Paper to be presented at the DRUID Academy Conference 2016 in Bordeaux, France on January 13-15, 2016

Structural dynamics of the French aerospace collaboration network.

Johannes Van Der Pol
University of Bordeaux
GREThA UMR-CNRS 5113
johannes.van-der-pol@u-bordeaux.fr

Abstract

State-of-the-art: The analysis of collaboration networks has resulted in a better understanding of the sharing and difusion of knowledge between firms. Structural analyses have highlighted the importance of the structure of the network in the diffusion of knowledge that firms use to innovate. This literature is build upon the Schumpeterian hypothesis that new knowledge is created by the recombination of existing knowledge. Research gap: A large majority of the empirical research focuses on high technology sectors such as biotech, pharmaceuticals and software. These sectors are characterized by a high level of competition and are highly atomized. This paper extends the existing literature by studying a specific sector which has low competition on the national level and is highly organized. In addition, the scale free structure is analyzed from a dynamic perspective. An inverted U-shape relation between the probability to cooperate and technological proximity between firms has been theorized and found empirically by Hanaki 2012 using a probit analysis. This paper uses a new method more appropriate for the analysis of networks since it allows for the dependance of observation while standard econometric models do not. Finally we extend the existing literature on the link between network position and performance of the firms by identifying the impact of technological diversity and several network position characteristics. Theoretical arguments: Firms benefit from the recombination of knowledge from inside and outside sources. This should imply that firms pick their collaborators based on the knowledge ressources inside the firm. Technological proximity should hence rule network formation. Following the Schumpeterian hypothesis that knowledge is created by the recombination of existing knowledge one would expect that firms with access to a higher diversity of knowledge or a privileged position in the network would outperform firms in less favorable positions. Method: Patents deposited by French firms between 1980 and 2013 from the ORBIT database are used to generate the collaboration network. The priority dates of the patents are used for the dynamic analysis. A 5-year sliding window is used to analyze the evolution of the structure. Using Gulati's 2012 method we check for small world features, we then check for scale-free features. After computing a measure of technological proximity an Exponential Random Graph Model is used to explain the network structure and check if technological proximity can predict network formation. We then add financial information using the AMADEUS database to our network to perform a panel regression to explain the link between network position and performance. Results : Results of the network analysis show that the network is build up from clearly identifiable clusters. The overall structure is not a small world but reaches a scale-free structure. Moreover the underlaying powerlaw stabilizes around the year 2006. The network is hence an interconnection of stars, stars with a high number of links connecting with stars with a lower number of links.

These stars are central firms in the value chain, they assemble different parts of the airplane and hence accumulate knowledge from different sources in order to accomplish their assembly. The ERGM model confirms the previously identified network structure and shows that an inverted U-shape between the technological proximity and the probability for firms to collaborate exists.\\ The panel regression shows that firms with a central position a larger diversity in technologies perform better than firms less central and access to a lower diversity of technologies. Firms with a higher clustering perform better, showing that collaborators of a firm collaborate the result has a positive impact. The lower the distance between the firm and it's collaborator the better the firm performs.

Jelcodes:O31,D85

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. The networks have been studied extensively in the literature.(Geenhuizen, 2008), (van der Valk et al., 2011). 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 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 (Tsai, 2001) as well as survival and growth (Watson, 2007).

A particular focus has hence been observed on the analysis of the flow of knowledge between firms on both the empirical level (Hansen, 1999), and the theoretical level (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 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)

The structure of the network has indeed been identified as a vital variable. The structure of the network can define the efficiency with which knowledge flows throughout the network. In this light the small world structure has been identified as being the most efficient structure (Verspagen and Duysters, 2004). This observation is however the result of mostly theoretical work, (Baum et al., 2003) even though the small world structure has been observed empirically (Ahuja, 2000).

