Groupe de Recherche en
Économie Théorique et Appliquée

Explaining the structure of collaboration networks: from firm-level strategies to global network structure

Johannes VAN DER POL

VIA Inno, GREThA, UMR-CNRS 5113, Université de Bordeaux

johannes.van-der-pol@u-bordeaux.fr

Cahiers du GREThA
n° 2018-02
February

GRETHA UMR CNRS 5113
Université de Bordeaux
Avenue Léon Duguit - 33608 PESSAC - FRANCE
Tel : +33 (0)5.56.84.25.75 - Fax : +33 (0)5.56.84.86.47 - www.gretha.fr

Expliquer la structure des réseaux de collaboration: de la stratégie de selection de partenaires à la structure globale du réseau

Résumé

L'objectif de ce papier est de montrer comment les stratégies de sélection de collaborateurs influencent la structuration des réseaux de collaboration. L'analyse se construit en trois étapes. Une première étape identifie les stratégies de sélection de partenaires des entreprises, une seconde comment ces stratégies résultent en la création de clusters. Une dernière étape consiste en l'analyse de la structure globale du réseau qui est le résultat de l'interconnexion de ces clusters.

Dans le but de mettre en avant l'influence du secteur d'activité sur la structuration du réseau de collaboration, l'analyse porte sur le secteur aéronautique en Franc et le secteur des Biotechnologies en France.

Les résultats montrent que les stratégies des firmes sont les mêmes dans les deux secteurs alors que les structures globales des réseaux diffèrent fortement. Le réseau aéronautique est une structure core-peripherie avec des caractéristiques petit monde, alors que le réseau des biotechnologies ne présente pas de caractéristiques particulières. In fine, la différence entre les deux structures provient des caractéristiques des secteurs.

Mots-clés: SNA ; Analyse sectorielle ; Réseau de collaboration ; Biotechnologie ; Aéronautique ; ERGM ; Innovation

Explaining the structure of collaboration networks: from firm-level strategies to global network structure

Abstract

The aim of this paper is to show how firm-level partner selection strategies impact the structure a of collaboration network. The analysis is performed in three stages. A first stage identifies how partners select their collaborators, a second stage shows how these decisions result in clusters, and a final stage studies the global network structure that emerges from the interconnection of these clusters. In order to highlight the importance of the sectors' influence, the analysis is performed on the French Aerospace and the French Biotech collaboration networks. Results show that the firm-level strategies are the same in both sectors while the resulting global network structure is different (core-periphery structure with small-world characteristics for the aerospace network and no particular structure for the biotech sector). The difference in the global network structure can be explained by sectorial characteristics. These differences define the manner in which knowledge flows through the network.

Keywords: SNA; Sectoral analysis; Collaboration network; Biotechnology; Aerospace; ERGM; Innovation

JEL: L25 ; C23 ; D85 ; L14 ; C20

Reference to this paper: ***VAN DER POL Johannes,*** (2018) Explaining the structure of collaboration networks: from firm-level strategies to global network structure, *Cahiers du GREThA*, n°2018-02.

http://ideas.repec.org/p/grt/wpegrt/2018-02.html.

GRETHA UMR CNRS 5113
Université de Bordeaux
Avenue Léon Duguit - 33608 PESSAC - FRANCE
Tel : +33 (0)5.56.84.25.75 - Fax : +33 (0)5.56.84.86.47 - www.gretha.fr

1 Introduction

The recent spur of recombination of technologies by firms collaborating with one another is contributing to technologies becoming increasingly sophisticated and complex. Creating new technologies therefore requires knowledge from a multitude of technological domains. This makes it difficult, if not inefficient, for firms to be expected to keep up with the degree of diversity of knowledge by themselves. Consequently, firms turn to external sources of knowledge to complement their own. R&D collaborations are one way for firms to access external knowledge. Even though collaborations have a high risk of failing (Mowla (2012); Barringer and Harrison (2000)), the number of collaborations between innovating firms has been steadily increasing since the 1980's (Gilbert and Ahrweiler (2001)). Hagedoorn (2002) shows that collaborations related to innovation have also increased steadily. Moreover he shows that there are sectorial differences when it comes to these collaborations. The increase in the number of collaborations has been observed to be significantly more important in high technology sectors than low or medium technology sectors (Hagedoorn and Narula (1996)).

A hypothesised consequence of collaboration is the opportunity for firms to exchange knowledge during their collaboration (Rogers, 1995; Mowery et al., 1996; Powell et al., 1996; Powell and Grodal, 2005; Pyka, 1999; Singh, 2005). It is unrealistic to assume that transfers of knowledge resulting from a collaboration are synonymous with perfect absorption of all the knowledge held by either firms to the point that it can replace the activities of the collaborator.

However, firms can exchange information, knowledge and routines that might improve the industrial or innovative performance of the collaborating firms. For exemple, learning of a better software from a collaborator may improve productivity in related activities.

There appears to be two main advantages to collaboration; the value of recombining technologies, and the flow of knowledge that might occur during these collaborations. As such, collaboration has proven to be beneficial for the firm (McEvily and Marcus (2005)), its innovation (Kogut and Zander (2003); Tsai (2001)) as well as its survival and growth (Watson (2007)).

When taken together, collaborations in a given sector can be represented by a network. The hypothesis of knowledge flow implies that the knowledge exchanged during a collaboration can potentially reach all firms that are interconnected. This observation has drawn the attention of many researchers who analysed the structure of networks in order to find structures that would be efficient for this transfer (Verspagen and Duysters (2004)). Different structures have been identified to exist empirically: core-periphery (Newman (2011)), small-worlds (Algamdi et al. (2012); Gulati et al. (2012)) and nested split graphs (König et al. (2009)). The small-world structure has received considerable attention since it was argued to be the most efficient structure for knowledge transfers (Verspagen and Duysters (2004)) and identified in many empirical settings.

