that the same mechanism of search for high centrality can give rise to assortative networks if agents face limitations in the number of collaborations a firm they are able to maintain.

To investigate assortativity-disassortativity in our R&D networks, we use the assortativity mixing coefficient r proposed by Newman (2002). This quantity, as described by Eq. 3, is the Pearson correlation coefficient of the degrees at both ends of all links in the network:

$$r = \frac{4M^{-1}\sum_i j_i k_i - [M^{-1}\sum_i (j_i + k_i)]^2}{2M^{-1}\sum_i (j_i^2 + k_i^2) - [M^{-1}\sum_i (j_i + k_i)]^2}, \tag{3}$$

where j_i, k_i are the degrees of the firms at the ends of the i-th link, with $i = 1, ..., M$. The coefficient r ranges between -1 for a totally disassortative network to 1 for a totally assortative network; a network in which links are formed randomly would exhibit $r = 0$. We compute the assortativity mixing coefficient r on both the pooled and the sectoral R&D sub-networks. We follow the same procedure as in the previous section. The whole observation period is again divided into six sub-periods of 4 years each and all the observations of every firm's degree are taken together within each sub-period. The degree correlation coefficients are then computed for each sub-period. The results are reported in Table 6.

The pooled R&D network is assortative, as indicated by the low but positive assortativity mixing coefficient during the whole observation period (see Table 6). This means that, on average, high-centrality (low-centrality) firms tend to connect to other high-centrality (low-centrality) firms. Moreover, and differently from the network indicators studied in Sections 3 and 4, the assortativity coefficient does not reveal any rise-and-fall dynamics over time.

	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Pooled Network	0.167	0.110	0.119	0.195	0.170	0.035
Manufacturing Sectors						
Pharmaceuticals (283)	0.005	0.172	0.119	-0.049	-0.047	-0.043
Computer Hardware (357)	-0.188	-0.179	-0.192	-0.133	-0.103	-0.145
Electronic Components (367)	-0.174	-0.151	-0.194	-0.094	0.023	0.267
Communications Equipment (366)	-0.233	-0.149	-0.147	-0.143	-0.077	-0.312
Medical Supplies (384)	NaN	-0.165	-0.155	0.106	-0.184	-0.108
Laboratory Apparatus (382)	NaN	-0.199	-0.134	-0.153	-0.159	0.018
Motor Vehicles (371)	-0.174	-0.309	-0.099	-0.071	-0.023	0.078
Aircrafts and parts (372)	-0.132	0.054	-0.182	0.035	0.019	0.804
Inorganic Chemicals (281)	-0.445	-0.228	-0.243	-0.188	-0.146	-0.239
Household Audio-Video (365)	-0.467	-0.368	-0.306	-0.329	-0.287	-0.342
Plastics (282)	-0.105	-0.249	-0.351	-0.437	-0.265	-0.151
Electrical Machinery NEC (369)	NaN	-0.250	-0.184	-0.283	-0.032	-0.134
Special Machinery (355)	NaN	-0.206	-0.223	-0.153	-0.214	-0.143
Crude Oil and Gas (131)	NaN	0.489	-0.017	0.383	0.255	-0.160
Naut./Aeronaut. Navigation (381)	NaN	-0.297	-0.318	-0.333	-0.217	-0.190
Organic Chemicals (286)	-0.458	-0.242	-0.206	-0.191	-0.190	-0.170
Service Sectors						
Computer Software (737)	-0.103	-0.074	-0.067	-0.029	-0.002	-0.105
R&D, Lab and Testing (873)	-0.024	-0.032	0.011	0.132	0.185	0.025
Telephone Communications (481)	-0.273	-0.178	-0.097	-0.035	-0.036	-0.279
Universities (822)	NaN	-0.133	-0.102	0.026	0.152	0.078
Investment Companies (679)	-0.057	-0.210	-0.193	-0.219	-0.187	-0.182
Professional Equipment Wholesale (504)	NaN	-0.128	-0.066	-0.168	-0.200	NaN
Engineer.,Architec.,Survey (871)	NaN	-0.275	-0.208	-0.130	-0.116	-0.207
Motion Picture Production (781)	NaN	-0.154	-0.081	-0.037	NaN	NaN
Management,Consulting,PR (874)	NaN	-0.288	-0.200	-0.221	-0.177	-0.135
Radio and TV Broadcasting (483)	NaN	-0.537	-0.173	-0.266	-0.250	-0.250
Cable and TV Services (484)	NaN	0.006	-0.101	-0.063	-0.287	NaN
Business Services (738)	NaN	-0.296	-0.247	0.382	0.087	-0.100
Electrical Goods Wholesale (506)	NaN	NaN	-0.305	-0.139	-0.100	-0.143
Electric Services (491)	NaN	-0.007	-0.235	-0.127	-0.107	0.664

Table 6: Assortativity mixing coefficient in the pooled and the sectoral R&D networks (SIC codes are in brackets). *Note*: missing values refer to sectors with not enough observations.

In contrast to the pooled R&D network, the sectoral R&D networks are disassortative: for most sectors and in most of the analyzed sub-periods, the assortativity coefficient is negative. For instance, when considering the 1990-1993 and the 1994-1997 periods, only 4 sectors out of 30 exhibit a non-negative assortativity coefficient (Pharmaceuticals, R&D-Lab-Testing, Aircrafts and Parts, Cable and TV Services). This indicates that in a sectoral R&D network, i.e. centered around a given industry, low-degree firms increase their tendency to connect to high-degree firms, and viceversa.

