NETWORKS OF INVENTORS AND ACADEMICS IN FRANCE

FRANCESCO LISSONI [1,2], *PATRICK LLERENA* [3], *BULAT SANDITOV* [3,4]

[1] Brescia University, [2] KITeS – Bocconi University, [3] BETA – University of Strasbourg, [4] UNU-MERIT, Maastricht University

(PRELIMINARY DRAFT. PLEASE DO NOT QUOTE WITHOUT PERMISSION FROM THE AUTHORS)

Abstract

The paper describes networks of inventors and the position of academic inventors in France. The data on university inventors from the KEINS database complemented by new dataset on the inventors from CNRS covers all patent applications filed at the European Patent Office from 1978 to 2004, and identifies who, among the inventors, were academic scientists in active service in 2004-2005 (KEINS). Structural properties of the networks of inventors and their dynamics are explored. Within such networks, academic scientists occupy central positions, as they stand in between other inventors, either taken as individuals or teams.

1. Introduction

University patents have been studied intensively over the past twenty years. Recent contributions from Europe have revealed that, no matter how small patent portfolios of European universities may be compared to those of their US counterparts, European academic scientists contribute greatly to inventions that are then patented by business companies and, to a lesser extent, by large public research organizations (Verspagen, 2006, Lissoni et al. 2008).

However, the role of academic scientists who contribute to patenting as inventors matters not only for the sheer number of the patents they produce, and their weight on total patents in a country and/or technological field. More important it is the relationship that academic scientists entertain with their co-inventors, many of whom come from industry or were students at the time when the patent was taken. More broadly, one may wish to investigate the academic inventors' standing in the overall technological community.

Such standing and the interaction with co-inventors may be revealing of how much information and knowledge academic inventors are in the position to pass on, or absorb from, technologists active in the invention process. Are academic inventors central to the communities of researchers contributing to the advancement of a given field? Is their presence necessary to connect people who would otherwise never exchange information and knowledge?

In this paper we build upon Balconi's et al. (2004) methodology in order to map the networks of inventors in France. We exploit the French section of the KEINS database complemented with the novel dataset on CNRS inventors collected following the KEINS methodology.

We begin with re-examination of earlier findings that social networks of inventors are "small worlds", having prominent local structure as characterized by high clustering coefficients and short social distances. We show that examining the "small world" features of social networks constructed from the "event-actor" data (patents-inventors data in our case) a researcher must carefully choose the proper random network against which small worlds properties of the observed network are tested. We find that inventors' networks in some technological fields are, indeed, small worlds, while in other fields not (too fragmented). Small worlds are found in those technological fields where presence of academic and CNRS inventors is significant.

We then apply a set of measures derived from both the classical and recent literature on social networks (Wasserman and Faust 1994; Blondel et al. 2008), in order to further explore both the structural properties of the network and the position and role of academic inventors therein. We find that that both academic and CNRS inventors are highly central actors both locally and globally. We use a modularity algorithm to uncover the block-model structure of the network, and find that the boundaries of tightly-knit clusters of inventors of which inventors' networks consist often coincide with firms' boundaries. This may explain why academic and CNRS inventors which are more likely to behave as "free lancers" for different firms exhibit higher centrality scores. We, indeed, find that academic and CNRS inventors are more likely to occupy important positions between clusters.

In what follows we briefly survey the existing literature on academic inventors (section 2) and describe the data (section 3). We then proceed to analyse the structural properties of inventors' networks (section 4) and the position of academic inventors (section 5). The last section summarize our results and draw some conclusions.

2. Background literature

3. Data

Data for building the networks of French inventors come from the EP-INV database produced at KITeS-Università Bocconi, which contains all the patent applications filed at the European Patent Office (EPO) since 1978, reclassified by applicant and inventor. Originally based upon legal information produced by EPO for patent attorneys, the EP-INV database is now currently updated by reading, cleaning, and matching applicants' and inventors' names and addresses as published in PATSTAT, the EPO worldwide statistical database[1]. The methodology of reclassification by inventor is explained in detail in Lissoni et al. (2006) can be summarized as follows:

- First, names and addresses of inventors have been standardized (in order to assign a unique code to all inventors with the same name, surname, and address);
- Second, for all pairs of inventors from the same country with the same name and surname, but different addresses “similarity scores” have been calculated;
- Third, a threshold value for the similarity score, over which two inventors in a pair are considered the same individual, was identified and all pairs having the score above the threshold value were assigned the same unique inventor’s code[2].

An important subset of the EP-INV database is the KEINS database, which identifies "academic inventors" and provide information on their affiliation as well as (if available, depending on the country) on discipline, gender, date of last promotion, and date of birth. Academic inventors are identified by matching inventors' names with names of professors on active service either on 2004 or 2005, depending on the country. Matches were then filtered by emailing or phoning the matched academics, asking for confirmation of their inventor status (see Lissoni et al., 2006, for methodological details; see also Lissoni et al., 2008).

The KEINS database covers not only the patents signed by academics and filed by their universities, but also all patents signed by academics and filed by companies, public or private research organizations, government, and individuals (such as the inventors themselves). For France, the KEINS

[1] EPO Worldwide Patent Statistical Database (http://forums.epo.org/epo-worldwide-patent-statistical-database/)
[2] In the case of France, the threshold was set at the median value of the similarity score distribution for all French inventors

database contains information on academic inventors from the hard sciences, medicine and engineering, who were active as *maitre de conference* or professors in any university of the country in 2004 and at least one patent signed after 1993 and before 2005. This make the data increasingly prone to return negatively biased estimates of academic patenting activity the farther back we go in time and highly unreliable for years before 1994 or after 2004.

For the purposes of this and related papers, the French section of the KEINS database was complemented with information on CNRS inventors, also collected following the KEINS methodology. Data from CNRS used for name matching include all *chercheurs*, *ingénieurs de recherche*, and technical staff on duty in 2007, in all the hard sciences as well as medical and engineering disciplines. This means that CNRS patent data suffer more of academic patent data of underestimation for early years. [Llerena 2010, (Thibaut 2009, Guarisco 2009 - ?)].

Table 1 reports the populations of patent applications, inventors, academics and CNRS researchers, as from the database described above (for France). It also reports the number of academic and CNRS inventors, and their weight over the total population of university professors and CNRS researchers, which stand respectively at 3.75% and 4.72%. We observe that the percentage of academic inventors over the total tenured academic staff is similar to what found for other countries in the KEINS database (Lissoni et al. 2008; Lissoni et al. 2009), while the share of CNRS inventors among all CNRS researchers is somewhat higher. This may be due to composition effects: as opposed to CNRS, universities host many tenured staff members who are not full-time researchers or do not perform any research at all, as they are mostly engaged in teaching and, in the case of the medical sciences, clinical activities.

Table 1 EPO patent applications, inventors, academics, and CNRS researchers

Inventors						
Nr. inventors (1994-2004)	*All academics (1)*	*Academic inv. (2)*	*(2)/(1)*	*CNRS research. (3)*	*CNRS inv. (4)*	*(3)/(4)*
51403	32006	1201	3.75%	15570	735	4.72%

Patents						
Nr. Patents (5)	*Academic patents (6)*	*(6)/ (5)*	*CNRS researchers' patents, (7)*	*(7)/(5)*	*(6)+(7) **	*(6)+(7) / (5)*
78800	2715	3.45%	1715	2.18%	4044	5.13%

(1) Academics active in 2005 in a French universities, the hard sciences, medicine, and engineering (source: French Ministry of University)
(2) Subset of (1), who have signed at least 1 EPO patent application in 1994-2004 (source: KEINS database)
(3) Researcher active in 2007 in CNRS, in the hard sciences, medicine, and engineering (source: CNRS)
(4) Subset of (3), who have signed at least 1 EPO patent application in 1994-2004 (source: elaboration on CNRS and EP-INV data)
(5) Nr of EPO patents signed by at least one inventor with French address with at least one patent in 1994-2004 (source: EP-INV database)
(6) Nr of EPO patents signed by at least one French academic with at least one patent in 1994-2004 (source: KEINS database)

(7) Nr of EPO patents signed by at least one CNRS researcher with at least one patent in 1994-2004 (source: elaboration on CNRS and EP-INV data)
* When computing (6)+(7), care was taken to avoid double counting the patents signed by both academics and CNRS researchers.

Academic inventors are responsible for 2715 patent applications and CNRS researchers for 1715 applications, which amount respectively to 3.45% and 2.18% of all domestic patent applications at EPO. In total all academic patents (both university and the CNRS) account for 4044 patent applications, which is 5.13% of all patent applications filed at EPO by French inventors. Notice that the latter figure is less than the sum of the previous two. This is because several patents are signed jointly by academics and CNRS researchers, so we counted them only once in the calculation.

Table 2 EPO patents by technology: all, university, CNRS, and academic (university & CNRS)

		INVENTORS						
	TECHNOLOGICAL FIELDS	*All inv. (1)*	*Ac. inv. (2)*	*(2)/(1)*	*CNRS inv (3)*	*(3)/(1)*	*(2)+(3)*	*(2)+(3) / (1)*
1	*Electrical engineering. Electronics*	13610	217	1.59%	123	0.90%	340	2.50%
2	*Instruments*	9714	363	3.74%	178	1.83%	541	5.57%
3	*Chemicals. Materials*	8653	336	3.88%	259	2.99%	595	6.88%
4	*Pharmaceuticals. Biotechnology*	5980	396	6.62%	280	4.68%	676	11.30%
5	*Industrial processes*	8159	153	1.88%	97	1.19%	250	3.06%
6	*Mech. Eng. Machines. Transport*	10386	64	0.62%	22	0.21%	86	0.83%
7	*Consumer goods. Civil eng.*	5158	12	0.23%	5	0.10%	17	0.33%
		PATENTS						
		All pat. (4)	*Ac. pat. (5)*	*(5)/(4)*	*CNRS pat (6)*	*(6)/(4)*	*(5)+(6)*	*(5)+(6) / (4)*
1	*Electrical engineering. Electronics*	18237	385	2.11%	167	0.92%	504	2.76%
2	*Instruments*	10164	513	5.05%	210	2.07%	658	6.47%
3	*Chemicals. Materials*	12157	801	6.59%	654	5.38%	1336	10.99%
4	*Pharmaceuticals. Biotechnology*	7346	713	9.71%	523	7.12%	1119	15.23%
5	*Industrial processes*	10043	195	1.94%	126	1.25%	290	2.89%
6	*Mech. Eng. Machines. Transport*	13796	91	0.66%	28	0.20%	113	0.82%
7	*Consumer goods. Civil eng.*	7057	17	0.24%	7	0.10%	24	0.34%

(1) All inventors who have signed at least 1 EPO patent application in 1994-2004 (EP-INV data)
(2) Academics active in 2005, 2005 in a French universities (sciences, medicine, engineering) who have signed at least 1 EPO patent application in 1994-2004 (source: KEINS database)
(3) Researcher active in 2007 in CNRS, in the hard sciences, medicine, and engineering who have signed at least 1 EPO patent application in 1994-2004 (source: elaboration on CNRS and EP-INV data)
(4) Nr of EPO patents signed by at least one inventor with French address with at least one patent in 1994-2004 (source: EP-INV database)
(5) Nr of EPO patents signed by at least one French academic with at least one patent in 1994-2004 (source: KEINS database)
(6) Nr of EPO patents signed by at least one CNRS researcher with at least one patent in 1994-2004 (source: elaboration on CNRS and EP-INV data)
* When computing (5)+(6), care was taken to avoid double counting the patents signed by both academics and CNRS researchers.