In order to fully understand how these networks emerge, how their structure evolves and how this impacts the flow of knowledge, different aspects of collaboration should be studied. First, the identification of partner selection strategies must be identified. This first step explains how the choices made by firms result in the creation of different communities inside a network. For this purpose an Exponential Random Graph Model (ERGM) is used. An ERGM model is a modified logistic regression specified for the econometric analysis of network data. Second, one need to explain these communities and analyse how they interconnect to form a global network structure. Finally the global network structure must be analysed in order to understand fully how knowledge is created and diffuses in the network. This is the logic we will follow in this paper and we will apply it to the French aerospace and biotech sectors. These sectors have been chosen because they have similar characteristics from an innovation point of view (long innovation cycles and high technology). However they are different in their organisation, the aerospace sector is a highly organised production chain while the biotech sector is driven by fierce competition between large firms while smaller firms tend to be suppliers.

The first section of this paper will focus on the identification of partner selection strategies. The second section of the paper will study how these strategies result in the creation of communities (or clusters) inside the networks. Finally the last part of the paper will study how these communities interconnect and create a global network structure. This structure of each of the networks will be checked for both small-world and core-periphery properties and a comparison will be made between the two networks.

2 Data

The networks analysed in this paper were extracted from patent applications. Patent databases contain information on their original assignees; whenever there are two or more original patent assignees on the document I conclude that the patent is the result of a collaboration. Even though patents provide a useful source of information on collaborations they are not exhaustive in the identification of the collaborative behaviour of firms. Not all collaborations result in patent deposits for various reasons.

Since patenting behaviour differs from country to country and rules change from one patent office to another, I chose to consider solely patents deposited by French companies in France. For the biotech sector a financial database is used to extract firms that declare “research in biotechnology” as their main activity (NACE 7211). The patent portfolio of each of the identified firms is extracted from a patent database. This results in a dataset of 2061 patents, deposited in France by companies located in France. The dataset is restricted to all identified patents deposited between 1980 and 2014. The starting year was chosen because the number of collaborations before 1980 was too low to create a significant network.

The identification of firms working in the aerospace sector is slightly more complicated. Many of the firms active in the sector have their main activities in different domain than aeronautics (i.e metal works, composite materials or tyres). Therefore a query is build based on a combination of keywords and International Patent Classification codes (IPC) to extract patents related to airplanes. These IPC codes are added to the patent application by the patent examiners. There are over 70.000 different codes used to classify patent, for example, B64C is the codes used on all patents related to airplanes, helicopters and cosmonautics, B64C001 is specific to the fuselage of airplanes, B64C 25/36 is a code for tyres for airplanes. The patents that resulted from the query were analysed and the original patent assignees were extracted to create the collaboration network.

These codes are also used to compute an indicator of technological proximity following Jaffe (1986) and Breschi and Lissoni (2001). The latter gives a correlation coefficient which takes the value of 0 when two firms deposited no patent using the same classification and 1 if they only deposit patents using the same codes. Using this method I identified 11992 patents deposited in France by companies located in France for the aerospace sector.

3 Strategic partner selection and the emergence of clusters

The structure of a collaboration network is the result of individual partner selection strategies. The aim of this first section is to identify which partner selection strategies can explain the structure of the observed collaboration network in each of the sectors. In this paper I test three strategies : Triadic closure, technological proximity and the rich get richer principle. These strategies will be explained in the following subsection.

3.1 Partner selection strategies

Given that collaborations fail often, any bit of information about the reputation of a potential collaborator is valuable. Current collaborators can provide this information about their own collaborators. A firm recommended by a collaborator can be considered as a less risky choice when compared to an unknown collaborator.

In addition, given that they have a common collaborator, the chance that they speak a common language in their work ethics is also higher. Therefore, for two firms the probability of collaboration is higher if they have a common collaborator. The latter is referred to as triadic closure.

Technological proximity measures if firms work on similar technologies. Proximity is high when firms work on the same technologies; low when they work on different technologies. The relation between the probability to collaborate and technological proximity is no expected to be linear. Firms that are close will be competitors on the market; their probability will be low. As the proximity decreases, the probability should increase until proximity becomes too low. In short, I expect to find an inverted U-shape relation between the probability to collaborate and technological proximity. In this model the technological proximity was computed following Jaffe (1986) andBreschi and Lissoni (2001). Since we aim at measuring the proximity between two firms, we only used patents that were not deposited through collaboration, this ensures that the firms master the technology themselves.

The rich get richer principle is the idea that the probability of firms collaborating increases with the number of existing collaborations. This is also referred to as preferential attachment (Barabási and Albert (1999)). This can be due to the fact that some firms are larger and hence are able to sustain more collaborations, or this can also be the result of a reputational effect.

3.2 An Exponential Random Graph Model

In order to verify that these strategies can explain the observed network structure an econometric model is developed. Previous work has focused on the identification of partner selection strategies, notably **?**, in this paper the probability of collaboration based on different strategies is computed. However, the model does not account for the structure of the network. In other words, they considered all collaborations to be independent. In a network setting this may however not be the case. The reason A collaborates with B is often dependent upon collaboration between A and C. The observations are therefore not independent. An ERGM model takes these effects into account; it models the global structure of a network while allowing inference on the likelihood of a link between two nodes. It thus computes the odds of a collaboration given the current structure of the network.

These models are rather recent but are stating to gain momentum in economic literature (Brennecke and Rank (2017); Ter Wal (2013); Lomi and Pallotti (2012); Caimo and Friel (2011)). An ERGM is basically a modified logistic regression. The general form of an ERGM model is given by:

$$P(X = x|\theta) = \frac{1}{k(\theta)} \cdot exp(\theta_1 z_1(x) + \cdots + \theta_n z_n(x)\ )$$

where X is the simulated network, x the observed network and θ a vector of parameters, z_i is a difference variable and $k(\theta)$ the normalisation constant.

