



HAL
open science

Structural dynamics of the French aerospace sector: A network analysis

Johannes van Der Pol

► **To cite this version:**

Johannes van Der Pol. Structural dynamics of the French aerospace sector: A network analysis. 2015. hal-01284993

HAL Id: hal-01284993

<https://hal.science/hal-01284993>

Preprint submitted on 8 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Structural dynamics of the French aerospace sector: A network analysis[☆]

Johannes Van der Pol^a

^a*GREThA, UMR-CNRS 5113*

University of Bordeaux. Avenue Leon Duguit, 33600, Pessac, France

Abstract

The focus of this paper is on the link between network structure and the financial performance of the individual firm. Under the hypothesis that firms access diverse and valuable knowledge through collaboration we analyze how firms pick their collaborators and how knowledge flows impact the financial performance of the firm. First, the evolution of the structure of the collaboration network of the French aerospace sector is analyzed between 1980 and 2013. The global structure is identified and, using an ERGM and clustering identification, the structure of the network is explained. Second, a panel regressions identifies a link between the position of the individual firm inside the network and their financial performance.

JEL classification: L25; C23; D85; L14; C20

Keywords: Network analysis, ERGM, Network evolution, Dynamic network, Small world, Scale free, Technological diversity, Social network analysis, Firm performance

[☆]This project was made possible by support from the VIA-INNO project of the Groupe de Recherche en Economie Théorique et Appliquée (GREThA). The project also received financial support from the "Initiative d'Excellence" of the university of Bordeaux

Introduction

The technological landscape expands continuously. Every day new technologies are developed that are the result of the recombination of previous technologies. With the diversity in technologies increasing it becomes increasingly difficult for firms to master all the technologies they need to innovate, they hence seek collaborations (Pyka, 2002). When one puts all these collaborations together for a given technology, sector or any other criteria we observe an innovation network. During these collaborations firms can exchange knowledge and routines. This knowledge then diffuses throughout the network. The speed and efficiency with which the knowledge diffuses throughout the network depends upon its structure. A densely connected network allows for fast diffusion while sparsely connected network allow for a slower diffusion.

The structure of the network is indeed a vital variable when it comes to knowledge diffusion. In this light the small world structure has been identified as being the most efficient structure ((Verspagen and Duysters, 2004), (Watts, 1999b), (Morone and Taylor, 2004), (Alghamdi et al., 2012)). This observation is however the result of mostly theoretical work ((Cowan and Jonard, 2004), (Baum et al., 2003)) even though the small world structure has been observed empirically ((Ahuja, 2000), (Van der Pol et al. 2015)).

The differences in the structure of the network can be explained by differences in the knowledge base (Orsenigo et al., 2001), or simply by the fact that the sector is organized in a particular manner (production chain).

In order to better understand how sectoral specificities can shape an innovation network and how this impacts knowledge flows I chose to study the French Aerospace collaboration network between 1980 and 2014. An analysis of the evolution of the network is performed to identify the global network structure. Community detection helps to explain this structure. An ERGM model is then applied to verify different hypotheses that explain link creation: triadic closure, technological proximity and patent co-citations.

The importance of knowledge flows are not only relevant at the level of the global network structure but also at the level of the firm. Both the risks that collaborations might entail as well as the benefits from a knowledge diffusion point of view have been studied. In economic theory there appears to be a consensus on the idea that an innovation network is more than the simple sum of its components. The added value of the network resides in the sharing of knowledge between firms which results from the connections made between firms (or the employees inside the firms). Accessing different knowledge sources is thus beneficial for the firm (McEvily and Marcus, 2005), for innovation ((Kogut and Zander, 1992), (Tsai, 2001)) as well as

survival and growth ((Fernandez et al., 2000), (Watson, 2007)).

40 A particular focus has hence been observed on the analysis of the flow of knowledge between firms on both the empirical level ((Hansen, 1999), (Sorenson et al., 2006), (Wasko and Faraj, 2005)) and the theoretical level ((Cowan and Jonard, 2003), (Egbetokun and Savin, 2013)) allowing a better understanding of the manner in which firms aim to benefit from these knowledge flows. This nourishes the idea that firms
45 according to the network position not all firms benefit in an identical manner from knowledge flows. More central firms might have access to a higher diversity of knowledge while firms in densely clustered components might suffer redundancy in knowledge and a lack of diversity. (Burt, 2004)

In this light, the second aim of this chapter is to identify a link between the position
50 of the firm inside the network and the performance of the firm.

The document is organized as follows: Section 1 will present the macro analysis of the evolution of the network structure. Section 2 will provide the ERGM model, in section 3 the analysis of firms performance is conducted. Section 4 concludes.

1. The data

55 Since our focus is on knowledge flows we need data on collaborations that were initialized for the purpose of creating new technologies. We thus create an innovation network from patent data. Whenever two or more firms are present on the same patent a link is created between the firms. All patents were extracted from the Orbit database, the firm names in the dataset were treated by hand to remove any typos
60 and text lost in translation.

We restricted our focus on Patents deposited in France by French companies in order to avoid any problems with data from different patent offices. For instance, the USPTO tends to cite more intensely than the other offices while the German firms make a heavier use of utility models. Restricting our dataset allows us to avoid biases
65 in these aspects.

In order to select patents relative to airplane technologies a query was constructed using a combination of keywords and IPC codes. We found that using only keywords resulted in a heavy percentage of false positives while selecting patents according to NACE codes was too restrictive. The combinatory method allows us to focus on
70 all the different technologies that make up an airplane. After all, an airplane is the perfect example of a multi-technology product (Prencipe, 1997).

Building such a query does require specific knowledge about the technologies inside an aircraft and their corresponding keywords and IPC codes. The query we used



Figure 1: The aerospace collaboration network as of 2014. Node size is proportional to the number of collaborations, colors correspond to structural clusters identified by a maximization of modularity.

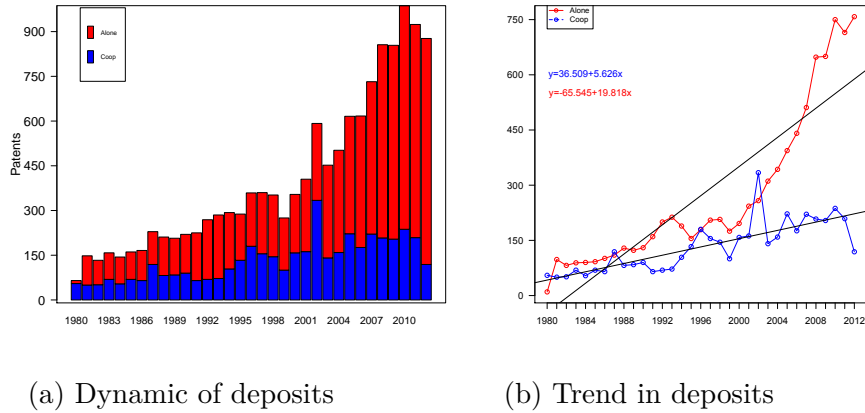


Figure 2: Evolution of the number of patents and the corresponding trend. A distinction is made between the number of patents deposited alone (red) and the number of patents deposited by collaboration (blue)

75 here was provided by the VIA-INNO platform¹ and is the result of repeated discussions between aircraft engineers and the platform to ensure viable results. The query resulted in a dataset of 11992 patents with a priority date between 1980 and 2013. 9544 (79,59%) patents were deposited by a single firm, 2448 (20,41%) patents were subject to a collaboration. From the 2448 patents we identified 4369 cooperations between 1309 companies during the time period (1,78 cooperations on average per patent). Aggregation of these collaborations results in the network in figure 1.