Table 2 provides a breakdown of these figures by technological field of patents (based upon the re-classification of IPC codes proposed by OST, 2004). It shows that the most science-intensive fields (*Instruments*, *Chemicals & Materials*, and *Pharmaceuticals & Biotechnology*) are also those where public research is the most relevant for inventive activity. In *Chemistry & Materials* the share of all academic (both university and CNRS invented) patents is about 11%; in *Biotechnology & Pharmaceuticals* the weight of academic patents over the overall national patenting activity is above 15%. As for *Electrical engineering & Electronics* and *Instruments,* the percentage value are lower not because of a lack of academic patents (as in *Mechanics* or *Consumer Goods*), but because of the large number of total patent applications in the fields.

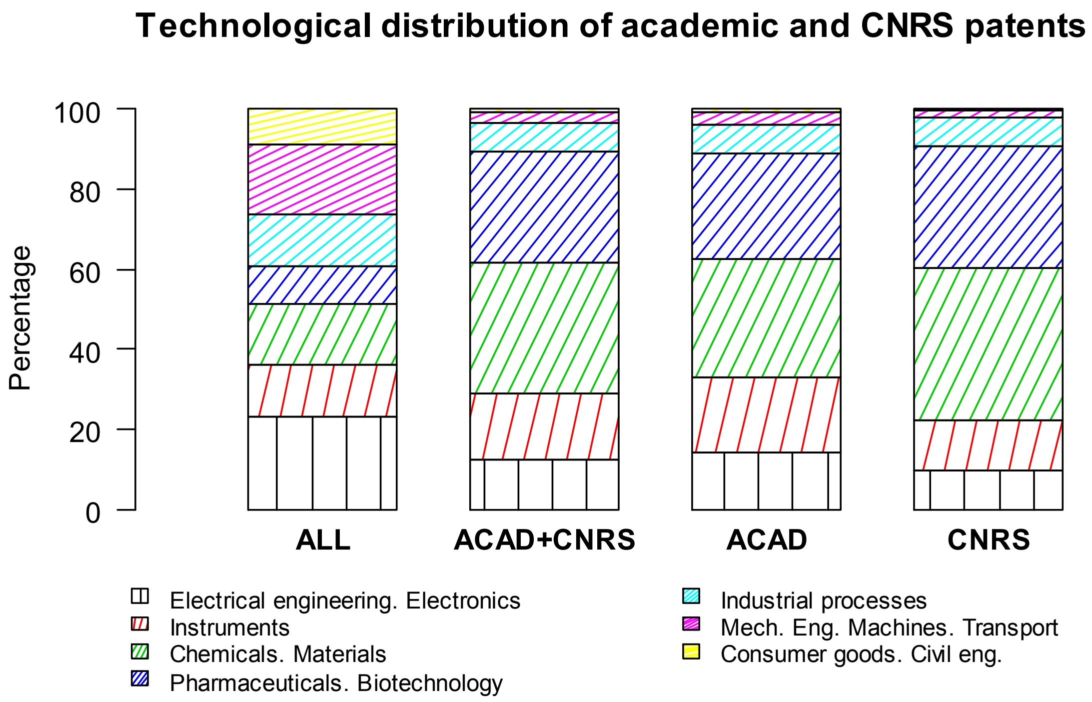


Figure 1 Distribution of patents across technological fields.[3]

The technological distribution of academic patents (Figure 1) resembles closely previous findings for the US (Mowery and Sampat, 2005). Academic patents concentrate in few technological fields: *Chemicals & Materials*, and *Biotechnology & Pharmaceuticals* and to less extent in *Scientific and Control Instruments*. Comparison of the CNRS-originated patents vis-à-vis university patents reveals that CNRS researchers tend to patent more in *Chemicals & Materials* and *Biotechnology & Pharmaceuticals*, while less in *Electrical engineering & Electronics* and *Instruments*.

Inspection of the applicants' identity reveals that by and large academic patents are owned by business companies, rather than universities or other public research organizations. Figure 2

[3] "ALL" - all inventors, "ACAD+CNRS"=academic and CNRS inventors, "ACAD" - academic inventors, "CNRS" - CNRS inventors.

describes the ownership distribution of academic patents in France. In about 14% cases patents are co-owned by applicants of different types (e.g. public research centre and university). These patents were counted as many times as there were different types of owners. Almost 60% of patent applications by either academic or CNRS inventors are owned by business companies; the share of large public research centres such as CNRS itself, INSERM and INRA is about 30%; while the universities' share is less than 10% (the few remaining patents are owned by individuals),. Further, CNRS controls a higher percentage of the patents signed by its employees than universities: in fact, more than 40% of patent applications with CNRS researchers as inventors belong to public research organizations (almost all of which to CNRS itself), while universities own only 11% of academic patents. Patents by academic inventors are also more likely to be held by companies (63%) than their CNRS equivalents (54%).

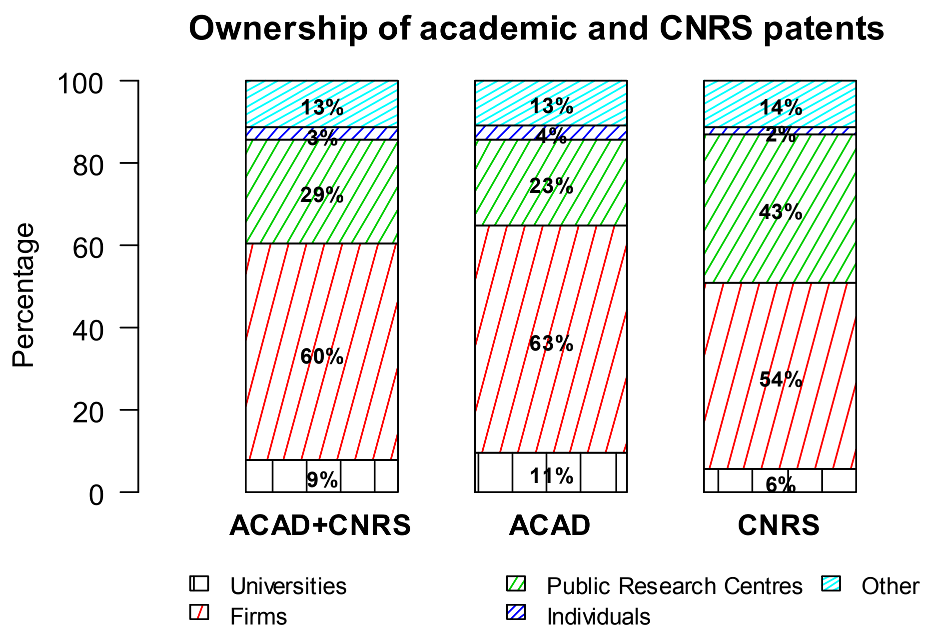


Figure 2 Ownership of academic and CNRS patents in France ("ACAD+CNRS"=academic + CNRS inventors, "ACAD" - academic inventors, "CNRS" - CNRS inventors). [4]

Lissoni et al. (2008) compared ownership pattern of academic patents in Europe versus the US, where university ownership rates are significantly higher (close to 70% according to Thursby et al (2006)), and concluded that the differences are largely explained by legal factors (related to IPR legislation), as well as by the institutional features of the university and innovation systems in the various countries. Here we stress that that most academic patents owned by business companies are

[4] "ALL" - all inventors, "ACAD+CNRS"=academic and CNRS inventors, "ACAD" - academic inventors, "CNRS" - CNRS inventors. Co-owned patents were counted as many times as there were different types of owners.

very much likely to be the result of research efforts conducted jointly by academic and industrial scientists. These patents result therefore from contacts between the academic and industrial communities. We turn now to explore the social structure resulting from such contacts, in terms of structural properties of the networks of inventors, and the position of academic scientists therein.

4. NETWORKS OF INVENTORS AS SMALL WORLDS

Patent and inventor data lend themselves to be transformed into “affiliation networks” (also known as "two-mode" networks). The latter are bipartite graphs with two types of nodes: actors and events, where actors are linked to events in which they participate. In our case actors are the inventors and the events correspond to patent applications.[5]

Affiliation networks, in turn, can be used for deriving corresponding "one-mode" networks in which actors are directly connected, based on the assumption that all actors affiliated to the same event are also in direct contact. Following the methodology outlined in Balconi et al. (2004) we projected the bipartite network of patents and inventors onto the set of inventors, under the assumption that all inventors of the same patent have direct links with each other. Such an assumption seems to be reasonable, if one takes into account that teams of inventors are on average rather small (less than 1% of all patents have more than 5 inventors, the average team size is in between 1.5 and 2.5 co-inventors, depending on the technological field.

In what follows, we first report some basic statistics on the French networks of inventors, by technology (section 4.1). We then discuss the small world properties of such networks for the technologies with the highest presence of academic and CNRS inventors (section 4.2). The networks were constructed using data on all patents from 1978 to 2004.