The model starts with an empty network and creates links between the nodes based on the values of the variables we have found from the empirical network. Using a Monte Carlo Markov Chain method the model simulates a network structure, changing the parameters of the model until the generated network is significantly identical to an observed network (on average). The model is estimated with a Markov Chain Monte Carlo procedure that identifies the parameters that maximise the likelihood of a graph (i.e the likelihood that the graph generated by the model is significantly identical to the observed graph).

ERGM models suffer from degeneracy issues which in many cases results in problems with finding the maximum likelihood. In order to solve this issue, a slightly modified version of the general model is used in this paper: the curved exponential random graph model. The modification consists of adding weights

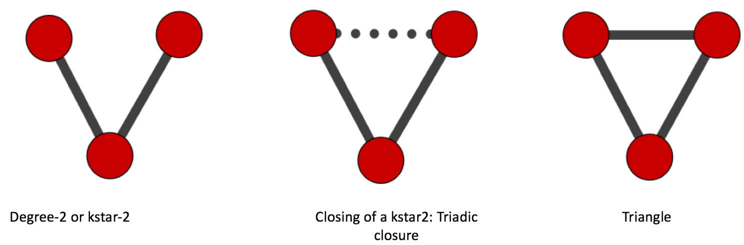

Table 1: Examples of the 2-kstar, triadic closure and triangle network substructures

on the degrees of the nodes (Lusher et al. (2012)). These weights keep the higher degree nodes from having too much influence resulting in degeneracy. Two main methods exist when it comes to solving degeneracy. The gwesp and gwesp.alpha parameters correspond to the first method, the altkstar parameter to the second method. In the end the model with the lowest Akaike criterion will be retained as the best model.

Results The *edges* variable plays the same role as a constant in a linear regression. A *kstar2* is a star with two links, in other words A connected to B and C, this variable is identical to degree2. The variable *triangle* identifies if triangles have an important role to play in the structuring of the network, and we do find them to be significant. More interestingly, when both the *degree2* (or *kstar2*) and the triangle variable have a significant effect we can conclude that in the model triangles have tendencies to be closed (Lusher (2011); Harris (2013); Ter Wal (2013)). This proves the first hypothesis that triadic closure has a significant effect in the structuring of the collaboration network. In both sectors, firms with a common collaborator have a higher probability of collaborating than firms without a common collaborator. We can explain this observation by the idea that firms will rely on information of their collaborators to pick new collaborators. In addition, there can be productivity gains associated with this behaviour since it is likely that the three firms will have fairly similar methods. From a network structure perspective, this behaviour results in the creation of triangles in the network. These triangles appear to be interconnected closely between firms with a certain level of technological proximity. The variable*edgecov.proximity2* represents the square of the technological proximity that has a significant impact on the structuring of the network. Therefore, there appears to be an inverted U-shape relation between the probability of link creation (i.e collaboration) and the technological proximity between firms. Firms too far from each other have low odds of collaboration as do firms that are too close. The firms that do interconnect have an intermediate level of proximity. This also appears to confirm the existence of a competition effect, which reduced incentives to collaborate between technologically close firms. Firms that are further away have less competition and hence appear to collaborate more.

	Dependent variable:		
	Network		
	(1)	(2)	(3)
edges	−7.267***	−1.121*	
	(0.228)	(0.626)	
kstar2	0.155***		
	(0.003)		
degree2		−1.336***	14.467***
		(0.255)	(2.970)
edgecov.citation			−20.934***
			(1.090)
triangle	3.428***	1.923***	1.726***
	(0.007)	(0.0001)	(0.0002)
gwesp	−0.439***		
	(0.166)		
gwesp.alpha	0.523		
	(0.385)		
edgecov.proximity2		1.565***	6.620***
		(0.271)	(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:			*p<0.1; **p<0.05; ***p<0.01

	Dependent variable:			
	Network			
	(1)	(2)	(3)	(4)
edges		−5.431***	−7.685***	−7.429***
		(0.124)	(0.298)	(0.006)
triangle			2.009***	1.736***
			(0.002)	(0.00001)
degree2	10.940***	0.426*	8.223***	1.528***
	(0.365)	(0.225)	(1.771)	(0.148)
degree3	11.820***	−0.006	3.176***	2.230***
	(0.462)	(0.215)	(1.072)	(0.134)
degree4	8.939***	−0.926***	8.122***	2.535***
	(0.781)	(0.226)	(1.790)	(0.226)
degree5	2.919***	−1.600***	4.038**	2.706***
	(0.942)	(0.258)	(1.568)	(0.322)
degree6	−0.332	−1.520***	6.770***	3.125***
	(0.778)	(0.262)	(1.411)	(0.311)
degree7	−4.771***	−2.521***	1.899	1.857***
	(1.533)	(0.406)	(1.455)	(0.611)
degree8	−1.030	−0.972***	4.026**	2.444***
	(0.916)	(0.221)	(1.754)	(0.318)
gwdegree			9.189***	
			(2.205)	
gwdegree.decay			−0.593***	
			(0.060)	
edgecov.proximity2	−2.629***	1.625***	1.959***	1.746***
	(0.006)	(0.132)	(0.350)	(0.008)
Akaike Inf. Crit.	11.909	8.370	7.337	7.161
Bayesian Inf. Crit.	11.981	8.451	7.445	7.251
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 1: Regression results for the aerospace sector (table on the left) and the Biotech sector (table on the right)

The triangles that are formed result in the creation of technological clusters. These clusters contain densely interconnected firms working on similar technologies. The *alkstar* variables that serve the purpose of reducing degeneracy also play a second role. They also hint to the fact that there are nodes with many links and nodes with few links. This is modelled more explicitly in the biotech sector by the enumeration of the different degree levels (*degree2*, *degree3* etc.). Interestingly, the strategies for collaboration appear to be the same in both sectors.

The fact that the degree variables are significant implies that these nodes have a structuring role to play in the network. This concurs with observations made by Cassiman and Veugelers (2002); Colombo (1995), which have noted that firms active in collaboration are mainly large firms, able to sustain many collaborations. The results of the ERGM model show that the structure of the network is built around these large firms.