85 Figure 2 shows the evolution of the number of patents deposited between 1980 and 2013. In figure 2a. we distinguish between patents deposited by one firm and patents that are the result of a collaboration. Figure 2b. shows a clear positive trend in both patenting and collaborative patenting in the aerospace sector. Similar observations have been identified in other sectors such as biotech and software by (Pyka and Scharnhorst, 2009), (Gulati et al., 2011) and (Salavisa et al., 2012) . The trend for patenting alone is however much higher than for cooperative patenting: 5.626 « 19.82.

90 We observe an important increase in the number of patents from the year 2000 onwards. This can be explained partially by the commercialization of the Airbus A380.

¹Plateforme d'intelligence économique labellisé centre d'investissement sociétale par l'initiative d'excellence de Bordeaux dans le cadre des investissements d'avenir de l'Etat Français (Website)

A particular aspect of the aerospace sector is the fact that there are mass patent deposits after the commercial release of an airplane which might explain some of the variance in the dataset.

95 The trends clearly show an increase in the number of collaborations over time. We will now turn our attention to the analysis of the network resulting from these collaborations.

2. The French aerospace network

2.1. Network structure identification

100 In order to identify the structure of the network we will track the evolution of the network from 1980 onwards. This will allow us to have a clear vision of the structuring of the network.

Figure 5 reports the number of new firms that enter the network each year. The variance is explained by the previously discussed patenting behavior in the sector. 105 The evolution of the number of nodes (figure 3) is computed using a sliding window of 5 years. This allows to keep track of the active firms in the network. This shows us that the network increases in size over the period with a decline during the last period (note that 2008 implies the frame 2008-2013). The decline can be explained by two factors. First, a small decline in the number of deposits in the last couple 110 of years (figure 2a). Second, the decline in the number of firms might be explained by the "Power8" program launched by Airbus in order to optimize their production chain which resulted in a decrease in the number of suppliers.

We will check for two particular structures that have been observed and studied both theoretically and empirically: A small world structure and a scale-free structure. 115 Both structures have different implications from a knowledge diffusion perspective. Previous research shows that small worlds are the optimal structure for knowledge flows. We will start by checking for a small world structure.

The literature on economic networks stresses the importance of the overall structure of the network for knowledge flows. Two general network structures have been 120 identified empirically and studied theoretically: the small world structure (SW) and the scale-free structure (SF). In this section we will check the global network structure for both the SW and the SF structure. Even though we know the network will exhibit only one of the two structures it is important to highlight why it has one structure and not the other. We will start by checking for a SW structure.

125 2.2. Small world identification

In order to check if our network has a small world structure we follow a methodology presented by (Gulati et al., 2012). Small world structures are defined by a

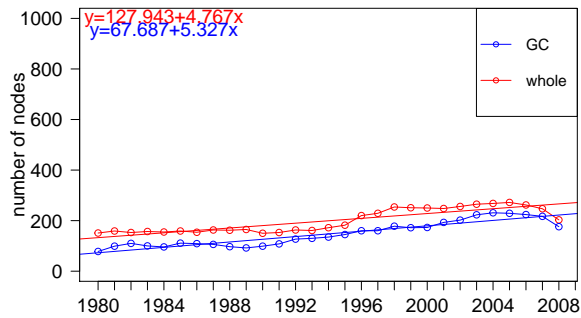


Figure 3: Evolution of the number of nodes with a 5-year sliding window (i.e 1980 → 1980-1984). "GC" is the giant component of the network, "whole" the giant component with all the smaller components

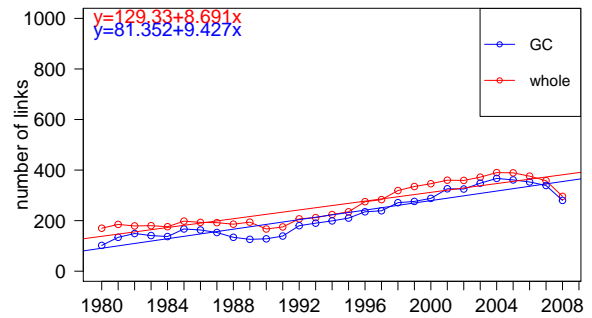


Figure 4: Evolution of the number of links with a 5-year sliding window (i.e 1980 → 1980-1984). "GC" is the giant component of the network, "whole" the giant component with all the smaller components

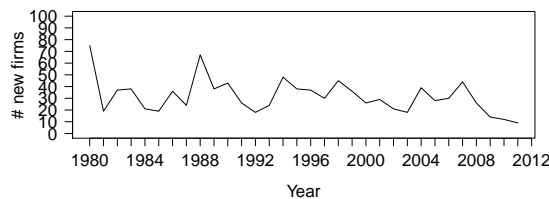


Figure 5: Evolution of the number of new firms entering the collaboration network each year.

low average distance and a high clustering coefficient. The Clustering coefficient of a network is defines as the ratio of observed triangles in the network to the number of possible triangles. The average distance is simple the average number of links between any two nodes in the network.

Since nodes can be added each year we need to make sure that a decrease in clustering is the result of less firms connecting in triangles and not the simple result of an additional node that reduces the overall clustering coefficient. The coefficients are hence normalized and compared to a random network with an identical number of nodes and links.

The theory behind small worlds is that random networks have low clustering while empirical networks have higher clustering. The latter is the results of social / economic / geographic / ... motivations of the entities inside the network. As such, a

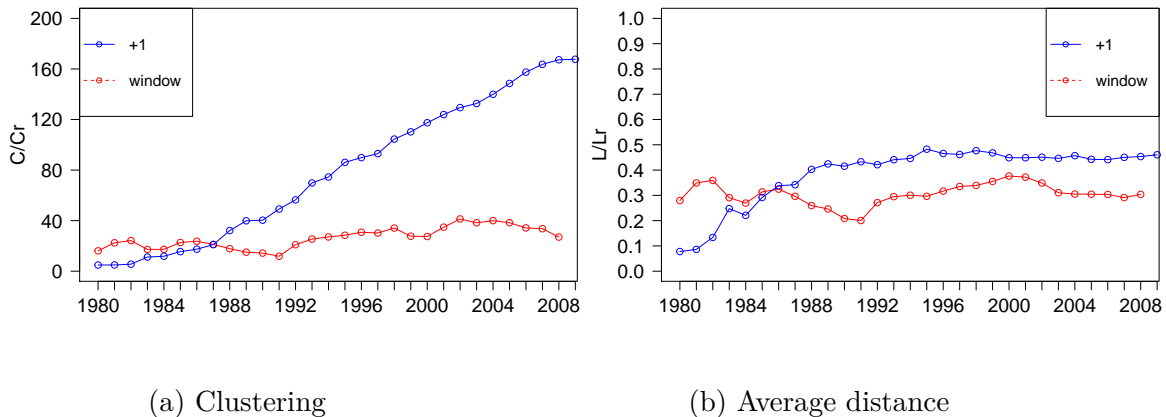


Figure 6: Evolution of the Small World indicators between 1980 and 2013. +1 indicates the additional method in which each year data is added while window indicates a 5-year sliding window.