4.1 BASIC EVIDENCE

Networks of inventors appear to be highly fragmented, as they are composed of a large number of distinct components. However, as shown in Table 3, in the five technologies with a higher presence of academic and CNRS inventors we see that the largest component is much larger than all others, including the second largest one. This suggests some resemblance with networks of scientific collaborations, as measured by co-authorship of scientific papers, which in turn present some

[5] Other typical instances of affiliation networks are found in corporate interlocks (Koenig and Gogel 1981), research collaborations (Powell et al. 1996), scientific co-authorship (Newman 2001), and underwriting syndicates (Baum et al. 2003).

characteristics typical of "small-world" networks (Newman, 2001)[6]. We will come back to this issue in section 5.

Table 3 Networks of inventors: size of the main components, and distribution of academic inventors therein

TECHNOLOGICAL FIELDS		*Nr. of inventors*						C1/ALL		C2/C1
		ALL (1)		C1 (2)		C2 (3)				
		(4)	(5)	(4)	(5)	(4)	(5)	(4)	(5)	
Electrical engineering. Electronics	**All. inv**	**23183**	**13610**	**6459**	**3978**	**928**	**567**	**0.28**	**0.29**	**0.14**
	Ac.Inv	*226*	*217*	*99*	*94*	*1*	*1*	*0.44*	*0.43*	
	CNRS	*166*	*123*	*71*	*49*	*0*	*0*	*0.43*	*0.40*	
Instruments	**All. inv**	**18419**	**9714**	**4542**	**2870**	**128**	**70**	**0.25**	**0.30**	**0.03**
	Ac.Inv	*370*	*363*	*149*	*147*	*0*	*0*	*0.40*	*0.40*	
	CNRS	*256*	*178*	*108*	*77*	*0*	*0*	*0.42*	*0.43*	
Chemicals. Materials	**All. inv**	**15908**	**8653**	**9611**	**5723**	**85**	**48**	**0.60**	**0.66**	**0.01**
	Ac.Inv	*345*	*336*	*276*	*268*	*0*	*0*	*0.80*	*0.80*	
	CNRS	*369*	*259*	*298*	*208*	*0*	*0*	*0.81*	*0.80*	
Pharmaceuticals. Biotechnology	**All. Inv**	**9134**	**5980**	**5213**	**3608**	**28**	**19**	**0.57**	**0.60**	**0.01**
	Ac.Inv	*398*	*396*	*232*	*232*	*0*	*0*	*0.58*	*0.59*	
	CNRS	*362*	*280*	*242*	*183*	*0*	*0*	*0.67*	*0.65*	
Industrial processes	**All. Inv**	**15677**	**8159**	**3203**	**2049**	**54**	**18**	**0.20**	**0.25**	**0.02**
	Ac.Inv	*155*	*153*	*85*	*84*	*0*	*0*	*0.55*	*0.55*	
	CNRS	*132*	*97*	*86*	*68*	*0*	*0*	*0.65*	*0.70*	
Mech. Eng. Machines. Transport	**All. Inv**	**19869**	**10386**	**1005**	**647**	**881**	**611**	**0.05**	**0.06**	**0.88**
	Ac.Inv	*64*	*64*	*2*	*2*	*2*	*2*	*0.03*	*0.03*	
	CNRS	*28*	*22*	*4*	*4*	*2*	*2*	*0.14*	*0.18*	
Consumer goods. Civil eng.	**All. Inv**	**10310**	**5158**	**201**	**150**	**171**	**137**	**0.02**	**0.03**	**0.85**
	Ac.Inv	*13*	*12*	*1*	*1*	*1*	*1*	*0.08*	*0.08*	
	CNRS	*9*	*5*	*0*	*0*	*0*	*0*			

(1) Overall network of inventors.
(2) Largest connected component of the inventors' network.
(3) Second largest component of the inventors' network.
(4) All inventors (including those who signed no patents after 1993).
(5) Only inventors with at least one patent in 1994-2004.

The largest components of networks in *Electrical engineering & Electronics* and *Instruments* collect 25% and 28% of all inventors, respectively, with the second largest component being just one or two tenth of it. The results are even stronger in *Chemicals & Materials* and *Pharmaceuticals & Biotechnology*, where the share of the inventors in the largest component is well over half of the total size of the network, and the second largest is follows at a ratio of 1:100. As for *Industrial*

[6] Comparing our result with Newman (2001) we shall keep in mind that our networks are less connected for a number of reasons: taken two comparable population of inventors and scientific authors active in related technologies and scientific disciplines, the number of patents per inventor is usually lower than the number of papers per author; and the average number of co-inventors in a patent is lower than the average number of co-authors per scientific paper. In addition, our networks spread through technological fields which are much larger and more heterogeneous than Newman's scientific disciplines.

Processes, we observe an intermediate situation, with the weight of the principal component in line with what found for *Electrical engineering & Electronics* and *Instruments* (20%), but a first/second component size ratio close to that of *Chemicals & Materials* and *Pharmaceuticals & Biotechnology* (2:100).

Academic and CNRS inventors are more likely to be found in the main components of the most science-based technological fields such as *Chemicals & Materials* and *Pharmaceuticals & Biotechnology*, with *Instruments, Industrial Processes,* and *Electrical Engineering & Electronics* following in decreasing order. Notice that the technological fields with the lowest presence of academic and CNRS scientists, namely *Mechanical Engineering* and *Consumer Goods*, exhibit a small size of the main component. .

Table 4 provides information on the main characteristics of the largest components in our networks. The first column in the table reports the number of inventors in the largest component (*N*). The second and third columns report two distinct graph centralization measures, namely "betweenness", "degree" and "closeness" centralization index (B_{CENT}, D_{CENT}, and C_{CENT}, respectively)[7].

Table 4 Networks of inventors: Properties of the largest components*

TECHNOLOGICAL FIELD	N	S	B_{CENT}	D_{CENT}	C_{CENT}	C	D	L
Electrical engineering. Electronics	6459	0.07%	0.194	0.009	0.101	0.345	35	12.4
Instruments	4542	0.11%	0.133	0.016	0.091	0.546	39	12.3
Chemicals. Materials	9611	0.06%	0.118	0.018	0.112	0.319	31	8.7
Pharmaceuticals. Biotechnology	5213	0.11%	0.115	0.014	0.120	0.390	28	8.8
Industrial processes	3203	0.15%	0.166	0.016	0.108	0.350	35	9.8
Mechanical eng. Machines. Transport	1005	0.50%	0.482	0.039	0.097	0.441	34	10.4
Consumer goods. Civil engineering	201	1.95%	0.390	0.076	0.181	0.306	11	5.3

N = Nr of inventors in the largest component
S = network density (number of links/max. number of links)
B_{CENT}= Betweenness centralization index of the largest component (Avg betweenness/Avg betweenness of N-node star graph)
C_{CENT}= Closeness centralization index of the largest component (Avg closeness centrality/Avg closeness centrality of N-node star graph)
D_{CENT}= Degree centralization index of the largest component (Avg degree centrality/Avg degree centrality of N-node star graph)
C = Clustering coefficient of the largest component (definition in main text)
L = Avg path length of the largest component
D = Diameter of the main component
* All inventors (including those who signed no patents after 1993).

[7] All network measures in the paper have been calculated using *igraph* package of R (Csardi and Nepusz 2006).

In social network analysis these measures are used to characterize the observed variation in actors' centralities across the network (Wasserman and Faust, 1994). They take values from 0 to 1 and indicate how far a network is from being as centralized as a star graph[8].

In general, large social networks based upon archival data, such as ours, exhibit a degree centralization index very close to zero: no single node is connected to all others, and most nodes exhibit more than one tie, which suggest a structure very far from that of a star graph. The main component of our network is no exception: there are no inventors who worked with a very large share of the total population of inventors, but many inventors have worked with more than one partner, either because they have joined different teams throughout their career, or, more commonly, because the only team they joined comprised more than two inventors. As a result, the degree centralization $\mathbf{D_{CENT}}$ is very low (close to zero).

However, we find that the betweenness and closeness centralization indices of our networks ($\mathbf{B_{CENT}}$ and $\mathbf{C_{CENT}}$) are rather high, which suggests the existence of a few inventors who occupy key positions in between other inventors or groups of inventors, without whom average social distances would be higher. As we will see in section 5, a disproportionate number of these in-between inventors are indeed academics and CNRS researchers.

The fourth, sixth, and eight columns of Table 4 report the clustering coefficient (*C*)[9], average length (*L*) and diameter (*D*) of our networks.

The clustering coefficient can be interpreted as the probability of two randomly chosen inventors to be co-inventors conditional on that they have (at least) one co-inventor in common. The two measures in Table 4, average distance *L*, and diameter *D*, are both related to the notion of the shortest path ("geodesic"): *L* is the length of the shortest path between two nodes on the graph, averaged over all nodes of the graph; *D* is the maximum of all geodesics of the graph (Wasserman and Faust, 1994). The values of such indicators suggest that networks of inventors, especially in

[8] A star graph of *n* nodes is made of one central node that is connected to all the other (*n-1*) nodes, none of which is connected to each other. That is, reaching a non-central node from another non-central node requires following a path through the central one. The degree centrality of node *j* is simply the number of ties pointing to the node. The betweenness centrality of node *j* is calculated as the number of shortest paths linking any two other nodes, which cross node *j* (node *j* stands "in between" the two other nodes). The degree (betweenness or closeness) centralization index of the network is calculated as the ratio between the average degree (betweenness or closeness) centrality of its nodes and the average degree (betweenness or closeness) centrality for a star graph network of the same size.

[9] We define the clustering coefficient of a network as in Newman et al. (2001):

$$C = \frac{3 \times number\ of\ triangles\ on\ the\ graph}{number\ of\ connected\ triples\ of\ vertices}$$

By "triangle" it is meant here a set of three nodes (a "triple"), each of which is connected to both the others; while a "connected triple" is a triple in which at least one node is connected to both the others. The numerator is multiplied by 3 because each triangle contributes to 3 connected triples of vertices, one for each of its 3 vertices; this adjustment ensures that the value of ***C*** will be always comprised between 0 and 1. This definition of clustering coefficient is sometimes referred as "transitivity" (or "transitivity ratio").

fields with strong presence of academics and CNRS researchers, may have some properties typical of small worlds (Watts and Strogatz, 1998).

First, very much like most large scale networks they are quite sparse: their network density (defined as the ratio of observed links to the number of all possible links between the network's nodes) is very low, lying as it does below 0.05% (the most dense network in *Pharmaceuticals & Biotechnology* has a density of only 0.045%). Even largest components which are generally denser than overall networks have density limited by 0.15%.