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 collaborations are formed and how they impact performance we chose to analyze the collaboration network of the aerospace sector in France between 1980 and 2014. The aim of this paper is to provide a macro analysis of the evolution of the structure of an innovation network. We will identify the structure of the network and provided a detailed explanation as to why this structure is observed. This will be

accomplished by both an Exponential Random Graph Model and community detection. We will then switch to the micro level and offer an analysis of the link between network position and firm performance. This paper extends the existing literature on the question of the evolution of collaboration networks and how firms can benefit from the knowledge flows inside the network.
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

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 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 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 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 here was provided by the VIA-INNO platform[1] 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

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

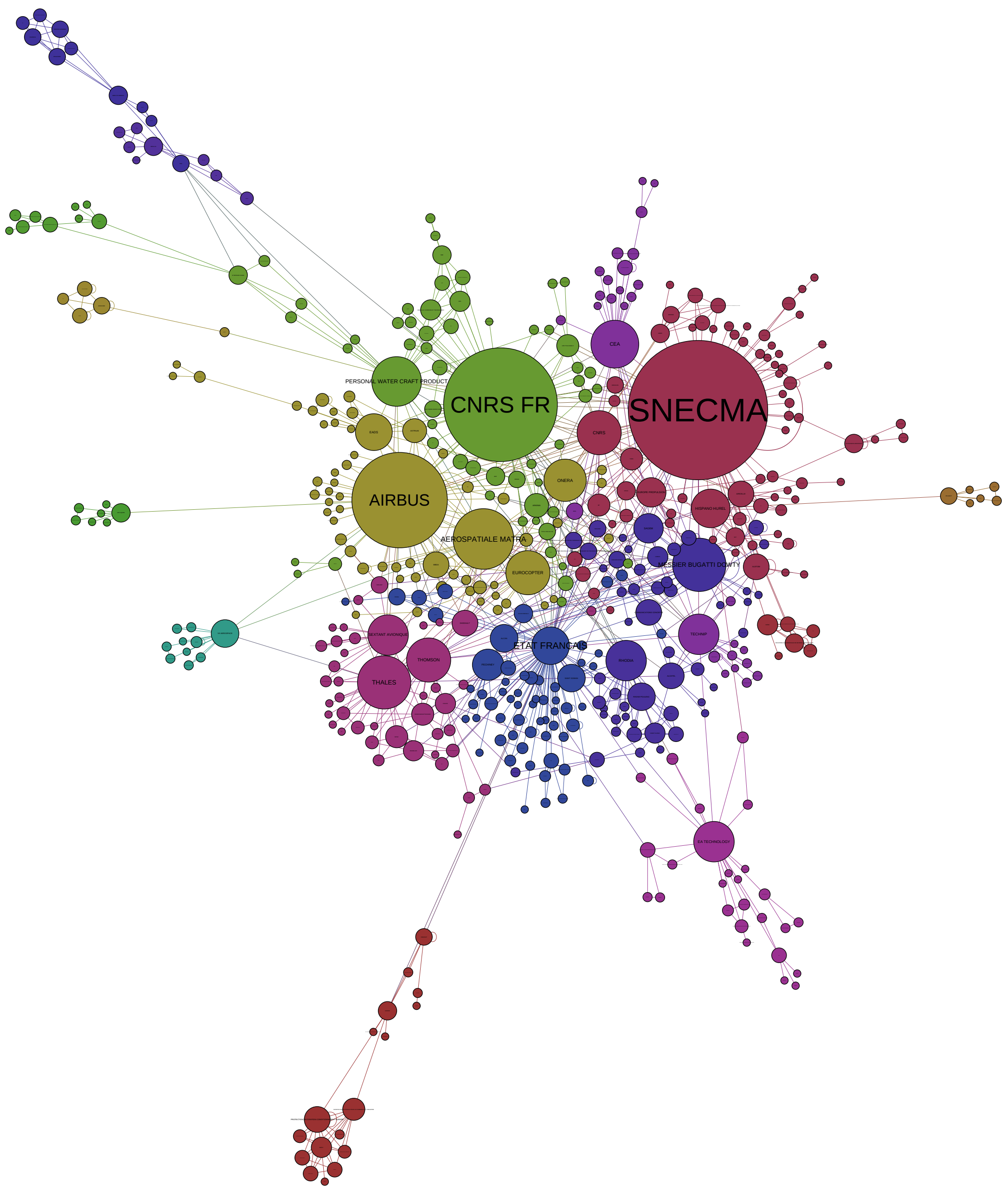
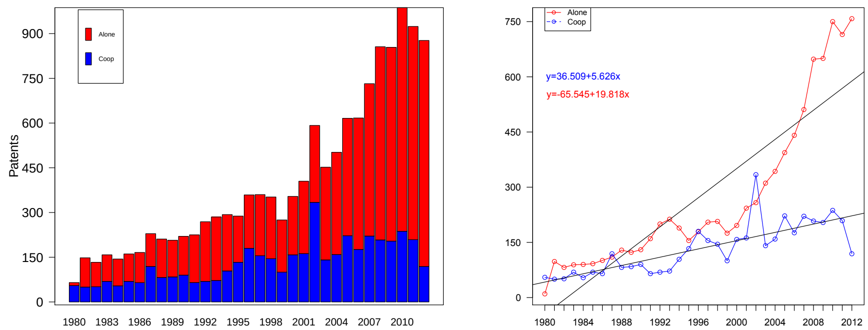