140 network is a small world if its clustering coefficient is higher than that of a random graph of identical dimension (i.e same number of nodes and same number of links). This would hence imply that the graph is not random and that there are some underlying rules dictating the creation of ties in the network.

As for the average distance, we want it to be roughly identical to that of a random graph. We note C_r (L_r) the clustering coefficient (path length) of the random network and C (L) the clustering (path length) of the empirical data.

We hence need to observe $\frac{C}{C_r} \gg 1$ and $\frac{L}{L_r} \approx 1$.

The evolution of the network was considered in two manner: using a 5-year sliding window and a method in which data was added year after year. The results are reported in figure 6a and 6b.

Figure 6a shows that the clustering coefficient trends strongly away from 1, indicating that the clustering observed in the networks increases faster than clustering in a random network of identical dimension. This is the case for both methods, showing that even when we remove firms that are no longer part of the network, the clustering stays higher than random.

The average distance however is too small, since $\frac{L}{L_r} < 1$. The low distance can be explained by densely connected clusters that are interconnected by large central firms, as we discussed before, the assemblers and the CNRS. The results allow us to conclude that the network is not a small world, not in the aggregate method nor in the time-window network. The aerospace sectors is hence structurally different from the Biotech sector (van der Pol, 2015b) and the computer sector (Gulati et al.,

2012).

The structure of the network seems to be highly correlated with the structural specificities of the aerospace sector. Indeed, knowledge stays within the clusters since
165 specific knowledge is developed inside each cluster. There is no need for knowledge to flow between all clusters since many parts are not technologically connected (wheels and wings for instance). Communication and knowledge flows are necessary between firms inside clusters since the parts developed by firms in clusters need to interact and need to be compatible. The most central firms hence benefit from the
170 most knowledge flows since they have to assemble the different parts of the plane. We hence find diverging conclusions from the results in (Gulati et al., 2012) who identified an inverted U-shape in the smallwordiness of their network. Our observations are the result of the specificities of the Aerospace sector. Gulati's observations were found in the automatized computer industry which resulted in a completely
175 different network structure.

We can conclude here that there is a high tendency for firms to cluster which confirms our previous observation that firms were organized in interconnected clusters. The structure also appears to stay relatively stable when it comes to these
180 two indicators, especially in the time-laps network. In the 90' has started a radical change in the organization of the sector resulting in many suppliers exiting the sector which has as a consequence a lower number of collaborators. These collaborators collaborate more intensively resulting in a more stable structure.

185 Since the network is not a small world we will check for another structure. We can see in figure 1 that the networks looks like an interconnection of firms with a high number of collaborations, this could point to what is called a "Scale Free" structure. We will check for this structure now.

2.3. Scale-free structure identification

Scale free networks represent a particular structure that has been observed empirically at different levels such as the world wide web, in lexicography or collaboration networks. A scale-free structure is a network with a particular (asymmetrical) degree distribution.

A network is build up from firms which can have a different number of links. We start by computing how many firms have one link, how many have two and so on, until we have classified all firms in the dataset. This gives us the degree distribution of the network. What we observe is that there are a few firms with a large number of links while there are many firms with a small number of links. We then transform

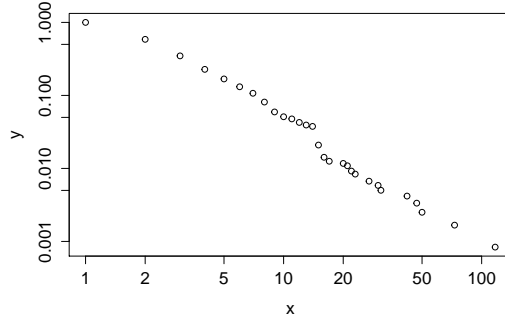


Figure 7: The cumulative frequency distribution of the aerospace network in 2014

this distribution into a cumulative distribution graph, as shown in Figure 7. (the scale is log-log). The idea behind a scale free network is that when we increase the degree by a factor of k then the frequency drops by a factor of k^α . This is verified for all degrees k in the network. Since this decrease is identical for all values of k , the distribution is called scale free. Typically a scale free distribution follows a power law:

$$P(x) = c \cdot x^{-\alpha} \tag{1}$$

190 Where c is a normalizing constant. The idea is to fit a function to the observed data, if a line is fitted we check the quality of the fit. In the case of discrete data we use a p.value to assert the quality (for continuous data the KS statistic would have sufficed).

The cumulative density distribution is computed for each 5-year window and fitted
 195 with a power-law and a log-normal function to the data. Since it is basically possible to fit a power law to any distribution, we need to check two values carefully. The first value is referred to as x_{min} . It represents the lowest density for which the function is fitted. If $x_{min} = 1$ then all possible densities.

are included in the fit. If $x_{min} = 5$ then density values 1 through 4 are ignored and
 200 the power-law is fitted to the data for densities > 5 . For instance, in figure 9 the cumulative frequency distribution of the network at is shown at 4 points in time. In 1986 the log linear fit starts at a density of 1, hence $x_{min} = 1$ while the power law fit starts at $x_{min} = 4$. In order to check if the fit is statistically valid a goodness of fit check is performed. A p.value is hence computed using bootstrapping with the
 205 R package *poweRlaw* (Gillespie, 2015). If the p.value exceeds a fixed critical value we can reject H_0 . Figure 9 contains the p.values in the lower left corner. In the first

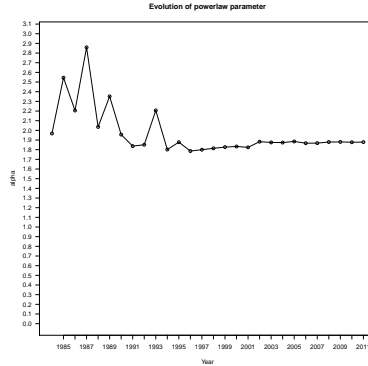


Figure 8: Scale-free structure stabilization

period, starting in 1986, we can clearly see that the log linear fits the data perfectly with an $x_{min} = 1$. The power-law however fits the data poorly. We hence conclude that the network is not a scale-free network in the early stages of the network.

210 As time passes the power-law becomes significant, in addition the parameter of the power-law, α , stabilizes from the year 2000 onwards (see figure 8). The constant "c" changes only with x_{min} (see the appendix for computation). Since this value also stabilizes we can conclude that the structure of the network stabilizes over time.

215 The fact that we have a scale-free structure here shows that the network is an interconnection of a small number of large stars with a large number of small stars. This implies that the probability that a firm will cooperate is proportional to the number of cooperations the firm already has. At the same time it implies that firms with many cooperations today have a high probability that they will cooperate in the future. New links are created between firms that already cooperate a lot inside
 220 the network. This proves the importance of core actors inside the network.