Second, their clustering coefficients appear to be very high. In particular, the values reported for *Chemicals & Materials* and for *Pharmaceuticals & Biotechnology*, the clustering coefficients for the inventors' networks are more than 1000 times larger than those we would find in a the theoretical random graph with the same number of nodes and ties.

Third, the average path length is rather short, as it never exceeds 13 steps and it is around 8 steps for the more science-based technologies.

4.2 Small world properties of inventor networks

The evidence we produced so far suggest that networks of inventors exhibit some small world properties, also typical of networks of scientists (Newman, 2001). To investigate further in this direction, we ought to produce a more proper test. In fact, we produce two: a small-world ratio measure adapted to networks of inventors; and a "rewiring-based" test .

4.2.1 A inventor-network-specific small world ratio

As suggested by Watts and Strogatz (1998), this should consist in comparing the observed network's clustering coefficient and average path length ("distance") to the values they would take in a benchmark random network (BRN), that is a graph with the same number of nodes and ties, where ties are distributed randomly. A synthetic way to conduct this comparison is proposed by Davis et al. (2003), who suggest calculating the following "small world ratio" Q:

$$Q = (C_{obs}/C_{BRN})/(L_{obs}/L_{BRN})$$

where C_{obs} and L_{obs} are respectively the clustering coefficient and the average path length in the observed network, and C_{BRN} and L_{BRN} the values of the same indicators in the benchmark random network. High values of Q suggest that the observed network has a higher clustering coefficient, but a similar average path length of a comparable random graph; as such they are indicative of small world qualities in the observed network.

The BRN used to calculate Q has to be generated by a stochastic process compatible with the nature of the data at hand. Accordingly, we follow Molloy and Reed (1995) in building a random two-mode inventor-patent network, which preserves the same patents-per-inventor and inventors-per-patent of the original network; and then project it onto the set of inventors, thus obtaining a one-mode network to be compared with the observed one. [10] [11]. We also limit our attention to technological fields which report a significant presence of academics and CNRS researchers (*Instruments*, *Chemicals & Materials*, *Pharmaceuticals & Biotechnology*, and *Industrial processes*).

Table 5 reports the structural characteristics of the simulated BRN averaged over 100 simulation runs (second line, in italics).

First, we notice a wide difference between the sizes of the first two largest components (columns **C1** and **C2**) in the observed network and the BRN. Random matching of inventors and research teams does not respect the existing boundaries between organizations, localities and technological niches typical of the real world, thus introducing a much higher level of connectedness hardly than observed networks can possibly achieve.

Further, connectivity of networks is achieved through most productive inventors, who participate in many patents. This effect, which is also present in observed networks, is disproportionally inflated in the BRN. While in the real world prolific inventors tend to move little in space and across firms or technologies, in the BRN the probability that the same inventors will engage in repeated collaborations is negligibly small. This explains why degree centralization of the main component ($\mathbf{D_{CENT}}$), a measure of variation in the number of an individual's collaborators, is somewhat higher in BRNs than in the observed networks. Larger main component and shorter social distances also results in BRNs exhibiting a higher centralization of the main component ($\mathbf{C_{CENT}}$).

---

[10] Our choice contrasts with the more common choice to found in the literature, where the "corresponding random graph" is usually the Erdos-Renyi (ER) one. In the ER model the random process is such that a tie between any pair of nodes is generated with equal probability, independently of the existence of other ties. The probability of a tie's existence is simply equal the network density. However, a difficulty arises when we wish to produce ER random graphs for networks of inventors, due to the latter's origin in a inventor-patent affiliation network (Uzzi et al., 2007). As in all affiliation network, connections are carried by events (in our case, the patents), so they come in a bunch. As a consequence, we cannot 'rewire' network ties one by one: inventors, in fact, establish contacts with whole teams of other co-inventors (as listed on a patent), hence rearranging each time a whole set of contacts with all co-inventors in the corresponding patent teams. This will necessarily result in high clustering coefficients, which in turn drive up artificially the small world ratio Q. It follows that the standard ER random graph, often used as a benchmark in studying small-world network structures, is not an appropriate model for networks of inventors (and, in general, for any other network constructed as a projection of a bipartite network).

[11] In network literature Molloy-Reed algorithm in which random graphs are generated for *a given sequence of nodal degrees* is referred as "configuration model". Newman et al. (2001) study somewhat different family of random graphs, random networks from a *given distribution of degree sequences*. Both models converge to the same result for networks as large as the ones discussed in this paper. See also Robins and Alexander (2004) and Kogut and Belinky (2008).

Table 5 reinforces our earlier statement about the presence of central inventors: observed networks have significantly higher betweenness centralization (**B_{CENT}**) than BRNs, which implies that some inventors are significantly more important for network connectivity than others. Typically such inventors connect communities of inventors otherwise disconnected or separated by long social distances (as in Breschi and Lissoni, 2009).

Table 5 Observed inventors' network vs. simulated random graph*

TECHNOLOGICAL FIELDS		C1	C2	B_{cent}	D_{cent}	C_{cent}	C	D	L	Q
Electrical engineering. Electronics	observed	6459	928	0.194	0.009	0.101	0.345	35	12.4	0.6
	simulated	*16922*	*7.9*	*0.068*	*0.011*	*0.193*	*0.262*	*14.7*	*5.5*	
Instruments	observed	4542	128	0.133	0.016	0.091	0.546	39	12.3	1.1
	simulated	*12955*	*8.3*	*0.089*	*0.015*	*0.206*	*0.216*	*14.7*	*5.4*	
Chemicals. Materials	observed	9611	85	0.118	0.018	0.112	0.319	31	8.7	1.6
	simulated	*13784*	*4.7*	*0.038*	*0.018*	*0.225*	*0.096*	*10.5*	*4.2*	
Pharmaceuticals. Biotechnology	observed	5213	28	0.115	0.014	0.120	0.390	28	8.8	1.8
	simulated	*7789*	*4.8*	*0.063*	*0.034*	*0.257*	*0.101*	*10.1*	*4.0*	
Industrial processes	observed	3203	54	0.166	0.016	0.108	0.350	35	9.8	1.4
	simulated	*10232*	*6.9*	*0.075*	*0.016*	*0.210*	*0.124*	*13.3*	*5.0*	

C1 = Nr of inventors in the largest component
C2 = Nr of inventors in the second largest component
B_{CENT}= Betweenness centralization index of the largest component (Avg betweenness/Avg betweenness of N-node star graph)
C_{CENT}= Closeness centralization index of the largest component (Avg closeness centrality/Avg closeness centrality of N-node star graph)
D_{CENT}= Degree centralization index of the largest component (Avg degree centrality/Avg degree centrality of N-node star graph)
C = Clustering coefficient of the largest component (definition in main text)
L = Avg path length of the largest component
D = Diameter of the main component
Q = Small worlds ratio (definition in main text)
* All inventors (including those who signed no patents after 1993).

Turning to the last three columns of Table 5 we first notice that the small world ratio (column **Q**) ranges from 0.6 in *Electrical engineering & Electronics* to 1.8 in *Pharmaceuticals & Biotechnology*. This, perhaps, allow us to say that the network of inventors in *Electrical engineering & Electronics* is not a small word at all (because for small worlds Q must be significantly larger than 1), but it does not provide any conclusive evidence on whether networks of inventors in other technologies have small-world structures.

This is because we do not know, a priori, how big Q should be "to qualify" a network as a small world one, nor we can be sure that Q , as calculated above, works well networks, such as ours, which are projections of affiliation networks.

4.2.2 A "REWIRING-BASED" TEST

In order to solve these doubts, we introduce an extension of the Watts-Strogatz model for one-mode projections of bipartite networks. In the same way we constructed the BRN above we, now introduce randomness into the observed networks. Again we start with the original bipartite network of inventors-patents and with some probability we randomly 'rewire' existing connection between inventor and patent. As in the Watts-Strogatz model, the higher the rewiring probability the closer we are to the corresponding random graph (in which both the clustering coefficient and the average distance are very low).

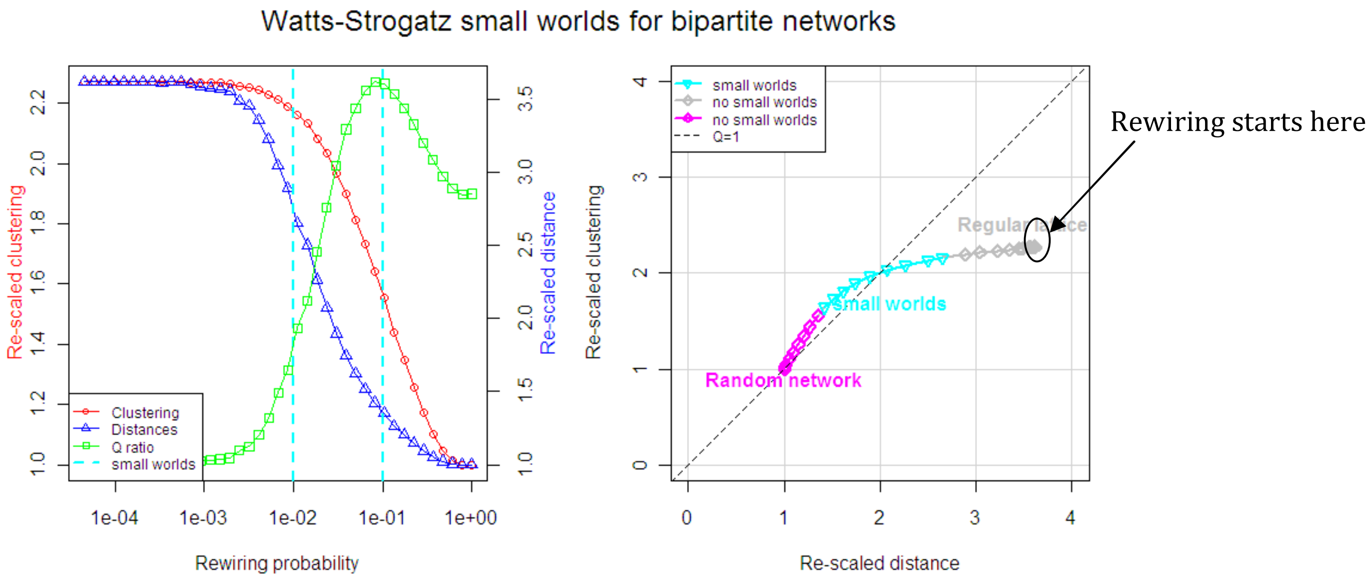


Figure 3 Small worlds networks (model).