Thus, R&D networks seem to have features of both technological and social networks, as they display both assortativity and disassortativity depending on the scale at which they are studied. To shed more light on the determinants of this phenomenon, we study the “local degree correlations” in the pooled R&D network. More precisely, Fig. 5 shows the average neighbors’

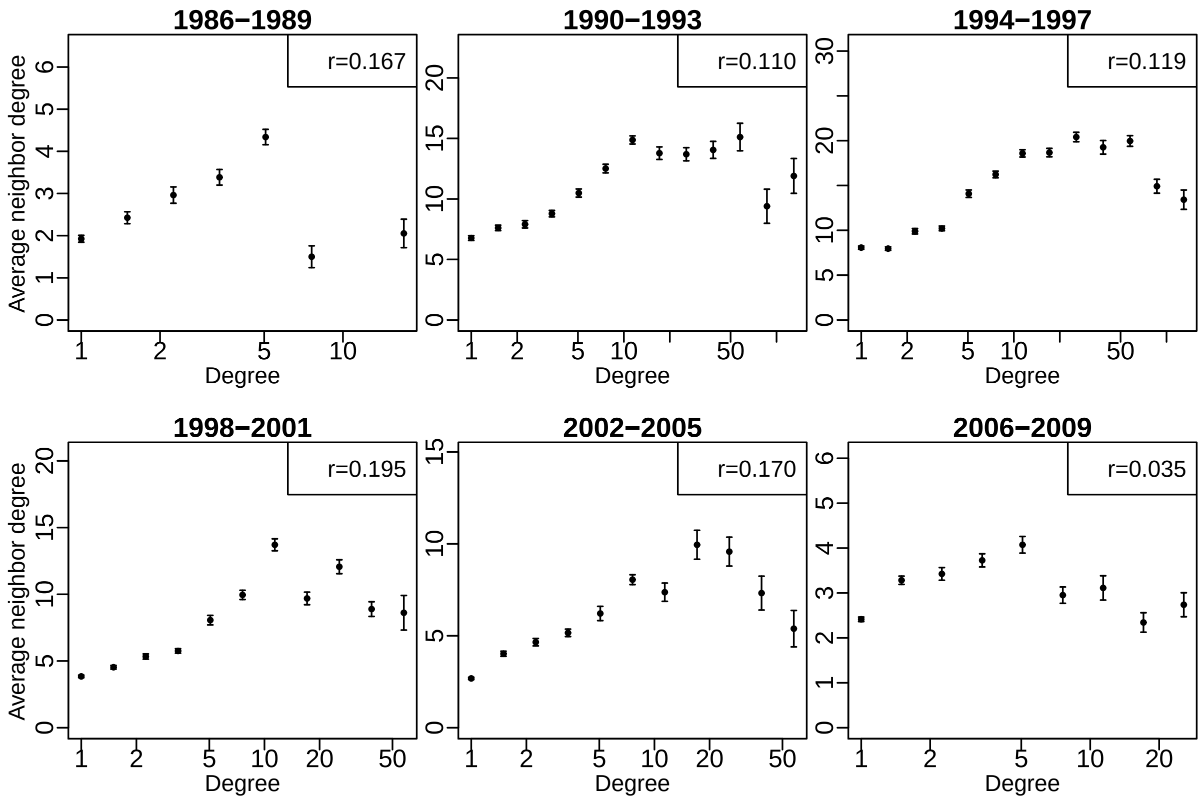

Figure 5: Local degree correlations (mean neighbors' degree VS degree) in the pooled R&D network. The error bars represent the standard error of the mean. *Note*: on the top-right corner of each plot we report the corresponding value of the assortativity mixing coefficient in the sub-period under analysis.

degree as a function of firms' degree, for the pooled R&D network, and for each of the six sub-periods considered in our analysis.

The plots show that the relation between average neighbors degree and node degree is strongly non linear in all the considered sub-periods. More precisely, node degree predicts quite well average degree of partners until high-degree nodes are taken into account. Then, a sharp decay occurs. This indicates that – when considering the pooled R&D network – firms with low and intermediate degree levels tend to connect with firms having similar degree, whilst high-degree firms display negative degree correlation. Moreover, the position of the maximum of these curves on the x-axis (i.e. the firm's degree) varies during the observation period and is positively correlated to the network size. Such a tipping point in the firm's degree is equal to 5 in the early period 1986-1989 and in the late sub-period 2006-2009, and it ranges between 10 and 20 in the other sub-periods. Interestingly, we find that the inverted U-shaped pattern of the local degree correlation curve holds for the sectoral R&D networks as well. The sharp decay in the local correlation curve is stronger in the sectoral R&D networks than in the pooled one.[5]. The above findings indicate that the transition from disassortativity to assortativity is the result of a composition effects of a non-linear relationship between the number of alliances of a firm and the one of its partners. In the pooled network sectoral hubs are poorly connected among them as indicated by the low average degree of their partners. In contrast, firms occupying low and intermediate positions in the sectoral degree distributions tend to form alliances with firms having similar degree in *other* sectors. This does not occur within sectors, where low- and intermediate-degree firms form alliances mainly with the sector hubs.

6 Small Worlds and Communities

Similarly to degree heterogeneity (cf. Section 4) the presence of *small worlds* in R&D networks has been analyzed by a large amount of theoretical and empirical works (see e.g. Cowan and Jonard, 2004, 2009; Fleming et al., 2007; Gulati et al., 2012; Uzzi et al., 2007). A network is a small world if it is characterized by two key features: high local clustering and low average path length (Watts and Strogatz, 1998). Local clustering measures the extent to which the neighbors of a node are in their turn connected among themselves. It is defined as the number of existing links between the neighbors of a focal node, divided by the number of all possible links between these neighbors; the measure is subsequently averaged over all nodes in the network. Average path length is defined as the average of all shortest distances, i.e. the lowest number of links that must be traversed to connect every pair of nodes in the network. In our R&D network representation, the first measure shows the extent to which a company's partners tend to be connected among themselves, while the second measure quantifies how long the average alliance chain from a firm to any other firm in the network is. Small world networks exhibit high clustering and short average path length, combining the qualities of both regular networks (typically characterized by high clustering and high average path length) and random networks (characterized by low clustering and low average path length). Previous empirical works have pointed out that the R&D network structure may follow a rise-and-fall dynamics. More specifically, Gulati et al. (2012) show that in the computer industry the excessive formation of ties can lead to the formation of a small world and then to its own decline.