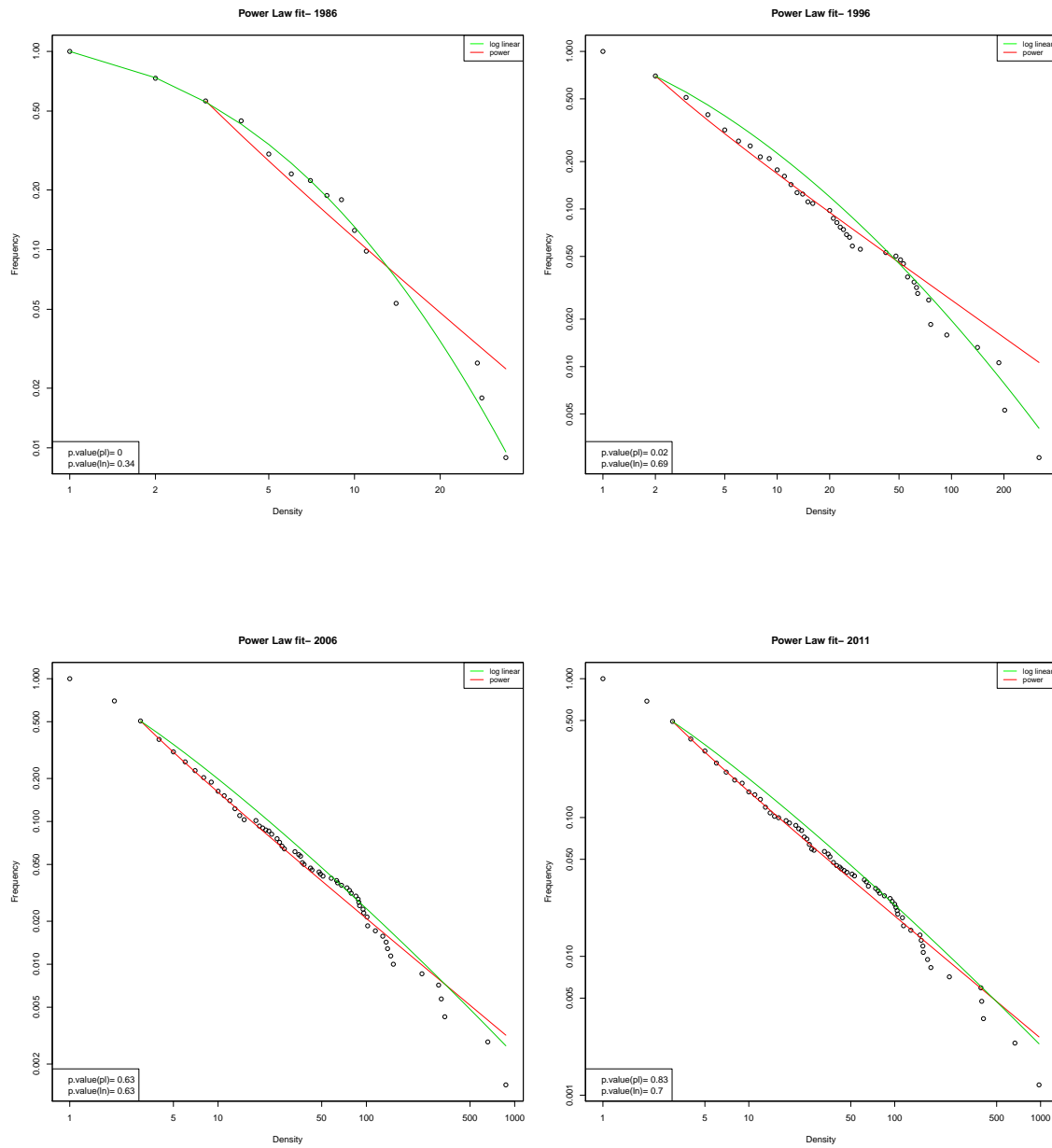


Figure 9: Power law fit over time

In order to understand this particular structure we turn to community identi-

fication. The identification of clusters inside the network might shed light on the interconnection of stars we have just identified.

2.4. Cluster identification

225 The previously identified dataset leave us with over 4300 collaborations. The collaborations allow us to generate a network by creating a link between all firms that have deposited a patent together. The result is shown in figure 1. The bigger the size of the node the more collaborations the firm has. The coloring is the result of a community detection algorithm based on modularity. Modularity measures how
230 well defined communities are inside a graph. Modularity gives a value between 0 and 1, the more the value tends towards 1 to more clearly defined the communities are (Newman and Girvan, 2004). For the result to be significant one expects a value of at least 0.6.

An algorithm introduced by (Blondel et al., 2008) was used to identify these com-
235 munities using the open-source program Gephi (Bastian et al., 2009).

This community detection algorithm identifies communities inside a network purely based on the structural properties of the network. It starts by assigning each node with a community, it then selects a node at random and create a community with one of it's direct neighbors. The neighbor with whom it will create a community is
240 the one that will maximize the modularity of the graph. This step is continued until maximum modularity is achieved. This method has the advantage of detecting automatically the number of communities (clusters) in the network while other methods ask the user for a fixed number of communities to be identified.

The results should however be handled with caution. The random component se-
245 lects a node at random. It is possible that different results emerge if a different node is chosen at the start of the algorithm. In fact, the sequence of choice of the nodes plays an important role in the detection of the communities. We hence ran the algorithm several times to make sure the same communities were detected on average.

250 The results are rather interesting given that the communities were clearly defined and easy to interpret. Different communities were identified around the following firms:

- Hispano Hurel: Nacelles
- Rhodia: Chemicals
- 255 - Thompson: Seating
- Messier Bugatti: Landing and braking.
- Pechiney Rhenalu: Structural elements (aluminium)
- Alcatel Lucent: Avionics and communication systems

260 These clusters suggest local technological development according to different parts included in the production of an aircraft. This allows us to understand the previously identified scale-free network structure. The large assemblers (Airbus, Snecma and Thales) and the CNRS have a large number of links connecting them with first order suppliers which in turn have their own clusters in which they are densely embedded. This observation coincides with the industrial organization of the sector, which is indeed rather hierarchical. Airbus, at the center, designs the aircrafts while externalizing large portions of the production process to first order suppliers (Frigant et al., 2006). The latter will work with other, second order suppliers. As such there are not many competitors but competition is tough between the few (Niosi and Zhegu, 2005). The sector has undergone a significant restructuring in the 90' and the 2000's resulting in the specialization of some suppliers while others diversified their production to include other sectors (Frigant et al., 2006). In addition, the sector has high barriers to entry, mainly because of high level of knowledge required. The sector need an influx of cutting-edge technologies and hence close collaboration with fundamental research. The collaboration network that we observe here reflects 275 these sectorial aspects: in a central position we find the CNRS (National Centre for Scientific Research) providing an influx of fundamental science to the large manufacturers and first order suppliers. While clusters exist around the first order suppliers connecting specialized and diversified suppliers. This results in a particular network structure that is made up from an interconnection of clusters. The overall structure of the network resembles a connected caveman structure (Watts, 1999a) in which each 280 specific part of the airplane is developed in it's own cluster. In terms of knowledge these firms need to collaborate with a large number of firms from different clusters in order to assemble an aircraft. While there is no need for direct knowledge flows between the landing and braking system and the nacelle manufacturer, Airbus needs knowledge on both technologies to assemble the final product. 285 The exception being that some firms connect all the clusters. Airbus has this central position since it needs to absorb knowledge from all clusters. Very little knowledge flows seem to exist between clusters, while there is a necessity for transfer intra-cluster. 290 Innovation in the aircraft industry is the result of an interplay of technology push and market pull (Dosi, 2000). On the one side aircraft manufacturers aim at making their aircrafts more cost efficient while there is a demand for governments to reduce noise and make planes more eco-friendly.

295 We now have an idea of the manner in which the network is structured and to what extend sectoral particularities play a role in the structuring of the network. I

now turn my attention to the mechanisms of tie formation at play in the formation of the network. We hence switch from a sectorial perspective to a micro level vision.