The left panel of Figure 3 shows a typical relationship between rewiring probability, clustering, and the average path length in the main component for an idealized network with 100 inventors collaborating on 100 patents, where each patent is produced by a team of 5 inventors and two consequent patents overlap by 4 inventors, so that corresponding one-mode projection is a ring structure as in Watts and Strogatz (1998). The diagram to the left closely replicates the figure in the original work of Watts-Strogatz (1998), which shows that small worlds emerge where corresponding curves for distances and clustering diverge: as the re-wiring probability increases, the scaled average distance L soon drops to the level of the random network while the clustering coefficient follows with some delay. For the modelled network rewiring results in networks with small world properties when the rewiring probability is in the range between 0.01 and 0.1 (marked by two dashed vertical lines at the diagram).

The right panel of Figure 3 presents the model in distance-clustering coordinates (scaled by average distance and average clustering coefficient of BRN). We start from the regular lattice (ring) characterized by high clustering and large average distances (North-East corner of the diagram). As the probability of rewiring approaches 1, we move steadily towards the South-West corner and we come close to the BRN, which is a point with coordinates (1,1) on distance-clustering plain. Small world structures lie between these two extremes.

This simple exercise suggests that the logic of the Watts-Strogatz model (1998) can be directly extended to our case, when 'randomization' is achieved through re-wiring of the original bipartite network as explained above, and as in Watts-Strogatz model it can be said that the set of networks produced by the re-wiring algorithm portrays a spectrum of network structures from ordered regular graphs to completely random BRNs giving the probability of re-wiring meaning of 'degree of randomness'.

Thus a test for small-world structure in the networks of inventors can be conducted as follows: we hypothesize that inventor networks, in order to exhibit small world properties, should also exhibit higher clustering and shorter distances, than their corresponding BRNs. As a consequence, when rewiring of a small-world inventor network we should observe only minor negative effect on the distances, but proportionally higher (negative) effect on clustering. By contrast, if the inventor network does not exhibit any clear small world feature, and it is composed of many isolated cliques, rewiring should have a strong effect on distances and a small effect on clustering.

The results of our experiments are shown in Figure 4. Networks of inventors in *Chemicals & Materials* and *Pharmaceuticals & Biotechnology* are somewhat closer to BRN, while networks in *Scientific & Control Instruments* and *Industrial Processes* are closer to regular networks. Yet all these networks show the same pattern of Figure 3(they move from North-East to South-West in the graph as rewiring proceeds), which suggests that they have small world structure. By contrast, the inventor network in *Electrical engineering & Electronics* exhibit a somewhat different pattern, in that rewiring at first increases both distances and clustering in the largest component.

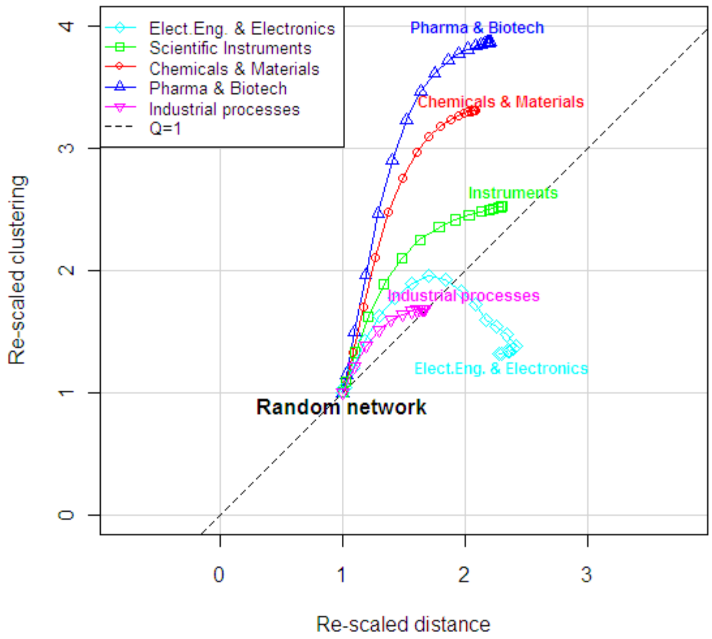


Figure 4 Small worlds in inventors' networks (networks of all inventors, including those who signed no patent after 1993).

Increasing distances derive from the growing size of the largest component (since rewiring increases the size of the largest component in all four networks, in the other networks the effect of introduced short-cuts offsets the effect of growing main component so that overall effect of rewiring on distances is negative. The largest component grow by connecting otherwise disconnected large tightly linked clusters which increases the clustering in the largest component. By contrast, in the other four networks all large clusters are part of the corresponding largest components, and main effect of rewiring in such networks is in breaking dense clusters within the largest component decreases clustering within the largest component.

5 Position and role of academic and CNRS inventors

So far, we have observed that four out of five of our inventor networks with high presence of academic and CNRS inventors have small world features. That is, they are composed of many cluster of tightly connected inventors, held together by a few inventors whose connections reach out the cluster, and contribute to keep average path lengths low.

We now investigate whether it is precisely academic and CNRS inventors who provide the out-of-cluster connections that typically transform a regular graph into a small world.

We first we examine some measures of academic inventors' position (centrality) in the largest component of each technological field, with respect to all the inventors in the component (section 5.1). Then, we identify cohesive subgroups and point at the importance of academic and CNRS inventors in providing a connection between them (section 5.2).

As mentioned in section 3, the data on academic and CNRS inventors becomes less reliable as we go back in time. To correct for such a bias all comparisons between academic, CNRS, and industrial researchers and their patents reported here are done on the subset of inventors having at least one patent signed between 1994 and 2004.

5.1 CENTRALITY

As we mentioned above, high centralization scores indicate the existence of asymmetries in positions within the networks, with a few inventors occupying central positions and many others at the periphery. We expect central inventors to play an important role in transferring knowledge and related resources, both symbolic and material (for example, they may be senior scientists in large R&D labs and decide over the allocation of funds to project; or they may have a say on the distribution of scientific credit, as discussed in Lissoni and Montobbio, 2010).

Table 6 Position of academic inventors in the main component*

TECHNOLOGICAL FIELDS		N	B_{CENT}	C_{CENT}	D_{CENT}
	All inv.	**3978**	**0.0024**	**0.0837**	**4.9**
Electrical engineering. Electronics	*Uni inv*	*94*	*0.0027*	*0.0811*	*5.5*
	CNRS inv	*49*	*0.0037*	*0.0856*	*5.5*
	All inv.	**2870**	**0.0034**	**0.0841**	**5.7**
Instruments	*Uni inv*	*147*	*0.0069*	*0.0840*	*6.5*
	CNRS inv	*77*	*0.0039*	*0.0844*	*5.4*
	All inv.	**5723**	**0.0011**	**0.1210**	**7.1**
Chemicals. Materials	*Uni inv*	*268*	*0.0019*	*0.1256*	*8.2*
	CNRS inv	*208*	*0.0019*	*0.1257*	*7.9*
	All inv.	**3608**	**0.0018**	**0.1186**	**6.4**
Pharmaceuticals. Biotechnology	*Uni inv*	*232*	*0.0034*	*0.1216*	*7.0*
	CNRS inv	*183*	*0.0026*	*0.1246*	*7.7*
	All inv.	**2049**	**0.0035**	**0.1098**	**5.6**
Industrial processes	*Uni inv*	*84*	*0.0081*	*0.1146*	*6.8*
	CNRS inv	*68*	*0.0038*	*0.1177*	*6.0*

B_{CENT}= Avg betweenness centrality of inventors considered
C_{CENT}= Avg closeness centrality of inventors considered
D_{CENT}= Avg degree centrality of inventors considered
* Inventors who signed at least one patent between 1994 and 2004.

Table 6 reports the average betweenness, closeness and degree centrality scores (respectively B_{CENT}, C_{CENT} and D_{CENT}) for academic and CNRS inventors as opposed to all inventors, within the main component of each technological field[12]. Academic and CNRS inventors tend to occupy more central positions than the average inventor, both locally (D_{CENT}) and globally (B_{CENT} and C_{CENT}).

Whatever the technological field considered, D_{CENT} for both academic and CNRS inventors is higher than average which implies that academic and CNRS inventors have a higher-than-average number of co-inventors.. This, in turn, may be a result of either (*a*) higher number of patents signed by the inventor, (*b*) larger invention teams (number of co-inventors per patent) in which the individual participates, or (*c*) higher mobility across different invention teams.

As for B_{CENT} and C_{CENT} they are also higher for academic and CNRS inventors than for the all inventors, with the only exception of *Electrical engineering & Electronics* and *Instruments* for C_{CENT}.

Table 7 reports some descriptive statistics on the size of inventors' teams, number of inventors designated on the same patent, and on inventors' productivity, defined as number of patents signed by each inventor (in commenting it, we will refer to teams that include at least one identified academic inventor as "academic teams").

Table 7 Size of inventors' teams and productivity of inventors, by technological field*

	Team size			
TECHNOLOGICAL FIELD	*All inventors*	*Academic Invs (1)*	*CNRS inventors (2)*	*(1) + (2)*
Electrical engineering. Electronics	2.01	3.01	3.43	3.06
Instruments	2.10	3.45	3.80	3.43
Chemicals. Materials	2.68	3.98	3.84	3.87
Pharmaceuticals. Biotechnology	2.51	3.63	3.98	3.68
Industrial processes	1.87	3.72	3.81	3.65
ALL TECHNOLOGIES	2.08	3.59	3.82	3.59

	Productivity			
TECHNOLOGICAL FIELD	*All inventors*	*Academic Invs (1)*	*CNRS inventors (2)*	*(1) + (2)*
Electrical engineering. Electronics	3.12	3.25	3.18	3.22
Instruments	3.59	3.47	3.50	3.48
Chemicals. Materials	4.72	4.32	4.36	4.34
Pharmaceuticals. Biotechnology	4.12	3.06	3.14	3.09
Industrial processes	4.00	4.60	5.07	4.78
ALL TECHNOLOGIES	2.79	2.84	2.95	2.88

* Only inventors who signed at least one patent between 1994 and 2004 and their patents..

[12] For a definition of D_{CENT} see footnote 4.