Figure 1: The aerospace collaboration network as of 2014. Node size is proportional to the number of collaborations, colors correspond to structural clusters identified by a maximization of modularity.



(a) Dynamic of deposits (b) Trend in deposits

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)

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.

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 (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.
We observe an important increase in the number of patents from the year 2000 onwards. This can be explained partially be the commercialization of the Airbus A380. A particular aspect of the aerospace sector is the fact that there are mass deposits after the commercial release of an airplane which might explain some of the variance in the dataset.
The trends clearly show in increase in the number of collaborations over time. We will now turn out attention to the analysis of the network resulting from these collaborations.

2. The French aerospace network

The network will be analyzed in two stages, first we will focus on different communities that make up the network. This will allow us to have a first insight into the construction of the network. Second, we will focus on the global network structure.

2.1. Cluster identification

The previously identified dataset leave us with over 4000 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 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 0.7
An algorithm introduced by (Blondel et al., 2008) was used to identify these communities 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 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 selects a node a 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. 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
- Thompson: Seating
- Messier Bugatti: Landing and braking.
- Pechiney Rhenalu: Structural elements (aluminium)

- Alcatel Lucent: Avionics and communication systems

These clusters suggest local technological development according to different parts included in the production of an aircraft. The large assembler firms (Airbus, Snecma and Thales) have a central position interconnecting the clusters. 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 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 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.
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.
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.

We now have an idea of the manner in which the network is structured. We will now turn our attention to the global network structure. The global network structure will allow us to compare this network with other networks. Also, since theoretical models state that certain structures are more or less efficient in terms of knowledge diffusion, the global network structure will give insight in the potential for diffusion in our network.

2.2. Network structure identification

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.

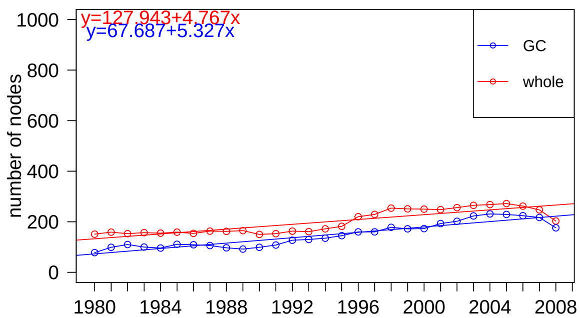


Figure 3: Evolution of the number of nodes with a 5-year sliding window (i.e 1980 → 1980-1984). "GC" is the giant component of the network, "whole" the giant component with all the smaller components

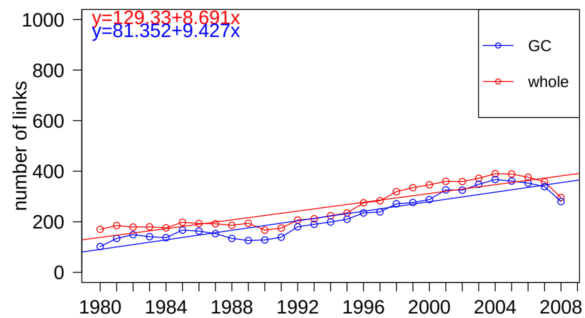


Figure 4: Evolution of the number of links with a 5-year sliding window (i.e 1980 → 1980-1984). "GC" is the giant component of the network, "whole" the giant component with all the smaller components