3. Tie formation mechanism identification

300 Ties inside a network are rarely created at random. The objective of this section is to test three tie forming mechanisms that explain the observed network structure.

Technological proximity between firms is a requirement for cooperation. If firms are too close together, they work on the same technologies and hence would not want to collaborate. As the technological distance increases the complementarity of the knowledge bases of the firms increases. This results in an increase in the probability of observing a collaboration. This complementarity does however reach a point where technologies becomes too distant and the complementarity decreases. This results in turn in a decrease in the probability of cooperation. These statements induce the first hypothesis to test:

Hypothesis 1 *There is an inverted U-shape relation between the probability of a collaboration and the technological proximity of two firms.*

We stated previously that social capital played an important role, especially since the "power8" program launched to streamline production. It would hence seem important for firms to closely collaborate and collaborate with firms they know. This allows a homogenization of work methods. We hence hypothesize:

Hypothesis 2 *Collaborators of collaborators have a higher probability to collaborate than firms without a common connection.*

We stated that firms inside clusters need to collaborate in order to render their products compatible. From a technological point of view this would imply that firms need to cite each-others patents in order to protect their technologies. I hence hypothesize:

Hypothesis 3 *Firms that cite each-other have a higher probability of collaborating than firms who do not.*

325 In order to test these hypotheses an Exponential Random Graph Model is used. Before going into the details of the model we need to specify some elements for hypothesis 1.

3.1. Technological proximity

Many measures of technological proximity exist, some are based on patent cita-
330 tions (Chang, 2012), (Marco and Rausser, 2008), (Mowery et al., 1998) while others
use IPC codes (Jaffe, 1986), (Breschi et al., 2003). The idea is that the different
technologies firms work on are not chosen at random, they co-exist because they
have factors in common (Teece et al., 1994). This idea has led to different measures
of technological proximity between firms, the most prominent was initiated by (Jaffe,
335 1986) further developed by (Breschi et al., 2003). Finer measures exist, see for in-
stance (Bar and Leiponen, 2012) or (Bloom et al., 2013).

For the present paper it is chosen to use an IPC based measure of technological
proximity. We will use a slightly different measure than the ones previously cited,
even though based on IPC codes. Our aim is to provide the likelihood of a cooper-
340 ation based on the technologies mastered by firms. Therefore we hypothesize that
firms cooperate on technologies that are closely related in order to ensure proper
incorporation of new technologies into an aircraft. As such having 1 technology in
common is motive enough for two firms to cooperate. If we were to use one of the
more common measures the prediction could be biased.

345

3.1.1. A measure of technological proximity

An IPC takes the following form: B64C1/18. Each part of the code (B, 64, C,
1,/18) indicates a practical classification. Hence B stands for Performing operations
and Transporting, B64 reduces the technologies to Aircraft, Aviation and Helicopters,
350 B64C denotes Airplanes and Helicopters, B64C1 are Fuselages, wings etc. B64C1/14
are windows. The longer the code the more precise the technology. We hence use
the full length of the IPC-codes. When a firm deposits a patent we can deduce from
the IPC codes what a firm has been working on and what technology it masters.
We base our measure of techno- logical proximity on an analysis of IPC codes. The
355 indicator of proximity computes the overlap in IPC codes between two companies.
Figure 10 shows two firms with 3 IPC codes. The numbers in the matrix corre-
spond to the level of proximity. If both firms work on B they will have an overlap of
1, if they both work on B64 the overlap is 2 and so-on. Figure 10. shows an example.

		Firm B		
		B64C/19	B53D/04	C01D/05
Firm A	B64C/19	4	1	0
	B53D/01	1	3	0
	C01F/03	0	0	2

Figure 10: Illustration of the proximity measure used in the ERGM

360 I defend the position that knowledge about one specific technology is enough to initiate a collaboration. The use of complete portfolios would induce a lot of noise in the data. In the end, firms cooperate often for a particular set of skills and not for all the skills used by a firm. A downside of this method is that the dataset is reduced to firms depositing both alone and by cooperation. We can only assume a
365 firm masters a certain technology if it has deposited a patent alone. Cooperation data is then needed to create a network. Firms that only deposit by cooperation are hence excluded from the dataset.

We hence computed a proximity matrix for 176 firms and generated the network that connected them.

370 3.2. Motivation for cooperation

In order to test for a link between technological proximity and probability to cooperate we need some indication that firms do in fact collaborate for knowledge and not just for financial needs.

To verify the latter we have a look at the IPC codes on the patents and try to
375 identify if firms work on specific IPC codes when they cooperate. In order to do so we classified firms in to three groups, 1) firms that exclusively deposit patents by cooperation, 2) Firms that exclusively deposit patents alone and 3) firms that do both. We then extract only the firms that deposit alone, and by cooperation. Firms that only deposit on their own are not connected in the network. Also, in the case of
380 a cooperation we are unable to identify which IPC code is mastered by which firm. An alone deposit is a proof that the firm masters the technology. From the resulting dataset we extracted all alone deposits and all the IPC codes in which a firm deposited alone. We did the same with all the co-deposits. We then have for each firm a list of the IPC mobilized when it deposits alone and a list with IPC codes mobilized when
385 it deposits together with another firm. We compared the list to identify IPC codes that were mobilized only by cooperation and only by alone deposit. The results show that for 27% of all co-deposits mobilize IPC that are not mobilized during alone deposits. In the remaining patents 54% of the IPC codes (on average) we not mobilized during alone deposits. These results clearly show that cooperations induce

390 firms to mobilize specific technologies and hence that cooperations are technologically oriented. We can hence start our analysis of technological proximity as a motivation for cooperation.

3.3. The Exponential Random Graph Model

An Exponential Random Graph Model models the global structure of a network while allowing inference on the likelihood of a link between two nodes. It is basically a modified logistic regression, the models are modified in the sense that they do not require a hypothesis of independence between observations. As such if firm A is connected to B and C , there is a high probability that B knows C through its connection with A . A link between B and C has hence a higher probability than B connecting with a random node. This implies that a link between two nodes depends upon the existence of a link between other nodes. Regular logistic regressions are unable to account for these aspects. These levels of dependence are however vital for the understanding of social and economic networks. In our case we hypothesize that firms with higher technological proximity will have a higher likelihood of connecting and hence proximity plays a vital role in the explanation of the global network structure.

The model to estimate is given in equation 2.

$$Pr(X = x | \theta) = P_{\theta}(x) = \frac{1}{k(\theta)} \cdot \exp(\theta_1 \cdot z_1(x) + \theta_2 \cdot z_2(x) + \dots + \theta_p \cdot z_p(x)) \quad (2)$$

Where X is the empirical observed network, x is the simulated network, θ a vector of parameters, z_i the different variables and $k(\theta)$ the normalizing constant. In short, the probability that the network generated by the model is identical to the observed network depends upon the given variables. If we consider that technological proximity has a role to play, we introduce it as a variable. The model will then generate links while increasing (iteratively) the probability that nodes with higher proximity will connect. This is repeated a certain number of times. If, on average, the network generated is equal to the observed network then we can conclude that proximity plays a role the structuring of the network. For a more complete explanation of ERGM models see (Lusher et al., 2012), (van der Pol, 2015a).

Each variable either increases or decreases the probability of observing the graph we are looking for. ERGMs are hence close to logistic regression except for the hypothesis of independence of observations.