It is possible that firms with a high degree are positioned inside the clusters themselves or that they play a role of gatekeeper, i.e they interconnect different communities. This question is the subject of the next section.

4 The interconnection of communities

In order to get a first idea of the interconnection of the communities in each network, we can proceed first by a visual inspection. Figure 4 provides the network of the biotech sector; figure 4 provides the collaboration network of the aerospace sector. Each node in the network is a firm and a link represents at least one co-deposited patent (i.e a patent with the name of both firms as original assignees). The colours of the nodes and edges represent the identified communities. These communities were identified using a modularity maximisation algorithm created by Blondel et al. (2008). The algorithm creates a random network with the same number of nodes and edges as the empirical network. It assigns random links between the different nodes of the network. It then compares the number of links each node has in the empirical network to the number in the random network. If there is a significant difference, the algorithm will consider that the connections are not random and hence that there is an empirically valid community. The algorithm does not require the user to provide a number of communities; the number is selected automatically as the number that maximises the modularity function. In short, modularity aims at maximising the number of links inside a community and minimise the number of links between communities. Using this method we can identify different communities that have been coloured for clarity.

Of course there is no guarantee that the identified communities make sense. Therefore some verification is required. This is accomplished by hand with the aim of finding a common feature between the firms (market tier, R&D focus). As we have shown in the ERGM model, firms regroup according to technological proximity (or competitive pressure). The identified communities reflect this, since the different communities in the biotech sector reflect the different market segments of the biotech industry (red biotech, blue biotech, yellow biotech and so on). In the case of the aerospace sector a similar conclusion is found; each community is built around firms that work on specific part of the airplane.

In the case of the aerospace sector, firms with a low number of links connect to a first order supplier. The latter has therefor a high degree. These first order suppliers are connected to the most connected firm: Airbus. This means that the structuring role of the density identified by the ERGM model can be explained by the presence of many second order suppliers with a low number of links, connected to a lower number of first order suppliers with high density which are in turn connected to the central firm with the highest number of links: Airbus.

To some extent, the same appears to apply for the biotech sector. Each community appears to have a central firm around which other firms interconnect. This explains the similar observation that the "rich get richer" principle applies to both sectors. Larger firms have more collaboration and hence the probability that they will collaborate in the future is higher.

The interconnection of the different communities is achieved by the first-order suppliers in the case of the aerospace sector. The results appear similar in the case of the biotech sector even though there are more firms playing the role of gatekeeper between communities.

In short, firms collaborate with firms that represent a relative low market threat and use common collaborators to find new partners. The latter results in the creation of triangles and to a larger extent, in the creation of clusters. The rich get richer principle identified in the ERGM model can be explained by the presence of larger firms inside the cluster which attracts a large number of firms with a small number of connection. The latter observation would lead to believe that both networks could have a similar global structure that could resemble a scale-free structure. A scale-free network structure is defined as a network structure in which there are a large number of nodes with a low degree and small number of nodes with a high degree. In addition, since there is a central firm interconnecting all clusters in the aerospace sector, we could expect that this network is characterised by a high level of clustering (i.e many triangles in the structure) and a low average distance between the nodes (the central firm connecting the communities reduces the distance one has to travel from one node in the network to another). This type of structure is known as a small world structure. This particular type of structure has implications for the efficient diffusion of knowledge in a network, because of this we will check for small world characteristics in both networks.

5 Collaborative behaviour at the sectoral level

5.1 A core of intensely collaborating firms: Scale-free networks

Some networks are defined by a densely interconnected core and a more or less sparsely connected periphery as shown in Figure 5.1. This type of structure has been identified in citation networks, the internet and lexicographical networks amongst others Estrada (2012). This particular structure results in a core of a few densely connected networks and a periphery of many sparsely connected nodes. Having this type of structure makes for a particular degree distribution when compared to network with a more homogenous distribution. We therefore require a method to identify a core-periphery structure statistically. Core-periphery networks are identified by their degree distribution, we therefore start by plotting the cumulative degree distribution of a network. Figure 5.1 gives an example of a degree distribution. This distribution gives the degree on the x-axis and the number of nodes with that degree on the y-axis. From this distribution we can see that the number of nodes with a high degree is low. In addition, the number of nodes with only a few links is high. This information alone is not sufficient to conclude that the network has a core-periphery structure Newman (2011). In order to get more precise information out of this data we are going to transform the degree distribution into a cumulative frequency distribution (Figure 5.1).

The Cumulative Frequency Distribution (CFD) transforms the degree distribution into a probability distribution. From this distribution we can read the