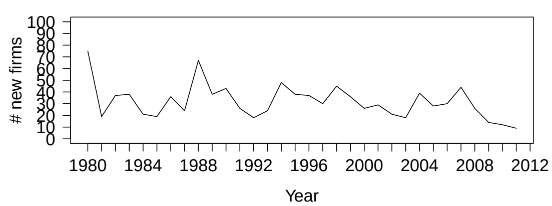


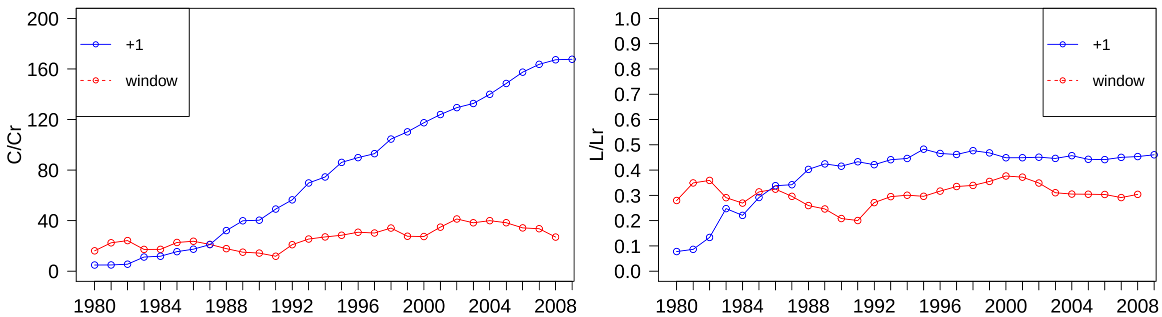
Figure 5: Evolution of the number of new firms entering the collaboration network each year.

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. 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 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. 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.

2.2.1. 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 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 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 Cr (Lr) 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 $\frac{C}{Cr} >> 1$ and $\frac{L}{Lr} \approx 1$.
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.



(a) Clustering

(b) Average distance

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.

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 $\frac{L}{Lr} < 1$. 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 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 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 observa-

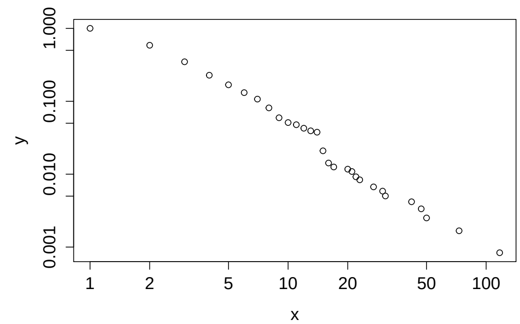


Figure 7: The cumulative frequency distribution of the aerospace network in 2014

tions 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 different network structure.

We can conclude here that there is a high tendency for firms to cluster which confirms our previous observation that firms where organized in interconnected clusters. The structure also appears to stay relatively stable when it comes to these 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.

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.

2.2.2. The scale free structure

Scale free networks represent a particular structure that has been observed empirically at different levels world wide web, lexicography, collaboration networks. A scale free structure is a network with a particular degree distribution.

In order to check for scale-free structure we try to fit a log-linear and a power law to the data. 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 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 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 a SFN. 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 the network. This proves the importance of core actors inside the network. Since different technologies are developed in different clusters we observe that there are some highly connected firms specialized in a particular technology (as can be seen in Figure 1.). These firms assemble the technologies developed with the downstream firms and hence have, and will always have, many cooperations. Economic network theory would suggest that firms with many cooperations have build up a reputation that is beneficial for any firm working with them. These large firms can be used as a reference for smaller firms that will benefit from the reputation of working with larger firms. In addition, larger firms have the ability to sustain a larger number of collaborations since they have more financial and productive resources.

The structure stabelizes over time. The parameter α which defines the function stabilizes through the end of the period as can be seen in figure 8. The constant "c" changes only with x_{min}, see Appendix for computation. Since this value also stabilizes we can conclude that the structure of the network stabilizes. We now have an idea of the overall structure of the network and why it has this structure. For the limited number of large firms there is a large number of small firms.