Small world properties in a network are often associated with the presence of community structures (Newman, 2004b), reflecting the tendency of nodes to divide into groups or modules. In a modular network, dense connections and high clustering are observed within each group, with only a few links connecting the different groups (Newman, 2004a). In inter-firm networks,

[5]We do not show here the local degree correlations for the sectoral R&D networks, but data and plots are available upon request from the authors.

dense groups are shown to facilitate information exchange among similar firms and support trust and cooperative behavior, while bridging ties connecting different groups favour information recombination between distant positions in the knowledge space (e.g. Granovetter, 1973, 1983; Tiwana, 2008).

In this section we analyze both the presence of small worlds and community structures in R&D networks. According to Watts and Strogatz (1998), the small world properties of a network have to be evaluated using a corresponding random network as the baseline. If the examined network is both large and sparse, i.e. $N \gg \langle k \rangle$, where N is the network size and $\langle k \rangle$ is the average degree, the basic requirement for small world is satisfied. Under this assumption, the values of clustering coefficient C and average path length L for the baseline random network will tend to: $C_R = \langle k \rangle / N$ and $L_R = \ln(N)/\ln(\langle k \rangle)$. The small world quotient Q_{SW} we use for our analysis is defined as:

$$Q_{SW} = \frac{(C/C_R)}{(L/L_R)}. \tag{4}$$

In our study, the condition of sparse network is always fulfilled for the pooled and the sectoral R&D networks (the average degrees are always smaller than 3, and much smaller than the corresponding network sizes, as reported in Table 4). Some of the sectoral R&D networks have relatively small sizes in the first (1986-1989) and in the last (2006-2009) observation periods (as can be seen from Table 1), but in these cases they exhibit an even smaller average degree $\langle k \rangle$, still validating the assumption of sparse networks. When computing the observed to random ratios, a small world network will show $C/C_R \gg 1$ and $L/L_R \simeq 1$, which is the case for all the R&D networks we analyze. The results of our computations are listed in Table 7. Once again, results are presented for six different sub-periods.

The small world quotient is computed separately for every year during the whole observation period, in both the pooled and the sectoral R&D networks, and then averaged within six sub-periods lasting 4 years each.[6] The evolution of this quotient over time reveals the presence of a rise-and-fall dynamics of the small world properties, in both the pooled and the sectoral R&D networks. The small world quotient rises to a peak in the "golden age" period and then decreases again. Moreover, this feature is common across sectors, generalizing the results of the work by Gulati et al. (2012), that was limited to the computer industry. With the exception of 6 sectors out of 30 (Medical Supplies, Universities, Aircrafts and Parts, Business Services, Crude Oil and Gas, Eletric Services), the small world quotient has a peak either in the 1990-1993 or in the 1994-1997 period. It should also be noticed that five industrial sectors (Motion Picture Production, Management-Consulting-P.R., Electrical Goods Wholesale, Nautical/Aeronautical

[6]We do not aggregate the observations inside every time period, because the small world quotient is a global network measure, and not an ego-network measure centered around single nodes.

	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Pooled Network	1.410	85.814	154.560	57.085	28.640	5.596
Manufacturing Sectors						
Pharmaceuticals (283)	0.000	23.434	34.030	14.241	5.468	2.628
Computer Hardware (357)	0.129	4.757	16.864	6.397	0.635	0.000
Electronic Components (367)	0.000	7.082	12.414	6.691	5.450	2.290
Communications Equipment (366)	0.000	2.278	5.545	1.283	1.688	0.000
Medical Supplies (384)	NaN	0.000	0.368	0.000	0.000	0.000
Laboratory Apparatus (382)	NaN	0.976	0.534	0.000	0.000	0.933
Motor Vehicles (371)	1.740	2.924	4.134	0.669	1.863	0.840
Aircrafts and parts (372)	1.313	4.319	4.021	1.748	1.323	1.738
Inorganic Chemicals (281)	0.000	0.410	0.000	0.000	0.000	0.000
Household Audio-Video (365)	0.000	1.746	5.316	3.475	1.924	0.000
Plastics (282)	0.000	0.380	0.080	0.000	0.000	0.000
Electrical Machinery NEC (369)	NaN	0.000	0.000	0.000	0.000	0.000
Special Machinery (355)	NaN	0.000	0.000	0.000	0.000	0.000
Crude Oil and Gas (131)	0.000	2.004	0.775	1.269	2.002	0.000
Naut./Aeronaut. Navigation (381)	0.000	0.000	0.000	0.000	0.000	0.000
Organic Chemicals (286)	0.000	0.000	0.000	0.000	0.000	0.000
Service Sectors						
Computer Software (737)	0.769	13.584	33.514	5.242	0.669	0.000
R&D, Lab and Testing (873)	0.000	4.155	12.404	0.864	1.668	0.636
Telephone Communications (481)	0.000	7.521	10.110	1.448	1.222	0.000
Universities (822)	0.000	1.456	4.489	0.863	2.135	0.594
Investment Companies (679)	0.000	0.884	0.452	0.126	0.400	0.576
Professional Equipment Wholesale (504)	0.000	0.594	1.131	0.000	0.000	NaN
Engineer.,Architec.,Survey (871)	NaN	0.450	0.000	0.000	0.000	0.000
Motion Picture Production (781)	NaN	0.000	0.000	0.000	NaN	NaN
Management,Consulting,PR (874)	NaN	0.000	0.000	0.000	0.000	0.000
Radio and TV Broadcasting (483)	0.000	0.000	0.000	0.000	0.000	0.000
Cable and TV Services (484)	NaN	0.320	1.454	0.000	0.000	NaN
Business Services (738)	NaN	0.000	0.000	0.778	0.389	0.000
Electrical Goods Wholesale (506)	NaN	0.000	0.000	0.000	0.000	0.000
Electric Services (491)	NaN	0.429	0.000	0.000	0.594	2.377

Table 7: Small world quotient of pooled and sectoral R&D networks (SIC codes are in brackets), for the *giant component*. The values are averages within each sub-period. *Note*: missing values refer to sectors with not enough observations.

Navigation, Organic Chemicals) display constant zero values for their small world quotients, meaning that there is no observed clustering in the corresponding networks. The sectors that deviate the most from the non-monotonic small world dynamics are mostly service sectors, which indeed tend to create more inter-sectoral alliances, rather than forming their own intra-sectoral network.