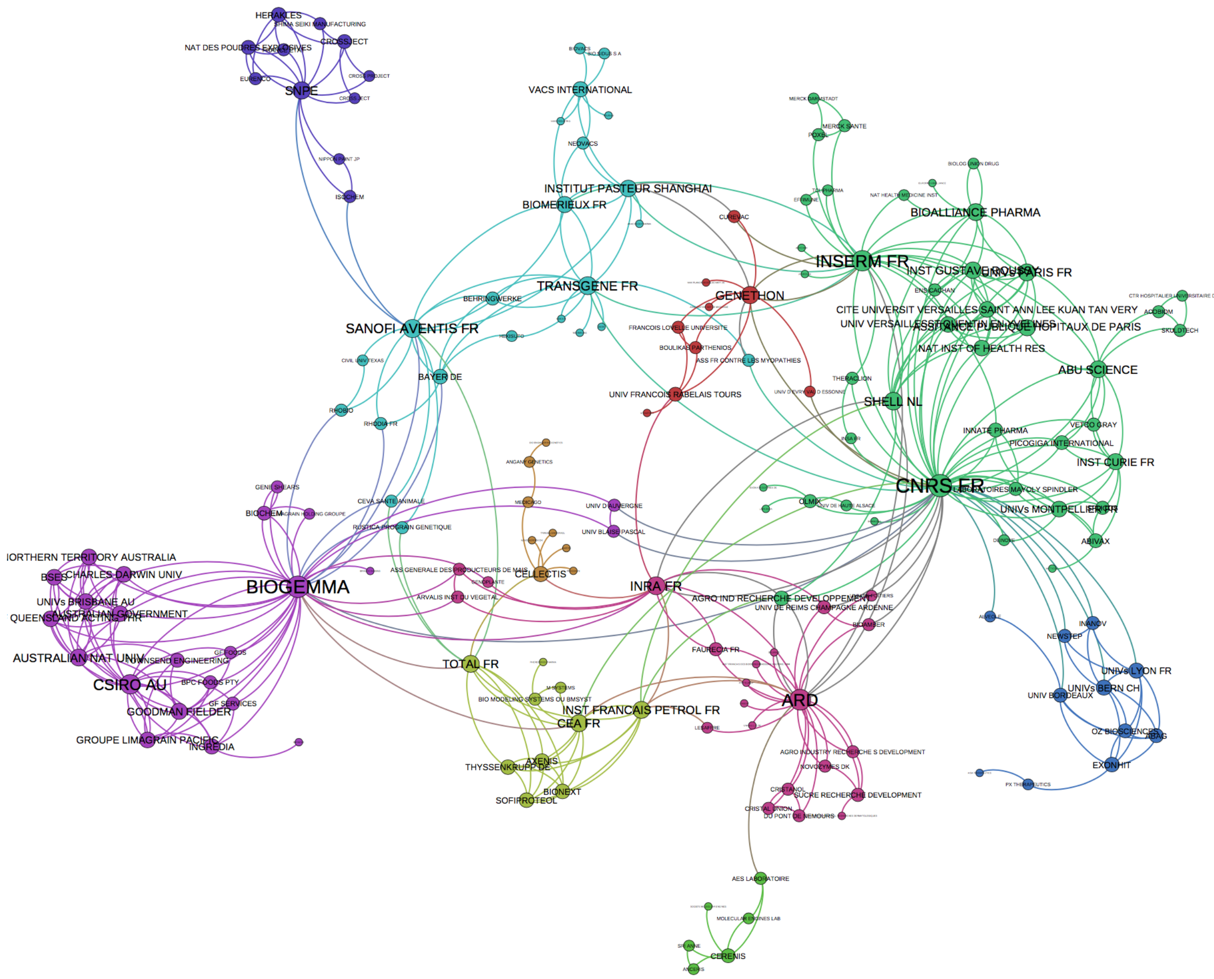

Figure 2: The collaboration network of the french Biotech sector

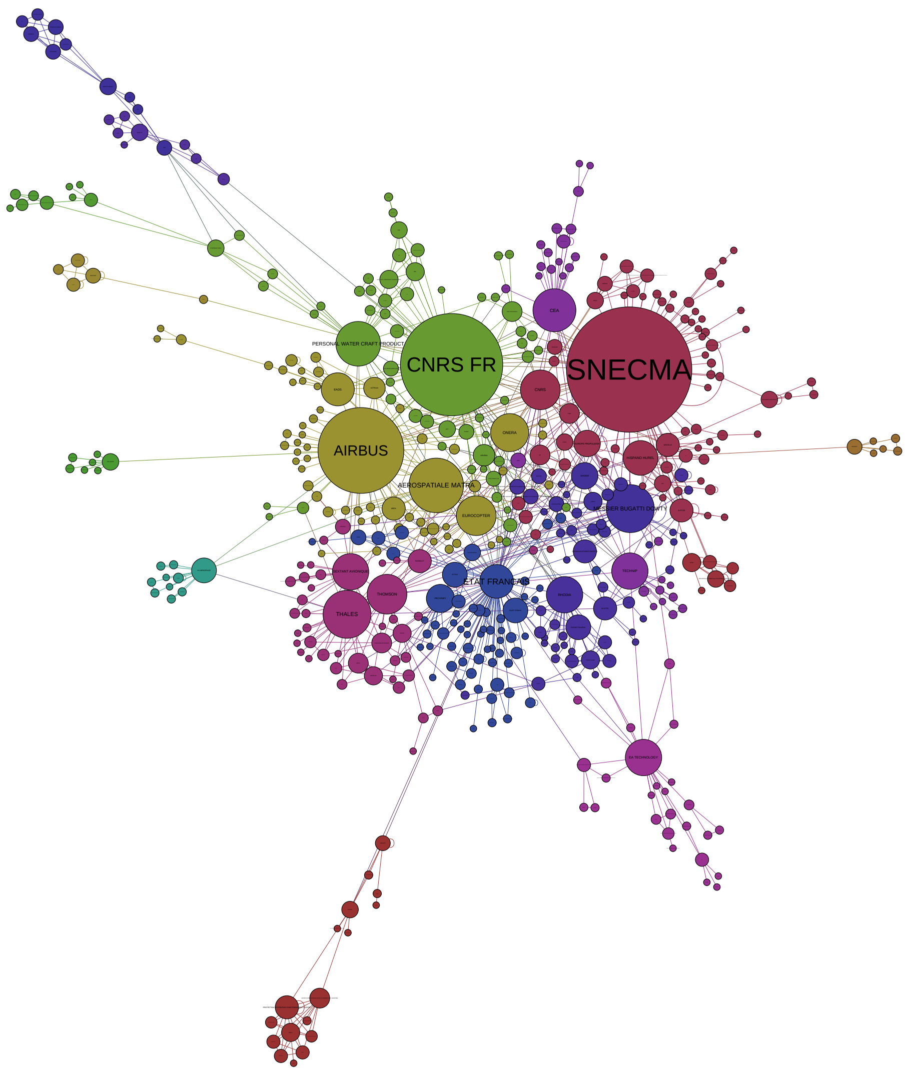

Figure 3: The collaboration network of the french Aerospace sector

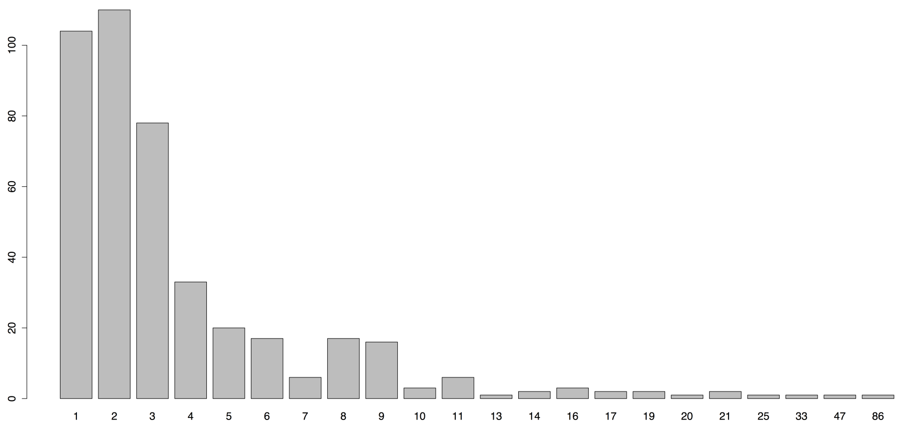

Figure 4: Example of a degree distribution

probability that a node taken at random from the graph has degree x. Note that this distribution is plotted in a log-log scale. The CFD then represents an equation linking frequency and degree. According to the network that is being represented the CFD highlights specific aspects of the network structure. In this example we can see that the relation between the density and the frequency is linear. As such the relation can be written:

$$ln(y) = a \cdot ln(x) + b \quad \forall \;\; a < 0 \tag{1}$$

This is the equation for the log-log scale. On the normal scale the form of this function is given by:

$$\begin{aligned} e^{ln(y)} &= e^{a \cdot ln(x) + b} \\ y &= e^{a \cdot ln(x)} \cdot e^{b} \\ y &= e^{ln(x^a)} \cdot e^{b} \\ y = e^{b} \cdot x^a y &= C \cdot x^a \end{aligned}$$

This highlights the fact that when we increase the density by a factor of k, the frequency drops by a factor k^a with $a < 0$. The latter is true for each value the density might take. For this reason, when the CFD of a network has a linear form on the log-log scale, the network is referred to as a scale-free network.

The scale-free network is not the only core-periphery structure. Exponential and log-normal distribution can also represent core-periphery structures. The main difference between the distribution is the manner in which the core transitions to the periphery. In a very abrupt case as in figure 5.1 the transition is very abrupt. 67% of the nodes has a degree of one while the 33% of the nodes

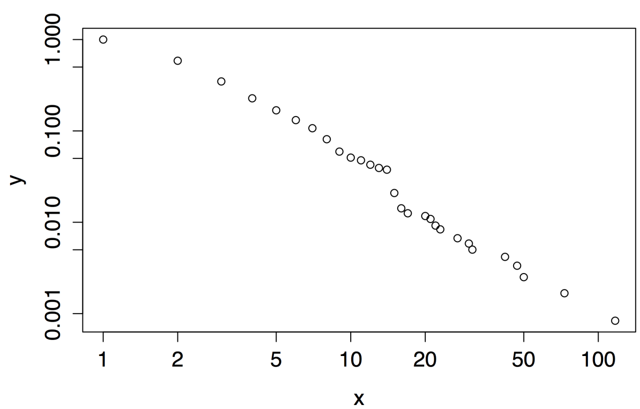

Figure 5: Example of a cumulative frequency distribution

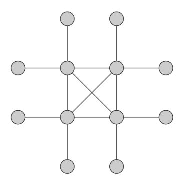

Figure 6: Core-Periphery illustration

have a degree of 5. The CFD will show a sharp drop in frequency between densities 1 and 5. The scale-free structure is a particular case in which the decrease in frequency is constant. Another case can be imagined in which there are many nodes of degree 1, 2, 3 and 4 making for a more dense periphery. This observation is close to the observation of Newman (2011) who described this idea when he observed Lorentz style curves in the distributions.