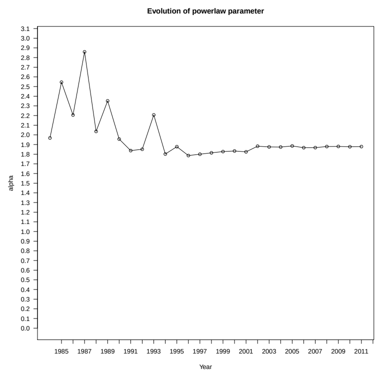
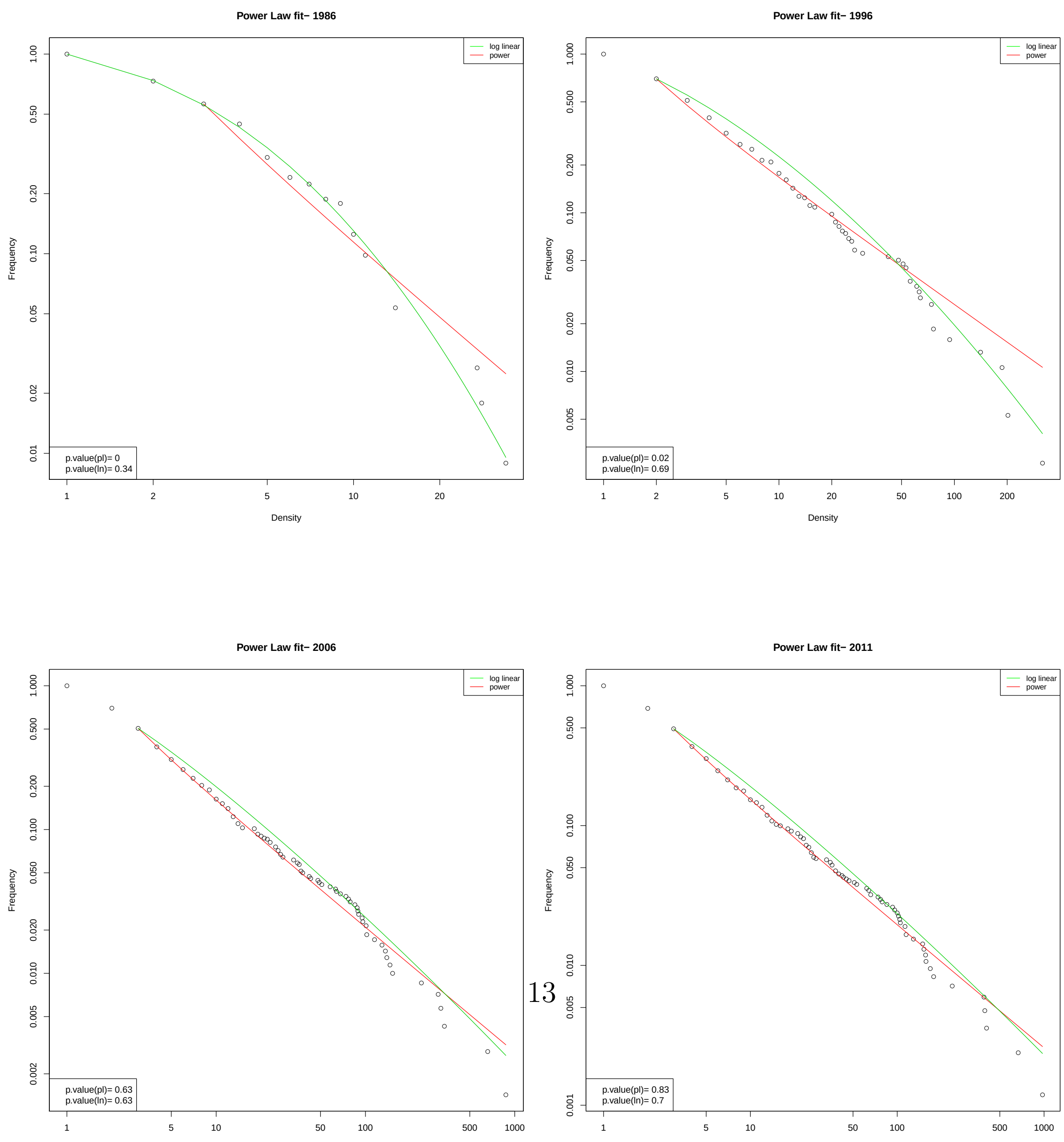