We now want to assess whether such emergence of small world properties in R&D networks is associated with the presence of modular structures. The standard approach to quantify this phenomenon, described by Newman (2004b), is to perform a partition of the network into communities, i.e. assigning a label to every node, in order to maximize the so called *modularity coefficient*. Such indicator of modularity is maximum if the chosen network partition perfectly reflects the positioning of links in the network, with all links occurring within communities and no links occurring between different communities. We do not intend to test several partitions to maximize the modularity coefficient of the network. We rather assume that a community corresponds to an industrial sector. Next, we partition the pooled R&D network by assigning every firm to its sector. Finally, we study the time evolution of the modularity coefficient in the pooled R&D network, computed by considering the sectors as communities. This way, we are able to evaluate the extent to which alliances are concentrated among firms belonging to the same sector. We call the modularity coefficient Q_M and define the *relative connectivity* c_{ij} between two industrial sectors i and j as follows:

$$c_{ij} = e_{ij}/a_{ij}, \tag{5}$$

where e_{ij} is the fraction of links in the network connecting any firm belonging to sector i to any firm belonging to sector j. The quantities e_{ij} (and consequently a_{ij} and c_{ij}) can be thought of as elements of a symmetric $n \times n$ matrix, where n is the number of sectors into which the R&D network is partitioned.[7] The row (or column) sums $a_i = \sum_j e_{ij}$ represent the fraction of links (alliances) involving at least one company in sector i. We then define $a_{ij} = a_i a_j$ as the expected fraction of links connecting firms in sector i to firms in sector j in a benchmark network having the same density and sector populations as the real network, but where alliances occur randomly between firms, independently of the sector they belong to. This way, c_{ij} is the ratio between the observed and the expected fraction of alliances connecting a firm in sector i to a firm in sector j. Values of c_{ij} greater than 1 suggest that the alliance probability between a firm in sector i and a firm in sector j is higher than one would expect with a random partner choice. On the contrary, when c_{ij} is smaller than 1, a firm in sector i forms alliances with firms in sector j with a smaller probability than a random partner choice.

Next, following Newman (2004b), we define the modularity coefficient Q_M as:

$$Q_M = \sum_i (e_{ii} - a_{ii}^2)/(1 - \sum_i a_{ii}^2), \tag{6}$$

where the index i spans all industrial sectors in the R&D network. The coefficient Q_M is equal to 1 in case of a perfect modular network, where alliances occur only intra-community and never inter-community. Likewise, Q_M is equal to -1 for a perfect anti-modular network, having only inter-community links, without any intra-community links. Q_M is equal to zero for a network where links are formed at random. The time evolution of the modularity coefficient Q_M of the pooled R&D network is reported in Table 8.

[7] To make sure that every alliance is counted once in the matrix e_{ij}, every link connecting sectors i and j is split in half between the elements e_{ij} and e_{ji}.

	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Pooled Network	0.237	0.220	0.228	0.220	0.218	0.277

Table 8: Modularity coefficients for the pooled R&D network (SIC codes are in brackets). The values are averages within each sub-period.

The coefficient Q_M ranges between 0.21 and 0.28, indicating the presence of a moderate modularity if compared to other examples of real networks (see Newman and Girvan, 2004). Furthermore, the modularity coefficient exhibits only small changes over the observation period and does not have a peak in accordance with the peak of the small world quotient. The rise-and-fall of the small world structure detected above is thus not associated to any rise-and-fall in the modular structure of the network.

To conclude, our results generalize the previous findings of Gulati et al. (2012). The rise-and-fall of small worlds is not a feature limited to few industries, but it is instead a general feature of sectoral R&D networks. Moreover, this property emerges also when alliances are considered independently of the sector to which the firms belong to. However, small worlds are not associated with the presence of a strong community division of the network when industrial sectors are used as communities. The emergence of small worlds might thus have other reasons, which will be further investigated in the next section.

7 Core-Periphery Architectures

Core-periphery networks are dominated by one group of highly inter-connected nodes (the *core* of the network), that have few connections to secondary nodes (the *periphery* of the network). In addition, the peripheral nodes are strongly connected to the core nodes, but poorly inter-connected between each other. Borgatti (2005) points out that such kind of networks are efficient because they can spread information quickly. A generalization of the concept of core-periphery architecture is the one of *nested* networks. A network is nested if the neighbors of a node with degree m are contained in the neighborhoods of all nodes with degree $m' > m$. The difference with core-periphery networks is that graphs with a nested neighborhood structure can feature not only two groups (the *core* and the *periphery*), but several densely connected groups of nodes, with increasing degree. In addition, each group is connected to the group of higher degree nodes. König et al. (2012) show that efficient R&D networks (i.e. networks maximizing industry profits) have a nested architecture, when marginal costs of collaborations are high. Interestingly, both core-periphery and nested networks can exhibit short path length and high clustering features that are typical for small worlds. In our case, given the absence of correlation between the emergence of small worlds and modular architectures in the R&D networks, the formation of core-periphery architectures could be the true reason for the emergence of small world properties reported in Section 6.

To quantify the presence of core-periphery architectures in our R&D networks we employ a slightly modified version of the core-periphery coefficient C_{cp} suggested by Holme (2005).[8] More precisely, we define the core-periphery coefficient C_{cp} of a network G as follows:

$$C_{cp} = \frac{c_c\left[G^{core}\right]/c_c\left[G\right]}{c_c\left[G_R^{core}\right]/c_c\left[G_R\right]}, \tag{7}$$

where $c_c[\cdot]$ indicates the closeness centrality of a network[9] and G^{core} is a subgraph[10] of the network G that maximizes this value of closeness centrality. The ratio between the closeness centrality of G^{core} and the closeness centrality of G is then divided by the mean value of the same measure mean value computed on 500 random networks of the same size and density as the network G. The values of the core-periphery coefficients C_{cp} for the pooled and the sectoral R&D networks are shown in Table 9. Values are reported – as usual – for the different 6 sub-periods. We do not pool the observations inside each of the 6 selected sub-periods, but we compute the value of the core-periphery coefficient separately for every year and then average over the duration of every sub-period.[11]

We clearly observe a rise-and-fall trend for the core-periphery coefficient, in both the pooled and the sectoral R&D networks, with a peak positioned either in the 1990-1993 or in the 1994-1997 period. The presence of core-periphery structures in the "golden age" is a common characteristic across all industrial sectors. One notable exception is the Pharmaceutical sector, whose core-periphery coefficient has a peak in the period 2002-2005. In addition, four small industrial sectors (Management-Consulting-PR, Business Services, Electrical Goods Wholesale and Organic Chemicals) exhibit core-periphery coefficients that are not peaked neither in 1990-1993 nor in the 1994-1997 periods.