With respect to overall productivity (“all technologies” in Table 7) we find that although academics seem to be more productive than average, more careful comparison of inventors’ productivity by field reveals that this is largely a composition effect. Academic researchers tend to be concentrated in technological fields such as *Pharmaceuticals & Biotechnology* and *Chemicals & Materials* (Figure 1), where the number of patents per inventor is higher than in other fields. However within these two fields their productivity is lower than average. Comparing academic and CNRS inventors we find that the latter are more productive than the former in most technologies (with the exception of *Electrical engineering & Electronics*).

As for size of research teams, we first notice that the mean size of inventors’ teams in science-intensive technologies (where academic patenting is more relevant), is larger than average, because inventions are more likely to result from a collective effort rather than from a solitary one. Second, in all technological fields the size of academic teams tends to be larger than the average team size for the field. This may suggest that, even within science-intensive technologies, patents produced by academic teams are more science intensive than others, and this intensity shows up in the team size. Before accepting this interpretation, however, one has to discard a more obvious one, which is purely statistical and arises from our definition of “academic team”.

In order to understand this point one has to realize that even if the size of a team were independent on the presence of an academic member, the probability to have at least one academic inventor in a team increases with the size of the team. This means that even if teams were assembled at random, the expected size of academic teams would greater than average. The magnitude of the bias is larger, the larger the share of academic inventors in the technological field.

In order to correct for this statistical effect we produce a “baseline” distribution of academic teams according to the procedure outlined in Appendix A. The average of sizes of “baseline” academic teams for each of technological fields is shown at Table 8 (in italics, below the observed team sizes).

Comparing these averages with the average sizes of academic teams we find that academic researchers are more likely to be a part of a larger inventors’ team in all technological fields. In *Chemicals & Materials* and *Pharmaceuticals & Biotechnology*, most science-intensive sectors, however the difference between observed and “baseline” averages is less pronounced than in other technologies, in these two technological fields including an academic researcher into the inventors’ team does not seem to make any significant difference in terms of team size; possibly, this is because such inclusion would not make any difference in terms of the team structure of the inventive effort.

Table 8 Adjusted size of inventors' teams, academic, university, CNRS; by discipline*

TECHNOLOGICAL FIELD		*All inventors*	*Acad. Invs (1)*	*CNRS invs (2)*	*(1) + (2)*
Electrical engineering. Electronics	Observed	2.01	3.01	3.43	3.06
	Baseline		*2.72*	*2.75*	*2.70*
Instruments	Observed	2.10	3.45	3.80	3.43
	Baseline		*2.90*	*2.96*	*2.87*
Chemicals. Materials	Observed	2.68	3.98	3.84	3.87
	Baseline		*3.44*	*3.46*	*3.38*
Pharmaceuticals. Biotechnology	Observed	2.51	3.63	3.98	3.68
	Baseline		*3.28*	*3.33*	*3.20*
Industrial processes	Observed	1.87	3.72	3.81	3.65
	Baseline		*2.53*	*2.54*	*2.52*
ALL TECHNOLOGIES	Observed	2.08	3.59	3.82	3.59
	Baseline		*2.85*	*2.87*	*2.82*

* Patents of inventors with at least one patent between 1994 and 2004.

Thus in these science-intensive sectors higher degree centrality of academic and CNRS inventors is related neither to the number of patents per inventor, nor to the size of the research teams. It points to the third explanation listed above – on average academic and CNRS inventors move across teams more often, while researchers in industry tend to work within same teams. This in turn may be due to:

- The size of research teams which involve academic and CRNS personnel. Such teams are likely to be teams attached to scientific, rather than purely technological projects; and to involve universities and the CNRS at the institutional level, thus guaranteeing their participants more freedom to pick up collaborators and partners than projects entirely controlled by business companies. As a consequence, we may expect a higher number of co-inventors per patent
- The tendency of academic inventors to work for a diverse set of inventors' teams. Most inventors are R&D employees of business companies; as such, they cannot easily change team, unless their employer is very large (i.e. hosts many different teams) or they change employer (so that, a fortiori, they change team). Academic and CNRS inventors, on the contrary, behave more like "free lance" inventors, who look for different sources of funding and partnerships, and change inventors' teams accordingly.

5.2 MOBILITY ACROSS ORGANIZATIONAL BOUNDARIES

Social network theory offers several approaches to identify tightly-knit subgroups in a network, such as traditional blockmodeling techniques (based on the notion of network position and structural equivalence: Wasserman and Faust, 1994), stochastic blockmodeling methods (based on estimation of particular stochastic process which generate observed network: Frank, 1995; Snijders and Nowicki, 1997) or structural cohesion methods (based on connectivity: Moody and White 2003). For our purposes, however, we have opted for a "community detection" approach that has become popular in interdisciplinary "network research" in the last decade. Community detection algorithms try to find division of a network into subgroups (partition) with the highest modularity, a measure which quantifies goodness of partition comparing observed density of connections within and between subgroups vs. expected densities in BRN.[13] Further, out of many available algorithms to optimize modularity we have chosen a recent "Louvein" algorithm (Blondel et al. 2008), which is best suited for one-mode projections of bipartite networks, as it our case[14].

A bipartite network of inventors and patents can be projected both onto the set of inventors (as we have done so far in the paper) as well as onto the set of patents. In most contexts the latter make little sense, however it can be useful for the purposes of this section of the paper.. First, while a standard result of blockmodeling consists, from a theoretical perspective, in a partitioning of a network into a set of separate clusters such that each given individual belong to one group only[15], it might be advantageous to have a set of overlapping communities of inventors, so that we can see which inventors are located at the intersects of several communities. This can be easily achieved by identifying clusters of patents first and then associating inventors' communities to the patent clusters: an inventor who has patents in several clusters will immediately appear to be at the

---

[13] Formal definition of modularity score for a partition of a network into communities $c_1, \ldots, c_M$ is

$$Q = \frac{1}{2m}\sum_{i,j}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(c_i, c_j)$$

where $\boldsymbol{A}$ is adjacency matrix, m is total number of edges, k_i is degree of node i, and $\delta(c_i, c_j)$ is the Kronecker delta (equal to 1 only if both i and j belong to the same community. With exception of small and trivial networks there is no exact analytical solution to the optimization problem, a number of approximate algorithms have been proposed and examined including original method of Newman and Girvan (2004). For a broad review of community detection methods see Fortunato (2010).

[14] Two reasons stay behind our choice: first, straightforward application of community detection algorithms might not work, due to the "two-mode origin" of our network; second, the methods developed specifically for two-mode networks (e.g. Barber 2007) are computationally intensive, which might be an obstacle given the size of the networks we want to analyze. However, in their study Guimera et al. (2007) found that modularity optimization on a *weighted* projection of a bipartite network generally performs well. The "Louvein" algorithm by Blondel et al. (2008) has certain advantages in this respect: first, it can work with weighted networks and second it is fast even on large networks (its authors have successfully applied to a mobile phone network of 2.6 million customers). In addition, the algorithm may help uncovering the finer structure of the clusters (i.e. clusters within clusters).

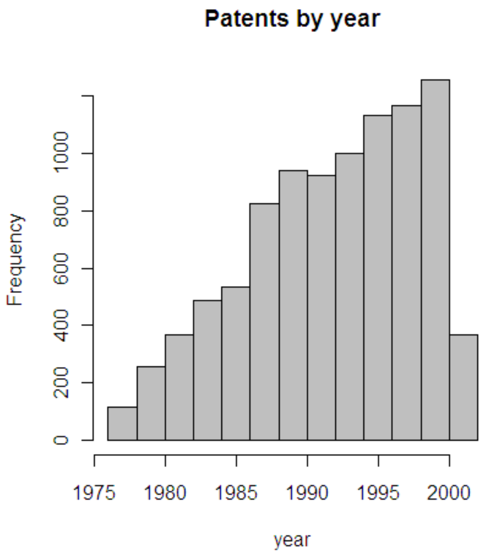
[15] Several methods for identifying overlapping communities have been developed recently (Palla 2005, Moody and White 2003).

intersection of the communities of inventors corresponding to the clusters. Further, from the practical perspective of identifying patent clusters, might make it easier to rationalize the results of the blockmodeling, because more information is available on patents (time, applicant, IPC class) than on their inventors.

We proceed as follows. First we construct a network of patents as a weighted projection of bipartite network of inventors and patents. We then analyse it by using a community detection method, which allows us to identify tightly linked patent clusters. The latter induce a partitioning of inventors consisting of overlapping communities. Then using characteristics of the patents (priority date and applicant) we can rationalize the nature of the communities and examine inventors who are located at the intersection of several communities.

We first illustrate the implementation of these steps on the largest component of inventors in *Industrial processes* technological field, and then move on to look at the results for all fields.

The largest component of the inventors' network in *Industrial processes* connects 3203 inventors, 2049 or about 2/3 of these inventors have at least one patent after 1993, among them there 84 academic inventors and 68 CNRS researchers (Table 3). The distribution of patents by priority date and applicant (top 15 applicants) is shown at Figure 5 (fractional counts for patents assigned to more than one applicant, patents owned by individuals rather than organizations ignored).



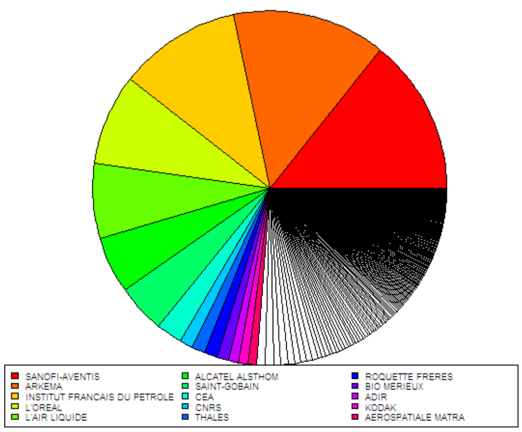


Figure 5. Distribution of patents in *Industrial Processes* by priority date and applicants.

We project the bipartite network of inventors and patents onto the set of 9494 patents, so that two patents are linked if the corresponding research teams share at least one inventor; the strength (weight) of a link between the two patents is equal to the number of inventors they share (self-links excluded).