Figure 8: Scale-free structure stabilization



3. Technological proximity as a preferential attachment mechanism

We now know that the network is a scale-free network. In order to understand how this structure came to be we hypothesize that technological proximity can explain why certain firms interconnect while others do not.
Firms need to cooperate to make their products compatible, especially in a value chain like the aerospace sector. The integration of composite materials for example asks for firms to collaborate closely so that all parts are compatible.
In order to explain from a micro perspective the structure of the network we provide an Exponential Random Graph Model which will take into account two types of covariates: structural and technological.

3.1. Technological proximity

Many measures of technological proximity exist, some are based on patent citations (Chang, 2012), (Marco and Rausser, 2008), (Mowery et al., 1998) while others use IPC codes (Jaffe, 1986). 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, 1986) further developed by (Breschi et al., 2003). Finer measures exist, see for instance (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 cooperation based on the technologies mastered by firms. Therefor 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.

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, 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 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 correspond 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: Proximity Measure

We defend the position that knowledge about one specific technology is enough to initiate a collaboration. he 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 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. We then used an ERGM model to check if firms with a high technological proximity have a tendency to work together rather than with firms with lower proximity.

3.2. Motivation for cooperation

In order to verify the latter we have a look at the IPC codes on the patents and try to 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 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 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 firms to mobilize specific technologies and hence that cooperation are technologically oriented. We can hence start our analysis of technological proximity as a motivation for cooperation.

3.3. The ERGM

An Exponential Random Graph Model models the global structure of a network while allowing inference on 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. The model to estimate is given in equation 1.

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

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. For a more complete explanation of ERGM models see (**?**), (van der Pol, 2015a). When checking for a goodness of fit we will generate a graph with boxplots showing if the observed network remains between the confidence intervals as shown in figures 11 and 12.