The above results confirm that – both at pooled and sectoral level – the small world prop-

[8]The difference is that we do not calculate the core-periphery coefficient only on the largest connected component of the network, but we take into account the whole network.

[9]The closeness centrality of a network is defined as the inverse of the sum of all shortest paths between any pair of nodes in the network. The idea behind this measure is to quantify how connected a network is. See Sabidussi (1966) for a more rigorous definition.

[10]There are many ways to divide a network G into subgraphs and then select the subgraph G^{core} with the maximal closeness centrality. Usually, one uses the computationally cheapest algorithm, which is a $k-$core decomposition of the network. G^{core} is then assumed to be the k-shell of the network with maximal closeness centrality. For the sake of brevity, we do not provide here any description of the $k-$core decomposition procedure, See Sabidussi (1966) for a detailed explanation, and Garas et al. (2012) for an extension to weighted networks.

[11]Similarly to the small world quotient, the core-periphery coefficient is not an ego-, but a global network measure (see Section 6).

	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Pooled Network	8.37	23.53	28.51	20.13	18.46	16.34
Manufacturing Sectors						
Pharmaceuticals (283)	0.97	11.41	7.59	12.69	12.88	3.81
Computer Hardware (357)	0.80	5.04	12.38	8.43	2.33	0.12
Electronic Components (367)	1.20	7.17	11.63	6.87	4.39	2.49
Communications Equipment (366)	0.19	3.13	11.05	4.11	2.30	0.77
Medical Supplies (384)	0.30	0.86	3.25	0.98	1.67	1.93
Laboratory Apparatus (382)	0.34	3.95	3.62	2.57	0.04	2.27
Motor Vehicles (371)	0.69	3.25	8.48	2.20	0.91	0.36
Aircrafts and parts (372)	1.87	5.75	9.12	3.18	1.01	1.88
Inorganic Chemicals (281)	0.17	3.60	1.89	0.06	0.07	0.09
Household Audio-Video (365)	0.50	2.98	10.12	5.43	2.96	1.95
Plastics (282)	0.28	2.37	3.80	0.07	0.08	0.16
Electrical Machinery NEC (369)	1.00	2.12	4.24	0.10	0.14	0.09
Special Machinery (355)	0.89	0.93	1.60	0.08	0.24	0.28
Crude Oil and Gas (131)	0.65	3.45	4.20	2.59	1.03	0.11
Naut./Aeronaut. Navigation (381)	1.00	0.78	1.20	0.14	0.19	0.24
Organic Chemicals (286)	0.63	0.11	0.06	0.15	0.13	0.16
Service Sectors						
Computer Software (737)	4.62	8.48	16.38	4.34	5.44	0.03
R&D, Lab and Testing (873)	0.16	4.01	11.36	10.69	5.47	0.49
Telephone Communications (481)	0.39	10.85	14.21	3.12	0.96	0.70
Universities (822)	0.81	0.95	8.47	2.35	1.93	2.35
Investment Companies (679)	0.26	5.75	7.84	7.68	4.91	2.33
Professional Equipment Wholesale (504)	0.58	1.94	2.75	0.15	0.38	0.35
Engineer.,Architec.,Survey (871)	1.00	3.12	2.85	0.05	0.15	0.19
Motion Picture Production (781)	NaN	0.78	1.45	0.30	0.66	0.55
Management,Consulting,PR (874)	1.00	0.19	0.03	0.05	0.05	0.12
Radio and TV Broadcasting (483)	1.00	2.40	2.62	0.13	0.45	0.64
Cable and TV Services (484)	NaN	0.81	4.43	0.36	0.48	0.78
Business Services (738)	1.00	0.46	0.04	1.25	0.72	0.67
Electrical Goods Wholesale (506)	NaN	0.23	0.04	0.18	0.29	0.38
Electric Services (491)	NaN	3.51	1.37	1.42	1.37	2.27

Table 9: Core-periphery coefficients for the pooled and the sectoral R&D networks (SIC codes are in brackets). The values are averages within each sub-period. *Note*: missing values refer to sectors with not enough observations.

erties detected in Section 6 are correlated to the presence of strongly centralized (core-periphery) architectures. Across sectors, firms show the tendency to organise their R&D collaborations in a core of densely connected companies and a periphery of companies that are linked to the core, but only weakly interconnected among themselves.

Next, we study whether the presence of core-periphery is related to the presence of a more general type of centralized architecture, i.e. nested architectures. In this way we also provide a test to some of the key predictions of the recent theoretical literature on R&D networks. There are several measures quantifying the extent to which a given network's neighborhood structure is nested. In this study, we use the measure generated by an algorithm called *BINMATNEST*[12].