We then apply the "Louvein" aggregation method. First, each of the 9494 patents is considered as a cluster of its own. Then they are placed one by one into a cluster which provides highest positive modularity gain until all such moves are exhausted. At this stage (level 1) there are 1082 clusters with the largest including 278 patents and the smallest 2 patents. Then network of clusters of patents is constructed and the aggregation procedure is re-iterated. The second iteration leaves us with 246 clusters (max size 495, min size 4). Further iterations are repeated until modularity stops increasing, which in our case occurred after 4 iterations. At the end we were left with 77 distinct clusters, with sizes ranging from 597 patents to just 4 patents. The distributions of cluster sizes at each level are shown at the right panel of Figure 6. The right panel of Figure 6 shows the final result as a adjacency matrix of network sorted by identified clusters (the clusters sorted in the decreasing order according to their size).

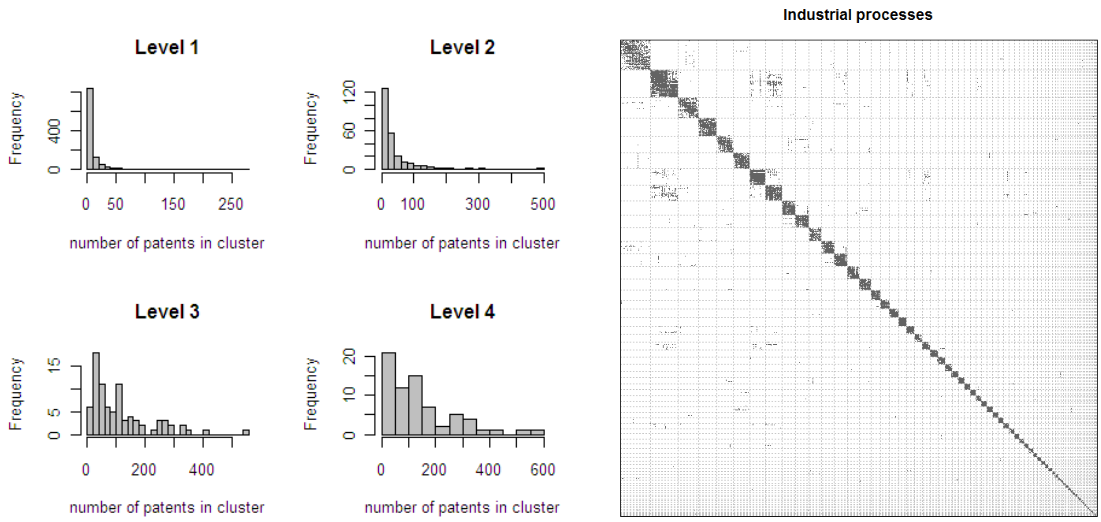


Figure 6 Distribution of clusters size by level (left) and block-matrix (right).

We then proceed to examine the priority dates and ownership of the patents in the various clusters. As far dates as concerned, we do not detect any pattern; that is, we conclude that cluster do not differ for the age of the patents therein, with the only exception of very small clusters. The left panel of Figure 7 reports the distribution of patents by priority date in the 20 largest clusters: we notice that,. while there is some variation in priority years among the clusters, all large clusters show a fairly similar time profile, one that does not significantly differ from overall distribution of patents by priority dates shown in the left panel at Figure 5.

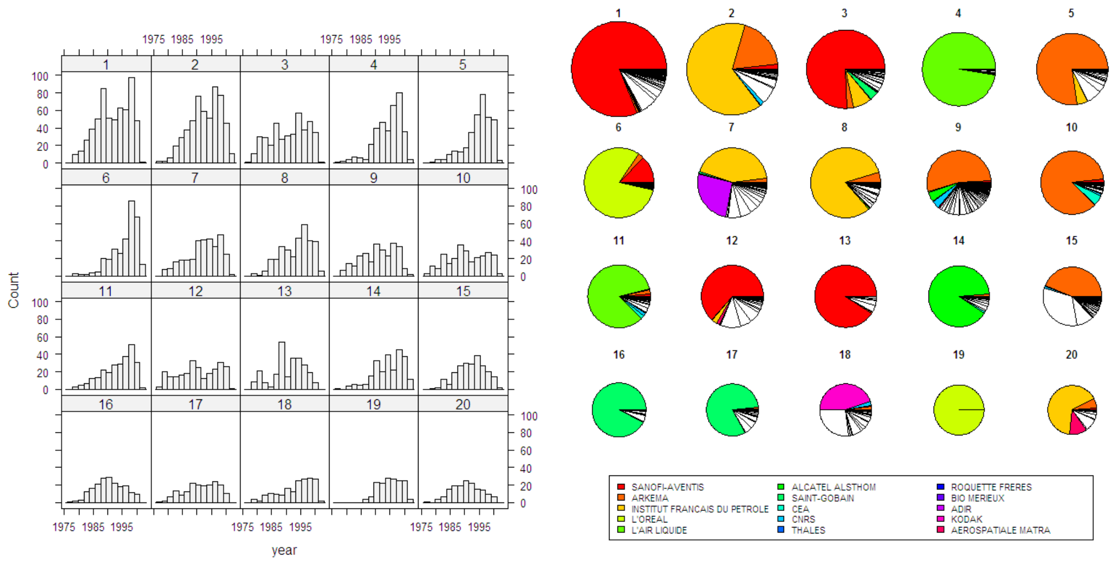


Figure 7 Distribution of patents in clusters by priority date (left) and applicants (right).

By contrast, the identity of patent applicants differs significantly across clusters. The right panel of Figure 7 reports the ownership shares for several large applicants, for the 20 largest clusters (the size of a pie chart is proportional to the size of the cluster it represents). Notice that most of the clusters are dominated by one applicant only; for example Sanofi-Aventis dominates clusters 1, 12 and 13; while cluster 19 consists exclusively of patents by L'Oreal. If we explore deeper into the hierarchy of the clusters (by breaking clusters obtained at iteration 4 back into those obtained at iteration 3 and so on) we would find even more homogeneous compositions.

This result is not entirely surprising, in that it confirms that a cluster of patents is tightly connected by the virtually the same set of inventors who work together time and again precisely because they all work for the same company. But it also confirms previous findings (e.g., by Breschi and Lissoni, 2009) which suggest that inventors move rarely across firms; and that the task of connecting the whole network falls entirely onto the shoulders of a few "mobile" inventors, that is on the few inventors who work for different applicants.

Such mobile inventors may either be R&D employees who change employer; but also researchers affiliated to organizations, such as universities or public research organizations, which engage in contract or collaborative research, and do not always share in the resulting intellectual, as we have seen to be the case with French universities and the CNRS. Therefore, we expect that a disproportionate number of mobile inventors to be academic and CNRS inventors.

5.2 Cohesive subgroups and the position of academic and CNRS inventors

First, we repeat the steps described in the previous section for the largest components of the networks of inventors in the other technological fields with both high presence of academic and

CNRS inventors and small world features. Blockmatrices for the weighted networks of patents are shown at Figure 8.

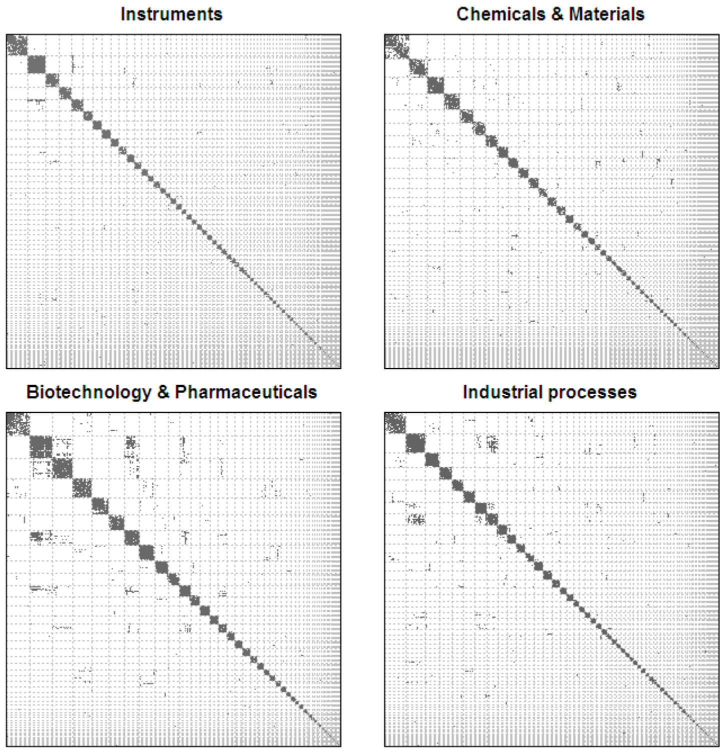


Figure 8 Blockmatrices for the largest components of networks of patents in four technological fields.

The most intuitive way to characterise the position of an inventor with respect to the various patent clusters consists in counting the number of distinct patent clusters the inventor belongs. Formally, let U be the set of all inventors' communities K, then the individual score of inventor i is

$$S_0(i) = \sum_{K,K' \in U} I(i \in K \cap K')$$

where $I(.)$ is the indicator function.

S_0, however, has two drawbacks. First, it does not take into account the uniqueness of the inventors' position. We may expect that an inventor who is the only one connecting two communities is in general more important than an inventor who connects communities which are connected by a number of other inventors. We can correct for this by dividing $I(.)$ by the number of all inventors lying at the intersection of the communities. Thus, when an inventor is found to be only one connecting two communities he/she will receive the full score; but in case the two communities were connected by a total k inventors, the individual score would be just $1/k$. We then sum we sum up the scores and obtain:

$$S_1(i) = \sum_{K,K\prime \in U} \frac{I(i \in K \cap K')}{|K \cap K'|}.$$

The second problem with S_0, one that neither S_1 solves, derives from the fact that it gives the same weight to any connection independent of the size of the connected clusters, while we may wish to attach more importance to connecting larger clusters rather than small ones. A measure that assigns different weights according to the size of the connected clusters is as follows:

$$S_2(i) = \sum_{K,K\prime \in U} \frac{I(i \in K \cap K')}{|K \cap K'|} |K| \cdot |K'|.$$

where $|K|$and $|K'|$ represent the numerosity of the connected clusters

Table 9 reports the S_0, S_1 and S_2 scores for academic, CNRS and other inventors who have signed at least one patent between 1994 and 2004. We notice that academic and CNRS inventors do not exhibit higher values for $\mathbf{S_0}$, which may be related to the fact that they sign a lower number of patents than most other inventors. However, it appears that academic and especially CNRS inventors are more likely to connect otherwise distant communities of inventors as witnessed by the higher values they score for $\mathbf{S_1}$;. We get a similar result when examining $\mathbf{S_2}$, which suggests that academic and CNRS inventors are more likely to connect larger communities than those connected by non-academic inventors. Interestingly, in Instruments and Industrial processes the relative ranking of academic and CNRS inventors reverts when we use size adjusted score $\mathbf{S_2}$, which indicates that on average academic inventors connect larger clusters than their counterparts from CNRS. In Pharmaceuticals and Biotechnology academic inventors rank higher than CNRS researchers both in $\mathbf{S_1}$ and $\mathbf{S_2}$, while in CNRS researchers have higher scores than academic inventors in Chemicals & Materials.