The periphery of such a network contains less nodes with a low density. The periphery is more interconnected than the scale-free network. The inverse would be true if we were to have a convex function. The shape of the adjusted function informs us about the type of core-periphery structure, ranging from sparse to dense. In order to conclude to a core-periphery structure we fit a particular function (power-law, log-normal distribution) to the data. The functions are fitted using a maximum likelihood estimation. We then use a bootstrapping method in order to assess the goodness of fit which provides us with a p.value. The null hypothesis (data comes from a power-law) is rejected when the p.value is below a threshold.

5.2 Interconnected clusters of firms: small-world networks

Small world networks are structures defined by a low average distance between nodes and a high clustering coefficient. The clustering coefficient measures how densely connected a network is; it does so by counting the number of triangles and dividing it by the possible number of triangles in the network. For instance, in a network with four nodes there are four possible triangles. If all four exist, the clustering coefficient would be $\frac{4}{4} = 1$, the highest possible score. In this example all possible links exist, the network is complete and therefore very dense. Saying that a network is highly clustered has no meaning, there needs to be a benchmark to which we can compare the clustering coefficient in order to judge if clustering is high or low. The reason why empirical networks exhibit clustering is because of social / economic / geographic / ... motivations. Individuals that are geographically close or work in the same company have a higher chance of knowing each-other and therefore to connect with each-other creating clusters. If these interactions were random there would be no (or very little) clustering. This is why a random networks i used as a benchmark (Erdös and Rényi (1959); Barabási and Albert (1999); Watts (1999)). For a given empirical network a random network with the same number of links and nodes is created and the clustering coefficient of this networks is computed. The coefficient of this network is then compared to the coefficient in the empirical network. A network is a small world if its clustering coefficient is significantly higher than that of a random graph of identical dimension (i.e same number of nodes and same number of links) Watts (1999); Gulati et al. (2012). This would imply that the graph is not random and that there are some underlying rules dictating the creation of ties in the network.