[12]The *BINMATNEST* algorithm, proposed by Rodriguez-Girones and Santamaria (2006), uses the unweighted adjacency matrix of the network to compute its nestedness score. The algorithm rearranges the adjacency matrix

	1986-1989	1990-1993	1994-1997	1998-2001	2002-2005	2006-2009
Pooled Network	0.977	0.997	0.999	0.997	0.996	0.996
Manufacturing Sectors						
Pharmaceuticals (283)	0.960	0.989	0.996	0.994	0.995	0.996
Computer Hardware (357)	0.960	0.992	0.995	0.984	0.950	0.940
Electronic Components (367)	0.981	0.984	0.993	0.984	0.969	0.926
Communications Equipment (366)	0.943	0.962	0.990	0.973	0.944	0.962
Medical Supplies (384)	0.998	0.944	0.963	0.954	0.946	0.947
Laboratory Apparatus (382)	0.961	0.943	0.965	0.930	0.951	0.924
Motor Vehicles (371)	0.938	0.961	0.964	0.950	0.962	0.946
Aircrafts and parts (372)	0.945	0.942	0.969	0.973	0.953	0.974
Inorganic Chemicals (281)	0.930	0.978	0.951	0.951	0.943	0.977
Household Audio-Video (365)	0.951	0.945	0.981	0.964	0.957	0.963
Plastics (282)	0.977	0.940	0.966	0.951	0.975	0.949
Electrical Machinery NEC (369)	0.939	0.947	0.950	0.962	0.961	0.987
Special Machinery (355)	NaN	0.927	0.940	0.984	0.931	0.953
Crude Oil and Gas (131)	0.939	0.922	0.950	0.945	0.960	0.936
Naut./Aeronaut. Navigation (381)	0.939	0.938	0.948	0.936	0.982	0.998
Organic Chemicals (286)	0.956	0.938	0.961	0.922	0.939	0.956
Service Sectors						
Computer Software (737)	0.981	0.992	0.997	0.985	0.950	0.946
R&D, Lab and Testing (873)	0.961	0.969	0.992	0.986	0.986	0.975
Telephone Communications (481)	0.945	0.954	0.981	0.950	0.947	0.976
Universities (822)	0.961	0.971	0.973	0.952	0.948	0.958
Investment Companies (679)	0.956	0.973	0.979	0.961	0.962	0.940
Professional Equipment Wholesale (504)	0.911	0.930	0.930	0.952	0.921	0.998
Engineer.,Architec.,Survey (871)	NaN	0.924	0.917	0.962	0.957	0.998
Motion Picture Production (781)	NaN	0.935	0.925	0.923	0.937	NaN
Management,Consulting,PR (874)	NaN	0.932	0.933	0.954	0.959	0.939
Radio and TV Broadcasting (483)	0.939	0.941	0.969	0.942	0.967	0.958
Cable and TV Services (484)	NaN	0.930	0.925	0.975	0.973	NaN
Business Services (738)	0.901	0.926	0.952	0.954	0.972	0.998
Electrical Goods Wholesale (506)	NaN	0.951	0.940	0.953	0.935	0.998
Electric Services (491)	0.956	0.918	0.969	0.977	0.957	0.939

Table 10: Nestedness coefficients for the pooled and the sectoral R&D networks (SIC codes are in brackets). The values are averaged in six sub-periods. *Note*: missing values refer to sectors with not enough observations.

For every analyzed network, the algorithm returns a nestedness score T_n, ranging from 0 (for a totally nested network) to 100 (for a completely random, non-nested network). In order to have a benchmark, the algorithm also builds and analyzes 500 random networks having the same size and density as the considered network. Instead of directly using the value generated by the

in such a way that all the "ones" (existing links) are concentrated in the top-left side of the matrix, and the "zeros" (missing links) in the bottom-right side. It then computes the optimal theoretical isocline separating the "ones" from the "zeroes" and counts the number of holes in these regions of the matrix – i.e. how many "zeroes" are in the region of the "ones", and viceversa. The number of such holes is proportional to the nestedness score computed by the algorithm: the more holes, the higher the nestedness score of the network. *Note:* the lower this score, the more nested the network is (and viceversa).

algorithm, we use a normalized nestedness coefficient C'_n, defined as:

$$C'_n = \frac{100 - T_n}{100}, \tag{8}$$

where T_n is the nestedness score generated by the *BINMATNEST* algorithm. Our normalized nestedness coefficient C'_n spans thus from 0, for a for a totally non-nested network, to 1, for a totally nested network. We calculate the coefficients C'_n throughout the whole observation period, for the pooled and the sectoral R&D networks, and average the results within six sub-periods lasting 4 years each. Results are shown in Table 10.

The values of the nestedness coefficients C'_n we report are extremely close to 1, during the whole observation period, both for the pooled and the sectoral R&D networks. This is surprising, if we compare such values with other studies of nestedness in real networks (e.g. Bascompte et al., 2003). All the values found in our R&D networks are significantly different from the average values of the random networks used as benchmark in the *BINMATNEST* algorithm. Moreover, the nestedness coefficient has a peak during the "golden age" for the pooled R&D network, as well as 9 out of 16 manufacturing sectors and 6 out of 14 service sectors. These results confirm not only that the pooled and the sectoral R&D networks are significantly nested throughout all the observation period, but also that their nestedness tends to increase during the "golden age", in correspondence to the emergence of the small world properties.

8 Implications

Three main implications arise from the evidence discussed in the previous sections.

First, our results provide strong support to the claim that several properties of R&D networks robustly hold across several manufacturing and service sectors. In addition, these properties are invariant across different scales of aggregation. In other words, they are the same if one considers the R&D alliances irrespectively of the sectors to which the firms belong (pooled network), or if one considers only alliances centered on a sector (sectoral networks). These properties do not only relate to basic characteristics of the networks like size, density, degree distributions. They also involve more complex features such as the presence of the small worlds and core-periphery architectures. From an empirical perspective, our results thus generalize previous findings in the literature, that were limited to the analysis of few sectors. From a theoretical perspective, the fact that many properties of the network hold irrespectively of the sector and of the scale of aggregation opens up the fascinating possibility that the same universal mechanism can be responsible for the emergence of those features. In this respect, our results also show that such a mechanism is probably different and more sophisticated than the preferential attachment described by Barabasi and Albert (1999). This is because the characteristics of the degree distribution observed in our R&D networks (cf. Section 4) can be hardly reconciled with the predictions of that model. Nevertheless, our results also show that not *all* properties of the network are invariant across different scales of aggregation. Sectoral networks are disassortative, i.e. characterized by a negative correlation across nodes degree, whereas the pooled network is assortative. This transition from disassortative to assortative networks is a fresh new stylized fact that should be taken into account in the theoretical explanations of R&D networks. It is important to remark that the contrast between disassortative and assortative networks has been so far stressed in relation to networks belonging to different domains (e.g. technological vs. social networks, cf. Newman, 2003). Our results suggest instead that the same type of network (network of R&D alliances) can be disassortative or assortative depending on the scale at which it is observed (i.e. taking into account the sectoral characteristics of the partners or not).