Table 5 S_0, S_1 and S_2 scores for inventors in the largest components*

TECHNOLOGICAL FIELDS		S_0	S_1	S_2
	Non-ac. inv.	0.083	0.038	336.9
Instruments	Ac. inv	0.075	0.066	972.5
	CNRS inv	0.130	0.117	542.8
	Non-ac. inv.	0.151	0.038	2968.1
Chemicals. Materials	Ac. inv	0.190	0.098	6665.7
	CNRS inv	0.313	0.132	7961.7
	Non-ac. inv.	0.160	0.035	1260.0
Pharmaceuticals. Biotechnology	Uni inv	0.151	0.087	2613.9
	CNRS inv	0.066	0.013	1423.1
	Non-ac. inv.	0.197	0.053	439.9
Industrial processes	Uni inv	0.250	0.207	1981.6
	CNRS inv	0.382	0.265	1655.9

*Only for inventors signed least one patent between 1994 and 2004.

These results suggest that both university and CNRS scientists in France occupy key positions in the network of inventors. They do not only breed more contacts than the average inventor in comparable technological fields, but also their contacts are more strategic. In particular, they play a key role in contributing to the small-world features of four out of the five technologies they contribute to, as they stand in between inventors and inventors' teams with no direct connection to each other, and possibly no connection at all.

6. Conclusions

The literature on social networks suggest that positions such as those taken by academic inventors in the networks examined here are indicative of an important role in the mediation and diffusion of information and knowledge. Breschi and Lissoni (2004), for example, make use of patent citations as indicators of knowledge diffusion and find that the distance between any two inventors in the same network greatly affects the probability of a citation link to exist between the two inventors' patents. To the extent that academic inventors, being central, are also relatively close to many other inventors in the network, one can expect their patents to be widely cited, and information therein widely diffused.

References

Balconi M., Breschi., Lissoni F. (2004), "Networks of inventors and the role of academia: an exploration of Italian patent data" , *Research Policy* 33/1, pp. 127-14

Barber M.J.(2007),"Modularity and community detection in bipartite networks", *Physical Review E*, 76, 066102

Blondel V.D., Guillaume J.-L., Lambiotte R., Lefebvre E. (2008), "Fast unfolding of communites in large networks", *Journal of Statistical Mechanics: Theory and Experiment*, pp. 1742-5468.

Breschi S., Lissoni F. (2004) "Knowledge networks from patent data: Methodological issues and research targets", in: Glänzel W., Moed H., Schmoch U. (eds), *Handbook of Quantitative S&T Research*, Kluwer Academic Publishers

Breschi S., Lissoni F. (2009) "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows," Journal of Economic Geography, 9(4), pp. 439-468

Burt R.S. (1992), *Structural Holes: The Social Structure of Competition*, Harvard University Press

Csardi G., Nepusz T. (2006), "The igraph software package for complex network research", *InterJournal*, *Complex Systems*, p. 1695.

Frank K.A. (1995), "Identifying Cohesive Subgroups", *Social Networks*, (17): 27-56Fortunato S. (2010), "Community detection in graphs", *Physics Reports*, 486, pp.75-174.

Guarisco S. (2009)

Guimerà R., Sales-Pardo M., Amaral L.A.N. (2007), "Module identification in bipartite and directed networks", Physical Review E, 76, 036102

Lissoni F., Llerena P., McKelvey M., Sanditov B., (2006b2008), "Academic Patenting in Europe: New Evidence from the KEINS Database", *Research Evaluation* 16(2), pp. 87-102paper presented at the UNI-KNOW / RIDE workshop <Universities as Knowledge Environments of the Future>, Chalmers University, Göteborg 11-12 December

Lissoni F., Lotz, P., Schovsbo,J., and A. Treccani (2009) "Academic Patenting and the Professor's Privilege: Evidence on Denmark from the KEINS database", Science and Public Policy 36/8, 2009, pp. 595-607

Lissoni F., Montobbio F. (2010)

Lissoni F., Sanditov B., Tarasconi G. (2006a2008), "The Keins Database on Academic Inventors: Methodology and Contents", *CESPRI WP* 181 (http://www.cespri.unibocconi.it/workingpapers)

Llerena (2010)

Molloy M., Reed, B. (1995), "A Critical Point for Random Graphs with a Given Degree Sequence", Random Structures and Algorithms, 6 , 161-179.

Moody J., White D. (2003), "Structural Cohesion and Embeddedness: A Hierarchical Concept of Social Groups", *American Sociological Review*, 68 (1): 1–25.

Mowery D.C., Sampat B.N. (2005), “The Bayh-Dole Act of 1980 and University-Industry Technology Transfer: A Model for Other OECD Governments”, *Journal of Technology Transfer* 30, pp. 115-127

Newman M.E.J. (2001), “The structure of scientific collaboration networks”, *Proceedings of the National Academy of Science USA* 98, pp.404-409

Newman M.E.J. and Girvan M. (2004), "Finding and evaluating community structure in networks", *Physical Review E,* 69, 026113.

Snijders T., Nowicki K. (1997) “Estimation and prediction for stochastic block models for graphs with latent block structure”, *Journal of Classification*, 14 (1997), 75 - 100.Thursby J., Fuller A., Thursby M. (2006), “US Faculty Patenting: Inside and Outside the University”, mimeo

Verspagen B. (2006), “University Research, Intellectual Property Rights and European Innovation Systems”, *Journal of Economic Surveys* 20 (4), pp. 607–632

Wasserman S., Faust K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge University Press.

Watts D.J. , Strogatz S.H. (1998), “Collective dynamics of 'small-world' networks”, Nature 393, pp. 440-442.

Zucker L.G., Darby M.R., Armstrong J. (2002), “Commercializing knowledge: University science, knowledge capture, and firm performance in biotechnology”, *Management Science* 48 (1), pp. 138-153

APPENDIX A: ACADEMIC INVENTORS' TEAM SIZE

Let q_m for m = 1,2.. be the observed distribution of sizes for all teams (academic and non-academic), i.e. the probability to extract at random from our data a team of size m, and p be the probability to select at random from our data an academic inventor.

If the inventors' teams were assembled at random, then probability for a team of size m to be "academic" (i.e. to include at least one university researcher) would be equal to

$$\Pr\{\text{team of size } m \text{ } is \text{ academic}\} = q_m(1-(1-p)^m).$$

It follows that the "baseline" distribution defined as the distribution of sizes of "academic" teams when research teams are assembled randomly, b_m, would be

$$b_m = \frac{q_m(1-(1-p)^m)}{1-\sum_k q_k(1-(1-p)^k)}$$

(in the limit of small p this expression reduces to $b_m = q_m m/\langle m\rangle$).

Then we can calculate it for all technological fields, compare to it our observed distribution of academic and CNRS teams, a_m, in the same field and examine whether the two differ.

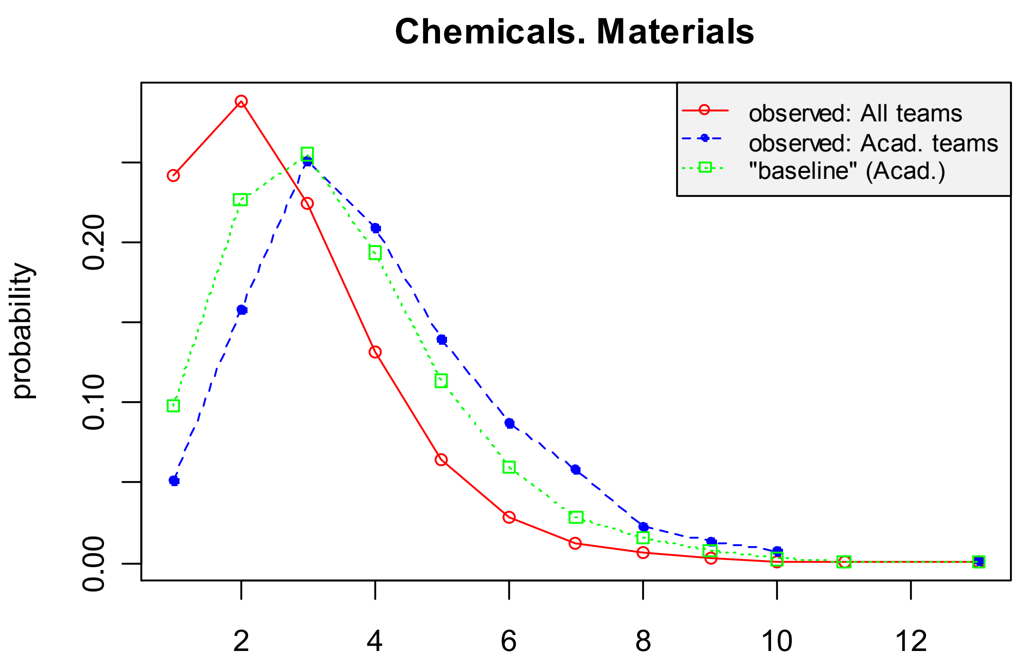


Figure A 1. Observed and "baseline" distributions of the team sizes in *Chemicals & Materials*.

As an illustration consider academic patents in technological field *Chemicals & Materials*. Figure A 1 shows the distribution of all teams (q_m), observed distribution of academic teams (a_m) and corresponding "baseline" distribution of (b_m). At the first glance the observed distribution of academic teams (blue solid circles), differ widely from the distribution for all teams (red hollow circles). However, once we make the correction for the statistical effect, and calculate the “baseline” distribution (green squares), the difference between academic teams and all teams “baseline” distribution become less apparent (although even the corrected “baseline” distribution is still more right-skewed than the observed one).