The same method is used for the average distance, one would want it to be roughly identical to that of a random graph (Watts (1999); Gulati et al. (2011)). We note Cr (Lr) the clustering coefficient (path length) of the random network and C (L) the clustering (path length) of the empirical data.

Thus, for a network to be considered a Small World we need to observe $\frac{C}{Cr} >> 1$ and $\frac{L}{Lr} \approx 1$.

5.3 Scale-free small worlds

The methods exposed in the previous subsections were applied to the network data for the biotech sector and the aerospace sector. The analysis was performed in a dynamic setting. At first, I consider a cumulative method in which each year links are added to the network, none are removed. This of course implies that links between firms at the start of the period are still present at the end of the period. I decided to test this method first because i have had the opportunity to discuss the question of collaboration is the aerospace sector with R&D decision makers from large companies of the sector. They concur to the fact that one-shot cooperations are rare. Even if no patent is deposited the collaborations are continuous, spanning often over different projects.

The analysis is also performed within a 5-years window in the aerospace network, meaning that all collaborations outside of the window are deleted from the network.

Figure 5.3 shows the fitted functions to the CDF for the aerospace sector. Whenever the reported p.value associated with the Kolmogorov-Smirnoff test (bottom left corner of each graph) exceeds a given threshold, the fit is statistically significant. We can then conclude that the degree distribution of the network comes from the same distribution as the one we check for. The green (plain) line is the log-normal function, the (dotted) red line is the power)law function. The results show that both functions are a good fit for the data. However, the power-law function is only fitted for part of the distribution, since it starts at a high density, therefore it only explains part of the network: the nodes with the highest number of links. The manner in which Airbus and its first order suppliers are interconnected follows a scale-free structure, the manner in which the first order suppliers are connected to the other suppliers does not. In the end the log-normal distribution fits the data much better. The structure of the collaboration network has a core-periphery structure. This implies that the core of the periphery of the network is rather densely connected. This is explained by the technological clusters. The core is less dense for the simple reason that the number of first order suppliers is low.

This core-periphery structure is stable over time, whether we look at the network through a five-year window or the cumulative method, the structure remains core-periphery. Given the hierarchical structure of the aerospace sector, this observation makes sense. The suppliers at the far-end of the network innovate with the constraints that their product need to be compatible for the final product (as specific part of the airplane), hence the importance of technological proximity. The resulting products then need to be recombined by larger suppliers which therefor require to master a larger diversity of knowledge in order to be able to reassemble the smaller parts. This process continues until the final assembler is reached: Airbus, which will have to largest number of links. In addition to working with the supplier it also has extensive links with universities and research institutions. No such central entity exists in the Biotech sector. The evolution of the structure of the biotechnology collaboration network shows that the network diverges from a core-periphery structure. In the early stages of the network a few large companies were interconnected creating a core. The periphery was sparse when compared to the network of the aerospace sector. As more firms enter the biotech sector, clusters are becoming increasingly well defined confining larger firms in their communities. The last graph in figure 5.3 shows that neither the power-law nor the log-normal distribution fit the data. The CFD does however present an interesting result, transition phases seem to emerge in the distribution. A first phase for low densities with sharp drops in frequency, a second for densities between 10 and 20 and a last non linear phase for the highest number of collaborations. This observations shows that there are many firms with a low degree (first phase) since there is a sharp drop in frequency between different density levels. In the middle, the drop is less important showing that there are not many firms with this intermediate level of collaboration. The final phase shows that there are few firms with many collaborations, and the scaling in their size in not linear. When analysing the final part of the distribution from a dynamic point of view, we can observe that these firms have increased the number of collaborations over time and some have managed to collaborate more than others. Int the end then, the Biotech sector is not a core-periphery network because it misses these intermediary firms and most central firm in the network. This concurs with result found by Zidorn and Wagner (2012) who showed that firms in the biotech sector tend to collaborate in order to specialise, they hence have no incentive to interconnect with clusters working on other market segments. From this perspective, the motivations for individual firms are identical in both sectors, clusters are created for the same reasons, but the interconnection of these clusters results in different overall network structures.

When we look at the small world indicators, we come to a similar conclusion. The biotech network does not present any small world characteristics. The adjusted average distance between the nodes is too high. This is explained by the same reason as for the core-periphery structure: the missing central firm. The presence of a central firm reduces the average distance between all nodes in the network. The aerospace network satisfies this condition. In both networks however, the clustering coefficient is high enough. As a final conclusion we can say that the biotech network has no canonical structure while the aerospace sector present both core-periphery and small world properties. From a dynamic perspective it reaches this structure around the year 2000. The latter is verified whether observing the network from a 5-year window or a cumulative perspective. The Aerospace network hence presents both core-periphery and small world properties. The structures are not mutually exclusive, they both highlight a different angle of the structure of a network.

From a knowledge flow perspective, it should be clear by now that the knowledge created in these networks has no reason to flow throughout the whole network. The creation and diffusion of knowledge is localised within the cluster to which it belongs.

6 Conclusion

The analysis of a collaboration network can be performed at different levels. Each level provides an insight into the understanding of the other levels. At the micro level we have shown that firms use triadic closure, repetition of collaboration and technological proximity as partner selection strategies. These strategies are used in both the biotech and the aerospace sector. These strategies result in the creation of clusters that can be explained by the specificity of each sector. In the aerospace sector, clusters emerge between firms working on parts on an aircraft that require technological overlap, while in the biotech sector collaboration is defined market segment. he inverted U-shape relation between the probability to collaborate and technological proximity can therefore be understood as both technological compatibility and competitive intensity. The real difference between the two sectors appears once we study the global