Second, the result that both the pooled and sectoral networks are organized into core-periphery architectures, and nested structures in particular, militates in favor of the predictions of the recent theoretical literature on R&D networks (e.g. Goyal and Joshi, 2003; Westbrock, 2010), and more precisely of the knowledge-recombination model of König et al. (2012). In this model, the efficient network structure is shown to critically depend on the marginal cost of R&D collaborations. In case of relatively costly partnerships, the resulting efficient R&D network exhibits a strongly nested neighborhood structure, as we observe empirically. Moreover, the presence of core-periphery architectures is also able to explain two network properties that received a lot of attention in the literature, namely the emergence of fat-tailed degree distributions and of small worlds. These properties are indeed the result of the organization of the R&D networks into core-periphery structures and cannot be instead related to other types of network characteristics (e.g. the presence of communities for small worlds, as we show in Section 6).

Third, our evidence indicates that the last three decades have witnessed a rise and fall of R&D networks. The foregoing rise-and-fall dynamics was previously emphasized in relation to the presence of small worlds in the computer industry (Gulati et al., 2012). We show that it is instead a general property of the R&D network dynamics (both sectoral and pooled ones). In addition, it concerns most network properties, even the more complex ones (presence of core-periphery and nested architectures).[13] Our results also show that the rise and fall of R&D networks was mainly driven by the entry and exit of firms participating into alliances rather than by the more or less intense activity of the incumbents (see Section 3). Moreover, during the growing phase, R&D alliances gave rise to network components of large size displaying the complex features discussed above. In the descending phase, the number of firms participating

[13] Indeed, the only exception to this general dynamics is represented by the assortativity (resp. disassortativity for sectors), that does not display any particular trend over time.

into alliances plummeted and the networks broke up into several components of small size.

Overall, the above facts suggest that theoretical explanations of the dynamics of the network should account for a significant role of the entry/exit of firms. In addition, they should be able to explain the ability of the network to self-organize into components having complex characteristics and the eventual breaking-up of them. Finally, as it is argued in Gulati et al. (2012), the rise and fall of R&D networks could be the sheer outcome of the knowledge recombination process associated with alliances embedded into a network. Indeed, the possibility of knowledge recombination fuels the growth of the network, either by combining heterogeneous knowledge bases (e.g. Cowan and Jonard, 2004; Gulati et al., 2012) or by granting access to multiple paths through which knowledge can reach the firm (König et al., 2011). The same process of knowledge recombination may however set the the premises for the subsequent breaking-up of the network. This is because recombination brings homogeneity into knowledge bases, consequently reducing the incentive for knowledge exchange and thus for alliance formation (Cowan and Jonard, 2004; Gulati et al., 2012). Likewise, in a large network, the number of additional paths to which a firm gets access with an alliance is higher if the alliance is created with a firm which is already part of its component (i.e. if the potential partner is already indirectly connected to the firm). In a situation where alliances are costly, this reduces the incentives to maintain bridging ties, thus contributing to the fragmentation of the network into many clusters which are sparsely connected among themselves (see König et al., 2011, for a model generating a similar dynamics).

9 Concluding Remarks

In this study we empirically investigated the dynamics and the properties of networks of R&D alliances in the period from 1986 to 2009. We studied both the pooled network, i.e. the network wherein alliances are considered independently of the sectors of the partners, as well as sectoral networks for a large number of manufacturing and service sectors. We show that many properties of the network robustly hold across different sectors. In addition, they are invariant at different levels of aggregation, i.e. when moving from the pooled R&D network to the sectoral ones. In particular, both pooled and sectoral R&D networks are organized into nested architectures displaying small world properties and fat-tailed degree distributions. However, not all R&D network properties are invariant across aggregation scales. In particular, we show that the pooled network is assortative, i.e. characterized by a positive correlation between the number of alliances of a firm and the one of its partners, whereas the opposite (disassortativity) occurs at the level of sectoral networks. Finally, we show that both the pooled and the sectoral R&D networks are characterized by a "rise-and-fall" dynamics. More precisely, during the period we analyzed, R&D networks showed a growing phase (1989-1997), where an increasing number of firms gave rise to network components having complex features, and then a decline (1998-2009) characterized by a breaking-up of the networks into smaller components.

Our work could be extended at least in two ways. First, one could track the evolution of the creation/destruction of single R&D alliances and test whether this can be predicted based on some topological properties of the existing networks, according to theoretical models of strategic network formation. Second, one could augment our data with information about firms' patenting. This would allow one to measure the evolution of firm technology portfolios as a consequence of R&D alliance formation.

Acknowledgements

We thank Claudio Tessone, Moritz Müller, Stefano Battiston, Lionel Nesta, Giorgio Fagiolo, Nobuyuki Hanaki, Sanjeev Goyal, Pier-Paolo Saviotti, and Pier-Paolo Andriani, for useful comments and discussions. In addition, we thank the participants to the ECCS 2012 Conference, in Brussels (Belgium), to the LATSIS 2012 Symposium, in Zürich (Switzerland), to the EMAEE 2013 conference, in Sophia-Antipolis (France), and to the WEHIA 2013 conference, in Reykjavík (Iceland), M. V. T, acknowledges financial support from the SNF through grant 100014_126865. A. G. and F. S. acknowledge support by the EU-FET project MULTIPLEX 317532.

References

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

Arora, A. and Gambardella, A. (1994b). Evaluating technological information and utilizing it : Scientific knowledge, technological capability, and external linkages in biotechnology. *Journal of Economic Behavior & Organization*, 24(1):91–114.

Barabasi, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286:509–512.

Bascompte, J., Jordano, P., Melin, C. J., and Olesen, J. M. (2003). The nested assembly of plantanimal mutualistic networks. *Proceedings of the National Academy of Sciences*, 100(16):9383–9387.