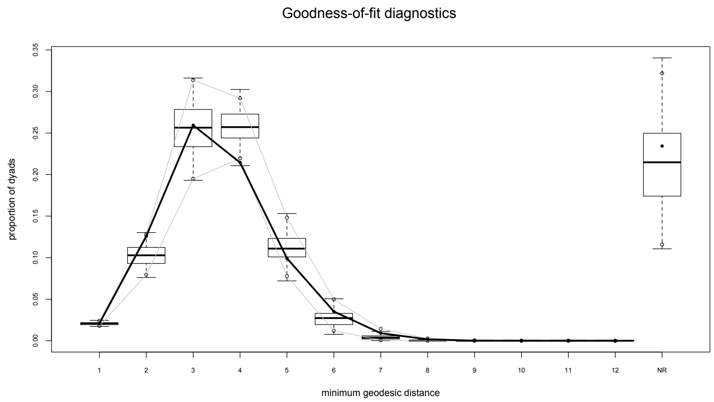


Figure 11: Goodness of fit ERGM model 1

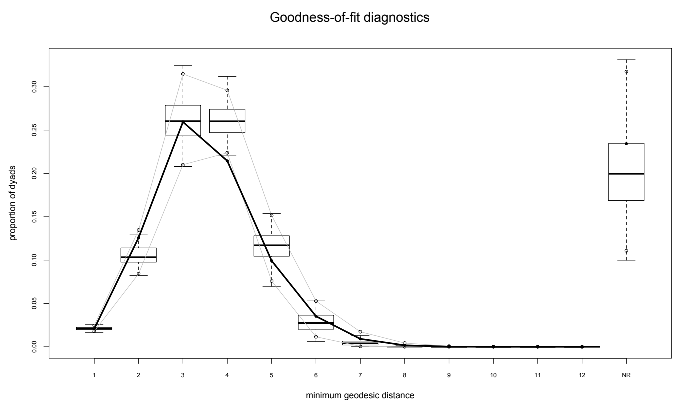


Figure 12: Goodness of fit ERGM model 2

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 1 shows the regression results. 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 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 technologies can consider that the other firm has nothing to offer them and hence prefer collaborating with a firm that has different technologies.
We have a large number of firms with a small number of links and a small number of firms with a large number of links. The degree variables of the model suggest that the firms with a small number of links tend to cooperate with each other, while the firms with a higher number of links collaborate between them. This is compliant with a production chain structure where the larger firms assemble a large diversity of products into a multi-technology product. The goodness of fit diagnostics (figures 10 & 11) show that the first and the last model are good fits for the observed network. This means that the variables we included in the model result, on average, in the observed network. The two grey lines are the confidence interval and a perfect fit would have the black line (the empirical network) on the center of each boxplot. In addition, the lower the AIC and BIC criteria the better the fit. The addition of proximity increases the fit of the model. We conclude that our model is a good fit.

We have studied the overall structure of the network and concluded that the 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 found that firms that are technologically close will have a high probability to work together. Now that we better understand the strategies of the firms we want to know how the network they created will influence their performance.

4. Impact of network position of the firm on performance

The objective of the 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 firms has access.

	Model 1	Model 2	Model 3	Model 4
Convergence	8	8	9	27
edges	-5.29^{***}	-5.20^{***}	-4.92^{***}	-4.61^{***}
	(0.19)	(0.13)	(0.14)	(0.13)
edgecov.proximity	0.47^{***}			
	(0.04)			
edgecov.citation	0.56^{***}	0.57^{***}	0.56^{***}	0.53^{***}
	(0.05)	(0.05)	(0.05)	(0.05)
degree4	-0.91^{**}		-0.83^{**}	-1.59^{***}
	(0.30)		(0.29)	(0.32)
degree5	-0.76^{*}		-0.66^{*}	-1.45^{***}
	(0.33)		(0.33)	(0.34)
degree6	-1.27^{**}		-1.15^{*}	-1.94^{***}
	(0.48)		(0.48)	(0.49)
degree7	-1.38^{*}		-1.26^{*}	-1.97^{**}
	(0.62)		(0.61)	(0.62)
degree8	-0.13		-0.03	-0.66
	(0.43)		(0.43)	(0.43)
edgecov.proximity2		0.09^{***}	0.08^{***}	0.07^{***}
		(0.01)	(0.01)	(0.01)
degree2				-0.87^{***}
				(0.22)
degree3				-1.32^{***}
				(0.26)
AIC	2661.67	2643.84	2636.20	2608.15
BIC	2722.81	2666.76	2697.34	2684.58
Log Likelihood	-1322.83	-1318.92	-1310.10	-1294.08

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1: ERGM model results

We will conduct a panel data analysis to estimate the influence of the position of the firm on it's 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:

$$ROA_t = \frac{Net\ Income_t}{Total\ Assets_t} \tag{2}$$

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 versions 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*. A count of the number of IPC counts in the direct neighborhood of the firm.
- *Clustering of the focal firm*. The ratio of the number of triangles for the focal firm and the number of possible triangles.
- *Centrality and average distance of the focal firm*. Average distance between the firm and the all other nodes in the network. Centrality is computed using a betweenness centrality measure.
- *Number of patents of the firm*
- *Number of technologies for the focal firm* The number of IPC codes the firm has deposited a patent in on it's own.

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 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 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.

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 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 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 collaborators. 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.

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.
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.

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.
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.

5. conclusion

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. 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 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 specialization 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.

Table 2

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

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

References

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.

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? 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 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.

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.

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. 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 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.

Geenhuizen, M. V., 2008. Knowledge networks of young innovators in the urban economy: biotechnology as a case study. Entrepreneurship and Regional Development 20 (2), 161–183.

Gillespie, C. S., 2015. Fitting heavy tailed distributions: The poweRlaw package. Journal of Statistical Software 64 (2), 1–16.
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: 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 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.

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.

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.

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.

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.

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.

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.

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.

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 analysis. Working Paper.

van der Valk, T., Chappin, M. M., Gijsbers, G. W., 2011. Evaluating innovation networks in emerging technologies. Technological Forecasting and Social Change 78 (1), 25–39.

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.

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. 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.

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

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

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

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

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

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