Table 2 shows the regression results, not that these coefficients cannot be interpreted as such. In order to compute the precise impact one needs to transform them

into odds.

425 We see that we used several factors to explain the structure. We stated that technological proximity was a decisive factor in collaboration between firms in the aerospace sector. The models shows that this is indeed the case. Firms with a higher technological proximity have a tendency to work together. More precisely the odds of a link between firms that are technologically close is higher than the odds of a link
430 between firms that are technologically far.

Moreover there appears to be an inverted U-shape to this relation as shows by the significance of the variable *proximity2*. This would imply that firms collaborate if they can learn from one another but if they are too close in terms of technology then the probability of a link deteriorates. Firms that are too close in terms of
435 technologies can consider that the other firm has nothing to offer them and hence prefer collaborating with a firm that has different technologies. Hypothesis 1 is hence verified.

The *altkstar* parameter checks (and controls) for the scale-free structure. Since the parameter is significant we see that the model has correctly identified the scale-free
440 structure we previously found.

Taking the *kstar2* and *triangle* parameter together allows for checking for triadic closure (Lusher et al., 2012). Since both the parameters are significant we conclude that firms with a common node have a higher probability of connecting than firms with no common node. It hence seems that the trust that diffuses through the network as well as the increased performance due to common practices is a motivator
445 for collaboration. Hypothesis 2 is verified.

Finally, co-citations are significant as well. Hypothesis 3 is hence also verified.

We have studied the overall structure of the network and concluded that the
450 structure has the features of a Scale Free network, implying that firms with a high number of collaborations is a predictor of collaborations. The network is build up from different clusters where different technologies are fostered explaining why the structure observed in not a small world. Moreover we identified three tie formation mechanisms at work that give the network the shape that it has. Firms strategically
455 cooperate in for technologies. These collaborations alter their position in the network and hence their exposition to knowledge flows. One could then argue that specific positions inside the network allows firms to be exposed to more and more diverse knowledge flow that they can take advantage off. In the following section I will aim at identifying a link between the position of the firm inside the network and the
460 impact on its performance.

Table 1

	<i>Dependent variable:</i>		
	Network		
	(1)	(2)	(3)
edges	-7.267*** (0.228)	-1.121* (0.626)	
kstar2	0.155*** (0.003)		
degree2		-1.336*** (0.255)	14.467*** (2.970)
edgecov.citation			-20.934*** (1.090)
triangle	3.428*** (0.007)	1.923*** (0.0001)	1.726*** (0.0002)
gwesp	-0.439*** (0.166)		
gwesp.alpha	0.523 (0.385)		
edgecov.proximity2		1.565*** (0.271)	6.620*** (0.345)
altkstar.1.6		-1.864*** (0.172)	
altkstar.1.7			-3.371*** (0.086)
Akaike Inf. Crit.	578722	617651	9813
Bayesian Inf. Crit.	578760	617689	9851

Note:

²⁰*p<0.1; **p<0.05; ***p<0.01

Table 2: ERGM model results

4. Impact of network position of the firm on performance

The objective of this section is to establish a link between financial performance and structural position. The structural position of the firm is important mainly because of knowledge flows. Innovations are achieved by the recombination of knowledge (Schumpeter, 1942). Since the knowledge stock inside a firm expands slowly and diversity decreases over time, external knowledge sources are important. The position of the firm inside the network defines the number and the diversity of knowledge sources to which the firm has access.

We will conduct a panel data analysis to estimate the influence of the position of the firm on its performance.

We hence need to find financial data for the identified firms. From our sample of 1309 depositors we need to eliminate all research institutions, financial institutions and government agencies. For the financial analysis these were removed. We were able to identify 676 firms in our set of 1309 firms. The financial performance of the firm will be measured by the Return On Assets (*ROA*) of the firms:

$$ROA_t = \frac{Net\ Income_t}{Total\ Assets_t} \quad (3)$$

The ROA seems the appropriate measure since the denominator of the ROA includes intellectual property and all capital mobilized for R&D activities. The data will be extracted from the Amadeus database. We run into a problem with the chosen timeframe. Since we have network data over 34 years we would have liked to have 34 years of financial data. This was however not possible due Amadeus' policy. Firms are automatically deleted from the database once they have not transferred any data for 3 years. This means that firms that changed their names during the 34 year period are no longer in the database. Using DVDs from a previous version of Amadeus (between 2000 and 2007) we were able to extract a relatively complete dataset over the years 2000 to 2012. We hence recomputed the network structure to include only links between firms in the dataset.

4.1. Variables and theoretic background

In this subsection we will discuss the different variables included in the analysis.

- *Technological diversity.* We mobilize the Schumpeterian hypothesis that innovation is achieved by the recombination of ideas. This hypothesis implies that firm exposed to a large variety of ideas will have a high potential for innovation

495 (Dosi, 2000),(Cowan and Jonard, 2007). In other terms, the advancement on
the inventive trajectory will be faster than firms with a low diversity. In the
latter case firms risk decreasing returns to innovation. We compute a variable
called "neighborhood diversity" which computes for each year the number of
technologies in the neighborhood of the firm ? Each technology will be an IPC
code. We do not include the IPCs of the focal firm. We wish to understand
500 the influence of the neighborhood, hence we correct for firm's diversity to avoid
polluting our data. This leads to the following hypothesis:

Hypothesis 4 *The technological diversity in the neighborhood of the firm has a positive impact on its performance.*

- *Clustering of the focal firm.* Two theories claim the importance of clustering
505 in a network. The two theories do however oppose each other when it comes
to the sign of the impact. A first theory suggests that having collaborators
work together results in a positive impact on innovation and performance .
The cooperations allow for a better understanding of the functioning of each
firm. This information will allow firms to better organize their innovative
510 activities.The effect is enhanced when cooperations are repeated over time,
the more they know about each other the more efficient the cooperation. The
other theory however suggests that a social lock-in might occur when firm
cooperate too often, they would rather work with people they know rather
than take the risk of finding a partner that is not efficient. This will result in a
515 reduction of the innovativeness of firms, by the means of a stagnation or even
reduction of the diversity of technologies. Instead of cooperating with a firm
that masters new technologies they keep cooperating with firms that master
the same technologies. This leads to the following hypothesis:

Hypothesis 5 *Clustering has a positive impact on the performance of the firm due a better mutual understanding of firms.*

- *Centrality and average distance of the focal firm.* A network connects firms by
creating paths between them. Knowledge flows between firms that are directly
or indirectly connected. A firm with a position on many of these paths has
access to more knowledge flows. This position is measured by the betweenness
525 centrality coefficient which takes into account the position of a firm on path
between other firms (Wasserman, 1994). The higher the centrality of the firm,
the more it is on the crossroads of knowledge flows. The higher the centrality
of the firm, the more it is able to benefit from diverse sources of knowledge.