Borgatti, S. (2005). Centrality and network flow. *Social Networks*, 27(1):55–71.

Bottazzi, G. and Secchi, A. (2003). Common properties and sectoral specificities in the dynamics of u.s. manufacturing companies. *Review of Industrial Organization*, 23(3):217–232.

Cantner, U. and Graf, H. (2006). The network of innovators in Jena: an application of social network analysis. *Research Policy*, 35(4):463–480.

Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4):661–703.

Cowan, R. and Jonard, N. (2004). Network structure and the diffusion of knowledge. *Journal of Economic Dynamics and Control*, 28(8):1557–1575.

Cowan, R. and Jonard, N. (2009). Knowledge portfolios and the organization of innovation networks. *Academy of Management Review*, 34(2):320–342.

Deeds, D. and Hill, C. (1999). An examination of opportunistic action within research alliances-the analysis of discrete structural alternatives. *Journal of Business Venturing*, 14(2):141–163.

Fleming, L., King, C., and Juda, A. I. (2007). Small worlds and regional innovation. *Organization Science*, 18(6):938–954.

Fleming, L. and Marx, M. (2006). Managing creativity in small worlds. *California Management Review*, 48(4):6–27.

Fruchterman, T. M. J. and Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11):1129–1164.

Garas, A., Schweitzer, F., and Havlin, S. (2012). A k -shell decomposition method for weighted networks. *New Journal of Physics*, 14(8):083030.

Goyal, S. and Joshi, S. (2003). Networks of collaboration in oligopoly. *Games and Economic behavior*, 43(1):57–85.

Goyal, S. and Moraga-Gonzalez, J. L. (2001). R&D networks. *RAND Journal of Economics*, 32(4):686–707.

Granovetter, M. S. (1973). The strength of weak ties. *The American Journal of Sociology*, 78:1360–1380.

Granovetter, M. S. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory*, 1:201–233.

Gulati, R., Sytch, M., and Tatarynowicz, A. (2012). The rise and fall of small worlds: Exploring the dynamics of social structure. *Organization Science*, 23(2):449–471.

Hagedoorn, J. (2002). Inter-firm R&D partnerships: an overview of major trends and patterns since 1960. *Research policy*, 31(4):477–492.

Hanaki, N., Nakajima, R., and Ogura, Y. (2010). The dynamics of R&D network in the IT industry. *Research policy*, 39(3):386–399.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):pp. 1163–1174.

Holme, P. (2005). Core-periphery organization of complex networks. *Phys. Rev. E*, 72:046111.

Jackson, M. O. and Rogers, B. W. (2007). Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97(3):890–915.

König, M. D., Battiston, S., Napoletano, M., and Schweitzer, F. (2011). Recombinant knowledge and the evolution of innovation networks. *Journal of Economic Behavior & Organization*, 79(3):145–164.

König, M. D., Battiston, S., Napoletano, M., and Schweitzer, F. (2012). The efficiency and stability of R&D networks. *Games and Economic Behavior*, 75(2):694 – 713.

König, M. D., Tessone, C. J., and Zenou, Y. (2010). From assortative to dissortative networks: the role of capacity constraints. *Advances in Complex Systems*, 13(04):483–499.

König, M. D., Tessone, C. J., and Zenou, Y. (2013). Nestedness in networks: A theoretical model and some applications. Forthcoming in Theoretical Economics.

Lee, Y., Amaral, L. A. N., Canning, D., Meyer, M., and Stanley, H. E. (1998). Universal features in the growth dynamics of complex organizations. *Physical Review Letters*, 81(15):3275.

Lissoni, F., Llerena, P., and Sanditov, B. (2013). Small worlds in networks of inventors and the role of academics: An analysis of france. *Industry and Innovation*, 20(3):195–220.

Newman, M. E. J. (2002). Assortative mixing in networks. *Physical Review Letters*, 89(20):208701.

Newman, M. E. J. (2003). Mixing patterns in networks. *Physical Review E*, 67(2):026126.

Newman, M. E. J. (2004a). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101(90001):5200–5205.

Newman, M. E. J. (2004b). Detecting community structure in networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 38(2):321–330.

Newman, M. E. J. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113.

Pastor-Satorras, R., Vazquez, A., and Vespignani, A. (2001). Dynamical and correlation properties of the internet. *Physical Review Letters*, 87.

Phelps, C. (2003). *Technological exploration: A longitudinal study of the role of recombinatory search and social capital in alliance networks*. PhD thesis, New York University, Graduate School of Business Administration.

Powell, W., White, D., Koput, K., and Owen-Smith, J. (2005). Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. *American journal of sociology*, 110(4):1132–1205.

Powell, W. W., Koput, K. W., and Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, 41(1):116–145.

Ramasco, J. J., Dorogovtsev, S. N., and Pastor-Satorras, R. (2004). Self-organization of collaboration networks. *Physical Review E*, 70(3):036106.

Reed, W. (2001). The Pareto, Zipf and other power laws. *Economics Letters*, 74(1):15–19.

Riccaboni, M. and Pammolli, F. (2002). On firm growth in networks. *Research Policy*, 31(8-9):1405–1416.

Rodriguez-Girones, M. A. and Santamaria, L. (2006). A new algorithm to calculate the nestedness temperature of presence-absence matrices. *Journal of Biogeography*, 33(5):924–935.

Rosenkopf, L. and Schilling, M. (2007). Comparing alliance network structure across industries: observations and explanations. *Strategic Entrepreneurship Journal*, 1(3-4):191–209.

Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, 31(4):581–603.

Schilling, M. (2009). Understanding the alliance data. *Strategic Management Journal*, 30:233–260.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440.

Tiwana, A. (2008). Do bridging ties complement strong ties? An empirical examination of alliance ambidexterity. *Strategic Management Journal*, 29(3):251–272.

Uzzi, B., Amaral, L. A., and Reed-Tsochas, F. (2007). Small-world networks and management science research: a review. *European Management Review*, 4(2):77–91.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, 393:440–442.

Westbrock, B. (2010). Natural concentration in industrial research collaboration. *The RAND Journal of Economics*, 41(2):351–371.