530 The average distance gives a measure of the average distance a firm is removed from all other firms in the network. The closer it is to all other firms the more beneficial the knowledge flows should be. An argument against this idea is that if the distance is too low there is a high risk of redundancy of information and hence low distance should have a negative influence on the performance of the firm. I hence test the following hypothesis:

535 **Hypothesis 6** *The more central the firm, the better the performance due to an increased access to knowledge flows.*

- *Number of patents of the firm* The number of patents gives an indication of the innovative dynamism of the firm. The more patents are deposited by the surrounding firms the more knowledge they accumulated. The following hypothesis is hence tested:
540

Hypothesis 7 *The more patents in the neighborhood of the firm the stronger the knowledge spillovers to the focal firm.*

- *Number of technologies for the focal firm* The more technologies mastered by the firm the easier it should be for the firm to learn new knowledge which should result.
545

Hypothesis 8 *The absorption capacity of the firm is positively related to its performance.*

4.2. Variable lags and panel regression

We have two different data sources. The financial data from 2012 comes from the performance in the year 2012, the patent data from 2012 does however result from
550 cooperations that took place any time before 2012. In order to perceive an effect of the cooperation on the performance we need to include lags in the patent-related variables. How far back the lags should go depends entirely on the type of information, some have a faster influence on the performance than other do. In terms of lag we will consider that a cooperation is initiated 3 years before the priority date of the
555 patent. This means that the transfer of some types of information may flow from that point on. The effects of the knowledge flow should be visible at about the date of priority of the patent. The effects of the production of the patented technology should be visible (if the technology is indeed put into production) at any point in time from $t - 1$ on.
560

Structural variables: Firms are influenced by the knowledge held within the firm at the moment of collaboration. The diversity is hence lagged to $t - 3$: firms connected

by a patent in 2010 cooperated in 2007 and are hence influenced by the diversity in the firm in the year 2007. However, since it takes time to absorb the knowledge and put it to use the impact on the *ROA* should be observed some time after the initialization of the cooperation, we will consider 3 years. Hence the variable Diversity is not lagged, the same is applied to the number of patents and the number of technologies. All the other variables are lagged at $t - 3$ since the knowledge flows may influence the performance from the start of the cooperation on.

We include 2 types of variables in this regression. Structural variables and technology variables. We have a panel of 1605 observations over a 10 year period. We use a standard linear panel regression to test the influence of the network on the performance of the firm. The previously discussed variables were included with the corresponding lags:

$$ROA_{t,t+1} = Clustering * density_{t-3} + Centrality_{t-3} + AverageDistance_{t-3} + Technologicaldiversity + Numberoftechnologies + Numberofpatents + Numberofcooperations$$

In a first regression we only integrated the variables relative to the position of the firm inside the network (model (1)), a second regression includes only the technology variables (model (2)), the last model show the regression with both types of variables (model(3))

In order to assess which type of regression is adequate for the data several statistical tests were performed. The Lagrange Multiplier Test (Breusch-Pagan) showed that there is presence of panel effects in the data, simple OLS regressions are hence rejected.

We then checked for time fixed effects in the data, by adding a dummy variable for each year and compared the regression results with an F-test, the results show that no time-fixed effects have to be included in the model. A fixed, random and pooled model were then tested against each other, the fixed effects was retained as the best model. Since the data presented serial correlation and heteroscedasticity, we used robust estimates.

The results of the regression are shown in table 2. All variables have a significant impact on the *ROA* of the with the exception of the number of cooperations and the number of patents. The latter observation is rather to be expected. Not all patents have the same value only a small portion of patents have an exploitable value. The number of cooperations shows that not all cooperations have a benefit in terms of knowledge flows. The number of collaborations is higher than the number of col-

laborators. As such it can be interpreted as the intensity of collaborations between firms, i.e how close firms are socially. The impact of social links is an order of magnitude lower than the impact of knowledge transfer by other objects and is difficult to capture.

605

The structural variables are all significant, showing that the position of the firm in the network does indeed have an impact on the performance of the firm. The adjusted clustering measure shows that firms with a higher clustering coefficient perform better. The collaboration of collaborators is hence a positive effect. The idea that working with people who already know each other seems to be validated.

610

In terms of knowledge absorption the central position of a firm is significant. The more central the firm is, the more knowledge it is able to absorb. The measure retained here is the betweenness centrality which measures the extend to which a firm is positioned on the a path between all the firms in the network. The higher the centrality the more favorable the position for knowledge absorption. The Average distance measures how far is firm is positioned from other firms, the further away the less knowledge the firm is exposed to. As such, the negative coefficient of this variable confirms the hypothesis that knowledge flows in the network have a decaying factor.

615

620

The technology related variables highlight the importance of technological diversity. Innovation literature puts forth the idea that innovations are achieved by the recombination of ideas. The diversity of technologies in the neighborhood of the firm should hence have a positive impact on the performance of the firm. The regression shows that this hypothesis is validated.

625

The final variable, the number of technologies mastered by the firm, has a negative impact. In our particular case, i.e the aerospace sector; the firms with the most technologies are suppliers with a specific position in the value chain. The regression show that specialized firms perform better than diversified firms, in a network. Specialized firms have to advantage of detaining valuable knowledge that can result in efficient innovations through collaboration. Diversified firms might be less interesting for cooperations and hence partner with less than optimal partners.

630

5. Conclusion and discussion

The production chain characteristic of the aerospace sector results in a network in which different clusters foster different technologies. These clusters are interconnected by a small number of large firms resulting in a Scale-Free structure. The

635

Table 3

<i>Dependent variable: Return on Assets</i>			
	Network var.	Techno. var.	Combined
Adjusted clustering	0.646** (0.313)		0.623* (0.322)
Centrality	0.890* (0.513)		0.941* (0.501)
Average distance	-0.328** (0.128)		-0.335*** (0.127)
Technological diversity		0.002*** (0.0004)	0.001*** (0.0005)
Number of technologies		-0.005*** (0.001)	-0.005*** (0.001)
Number of patents		0.004 (0.004)	0.003 (0.004)
Number of cooperations		0.001 (0.004)	-0.001 (0.003)

Note: *p<0.1; **p<0.05; ***p<0.01

specificities of the aerospace sector play a vital role in the shaping of the collaboration network. These results show that a small world structure is not always a desirable structure. Indeed, in theoretical models the structure is often sought after
640 and used as an exogenous structure for diffusion analysis. The present chapter shows that in sectors such as the aerospace sector, a small world structure would make no sense and would perform much less efficiently.

We have shown that technological proximity explains collaboration between firms but that it this behavior follows an inverted U-shape, implying that there is a butter zone
645 in which firms will want to work together. The analysis of the performance of the firm tends to indicate that a central position in the network goes hand in hand with better performance for the firm. This is explained by the access to knowledge flows by firms with a high centrality and a low average distance. The choice of partner is proven to be important for two reasons, the clustering of the firm and the specializa-
650 tion of the firm. If the partner evolves in an environment in which collaborators of collaborators collaborate, this will have a positive impact on it's performance. If the firm choses a specialized firm to innovate with this will also have a positive impact on the performance of the focal firm.

References

- 655 Ahuja, G., 2000. Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative science quarterly* 45 (3), 425–455.
- Alghamdi, M., McDonald, S., Pailthorpe, B., 2012. The Emergence of a Small World in a Network of Research Joint Ventures. University of Queensland, School of Economics.
- 660 Bar, T., Leiponen, A., 2012. A measure of technological distance. *Economics Letters* 116 (3), 457–459.
- Bastian, M., Heymann, S., Jacomy, M., 2009. Gephi: An open source software for exploring and manipulating networks.
- Baum, J. A., Shipilov, A. V., Rowley, T. J., 2003. Where do small worlds come from?
665 *Industrial and Corporate change* 12 (4), 697–725.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008 (10), P10008.
- Bloom, N., Schankerman, M., Van Reenen, J., 2013. Identifying technology spillovers
670 and product market rivalry. *Econometrica* 81 (4), 1347–1393.
- Breschi, S., Lissoni, F., Malerba, F., 2003. Knowledge-relatedness in firm technological diversification. *Research Policy* 32 (1), 69–87.
- Burt, R. S., 2004. Structural holes and good ideas¹. *American journal of sociology* 110 (2), 349–399.
- 675 Chang, S.-B., 2012. Using patent analysis to establish technological position: Two different strategic approaches. *Technological Forecasting and Social Change* 79 (1), 3–15.
- Cowan, R., Jonard, N., 2003. The dynamics of collective invention. *Journal of Economic Behavior & Organization* 52 (4), 513–532.
- 680 Cowan, R., Jonard, N., 2004. Network structure and the diffusion of knowledge. *Journal of economic Dynamics and Control* 28 (8), 1557–1575.
- Cowan, R., Jonard, N., 2007. Structural holes, innovation and the distribution of ideas. *Journal of Economic Interaction and Coordination* 2 (2), 93–110.

- Dosi, G., 2000. Innovation, organization and economic dynamics: selected essays.
685 Edward Elgar Publishing.
- Egbetokun, A. A., Savin, I., 2013. Emergence of innovation networks from r&d cooperation with endogenous absorptive capacity. Tech. rep., Jena Economic Research Papers.
- Fernandez, R. M., Castilla, E. J., Moore, P., 2000. Social capital at work: Networks
690 and employment at a phone center. *American journal of sociology*, 1288–1356.
- Frigant, V., Kechidi, M., Talbot, D., 2006. Les territoires de l’aéronautique: Eads, entre mondialisation et ancrage.
- Gillespie, C. S., 2015. Fitting heavy tailed distributions: The `powerlaw` package. *Journal of Statistical Software* 64 (2), 1–16.
695 URL <http://www.jstatsoft.org/v64/i02/>
- Gulati, R., Lavie, D., Madhavan, R. R., 2011. How do networks matter? the performance effects of interorganizational networks. *Research in Organizational Behavior* 31, 207–224.
- Gulati, R., Sytch, M., Tatarynowicz, A., 2012. The rise and fall of small worlds:
700 Exploring the dynamics of social structure. *Organization Science* 23 (2), 449–471.
- Hansen, M. T., 1999. The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative science quarterly* 44 (1), 82–111.
- Jaffe, A. B., 1986. Technological opportunity and spillovers of r&d: evidence from
705 firms’ patents, profits and market value.
- Kogut, B., Zander, U., 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization science* 3 (3), 383–397.
- Lusher, D., Koskinen, J., Robins, G., 2012. Exponential random graph models for social networks: Theory, methods, and applications. Cambridge University Press.
- 710 Marco, A. C., Rausser, G. C., 2008. The role of patent rights in mergers: Consolidation in plant biotechnology. *American Journal of Agricultural Economics* 90 (1), 133–151.
- McEvily, B., Marcus, A., 2005. Embedded ties and the acquisition of competitive capabilities. *Strategic Management Journal* 26 (11), 1033–1055.

- 715 Morone, P., Taylor, R., 2004. Knowledge diffusion dynamics and network properties of face-to-face interactions. *Journal of evolutionary economics* 14 (3), 327–351.
- Mowery, D. C., Oxley, J. E., Silverman, B. S., 1998. Technological overlap and inter-firm cooperation: implications for the resource-based view of the firm. *Research policy* 27 (5), 507–523.
- 720 Newman, M. E., Girvan, M., 2004. Finding and evaluating community structure in networks. *Physical review E* 69 (2), 026113.
- Niosi, J., Zhegu, M., 2005. Aerospace clusters: local or global knowledge spillovers? *Industry & Innovation* 12 (1), 5–29.
- Orsenigo, L., Pammolli, F., Riccaboni, M., 2001. Technological change and network dynamics: lessons from the pharmaceutical industry. *Research policy* 30 (3), 485–508.
- 725 Prencipe, A., 1997. Technological competencies and product’s evolutionary dynamics a case study from the aero-engine industry. *Research policy* 25 (8), 1261–1276.
- Pyka, A., 2002. Innovation networks in economics: from the incentive-based to the knowledge-based approaches. *European Journal of Innovation Management* 5 (3), 152–163.
- 730 Pyka, A., Scharnhorst, A., 2009. Innovation networks. *Innovation Networks: New Approaches in Modelling and Analyzing, Understanding Complex Systems*, ISBN 978-3-540-92266-7. Springer-Verlag Berlin Heidelberg, 2009 1.
- 735 Salavisa, I., Sousa, C., Fontes, M., 2012. Topologies of innovation networks in knowledge-intensive sectors: Sectoral differences in the access to knowledge and complementary assets through formal and informal ties. *Technovation* 32 (6), 380–399.
- Schumpeter, J. A., 1942. *Capitalism, socialism and democracy*. Routledge.
- 740 Sorenson, O., Rivkin, J. W., Fleming, L., 2006. Complexity, networks and knowledge flow. *Research Policy* 35 (7), 994–1017.
- Teece, D. J., Rumelt, R., Dosi, G., Winter, S., 1994. Understanding corporate coherence: Theory and evidence. *Journal of Economic Behavior & Organization* 23 (1), 1–30.

- 745 Tsai, W., 2001. Knowledge transfer in intraorganizational networks: Effects of network position and absorptive capacity on business unit innovation and performance. *Academy of management journal* 44 (5), 996–1004.
- van der Pol, J., 2015a. Introduction to network modelling using ergm. Working paper.
- van der Pol, J., 2015b. Structural dynamics of the french biotech sector: a network
750 analysis. Working Paper.
- Verspagen, B., Duysters, G., 2004. The small worlds of strategic technology alliances. *Technovation* 24 (7), 563–571.
- Wasko, M. M., Faraj, S., 2005. Why should i share? examining social capital and knowledge contribution in electronic networks of practice. *MIS quarterly*, 35–57.
- 755 Wasserman, S., 1994. *Social network analysis: Methods and applications*. Vol. 8. Cambridge university press.
- Watson, J., 2007. Modeling the relationship between networking and firm performance. *Journal of Business Venturing* 22 (6), 852–874.
- Watts, September 1999a. Networks, dynamics and the small world phenomenon.
760 *American journal of sociology* 105 (2), 493–527.
- Watts, D. J., 1999b. Networks, dynamics, and the small-world phenomenon 1. *American Journal of sociology* 105 (2), 493–527.

Appendix

Computation of the normalizing constant of the power-law.

$$765 \quad \int_{-\infty}^{\infty} p(x) = 1 \Leftrightarrow \int_{+\infty}^{x_{min}} p(x) dx = 1$$

$$\Leftrightarrow c \int_{+\infty}^{x_{min}} x^{-\alpha} dx = 1$$

$$\Leftrightarrow \int_{+\infty}^{x_{min}} x^{-\alpha} dx = \frac{1}{c}$$

770

$$\Leftrightarrow \left[\frac{x^{1-\alpha}}{1-\alpha} \right]_{\infty}^{x_{min}} = \frac{1}{c}$$

$$\Leftrightarrow \frac{1}{(1-\alpha)} [x^{1-\alpha}]_{\infty}^{x_{min}} = \frac{1}{c}$$

775

$$\Leftrightarrow c(\alpha, x_{min}) = \frac{\alpha-1}{x_{min}^{\alpha-1}}$$