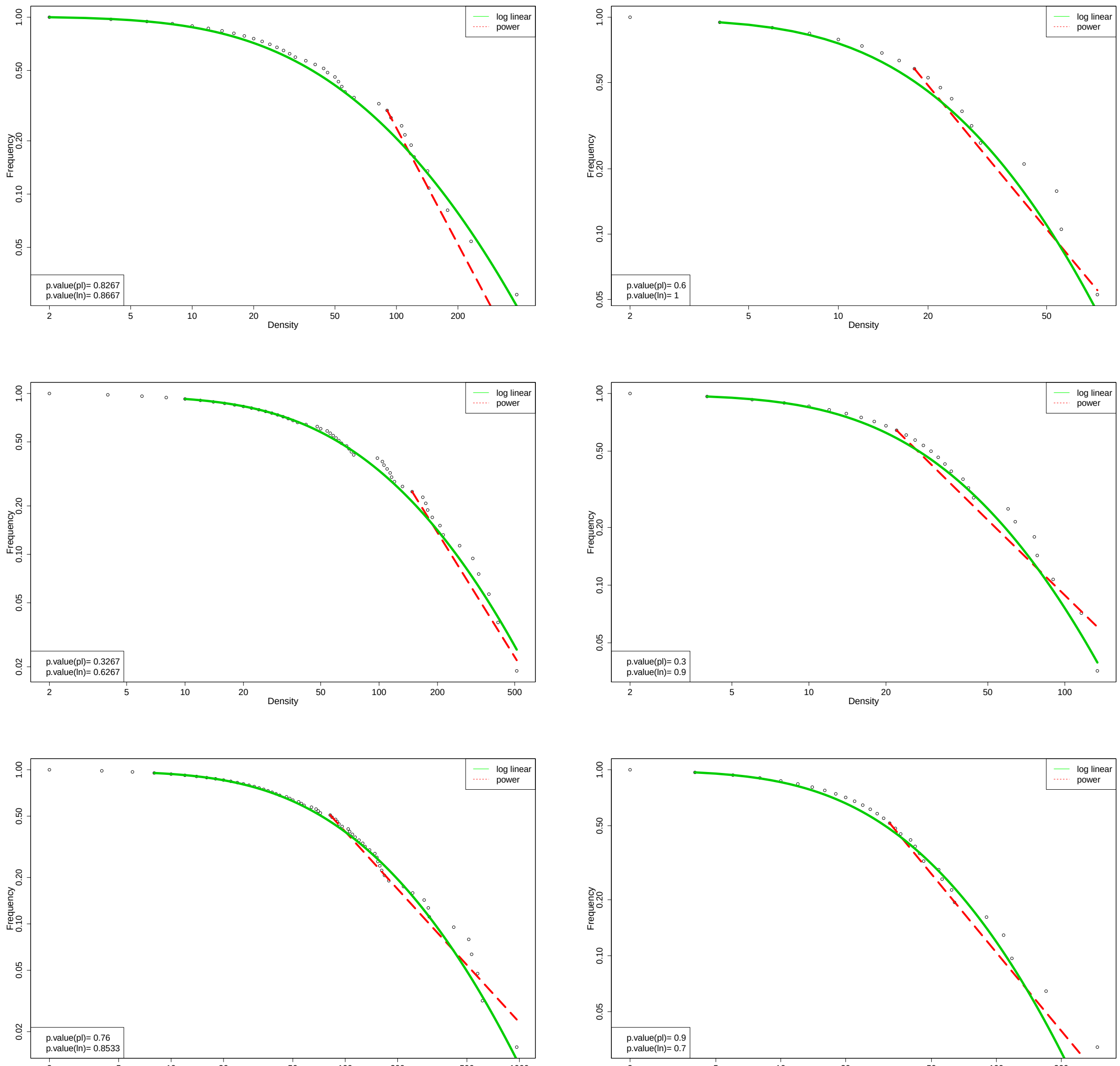

Figure 7: Core-Periphery identification in the Aerospace Collaboration network. The left column provides the window data, the right column the cumulative data.

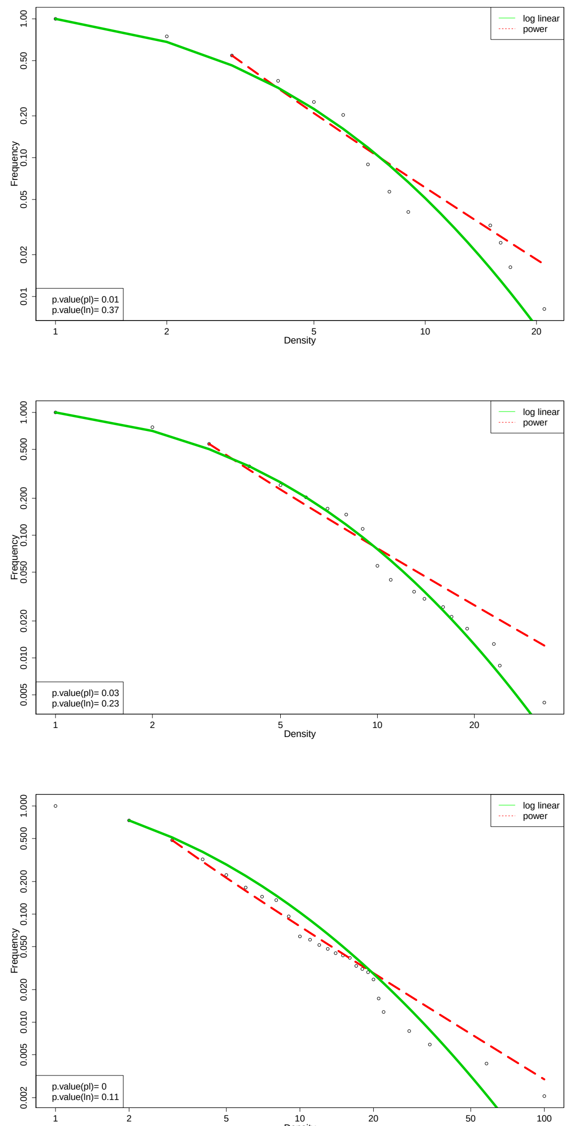

Figure 8: Core-periphery identification in the French Biotech Collaboration network. Data from the cumulative method only.

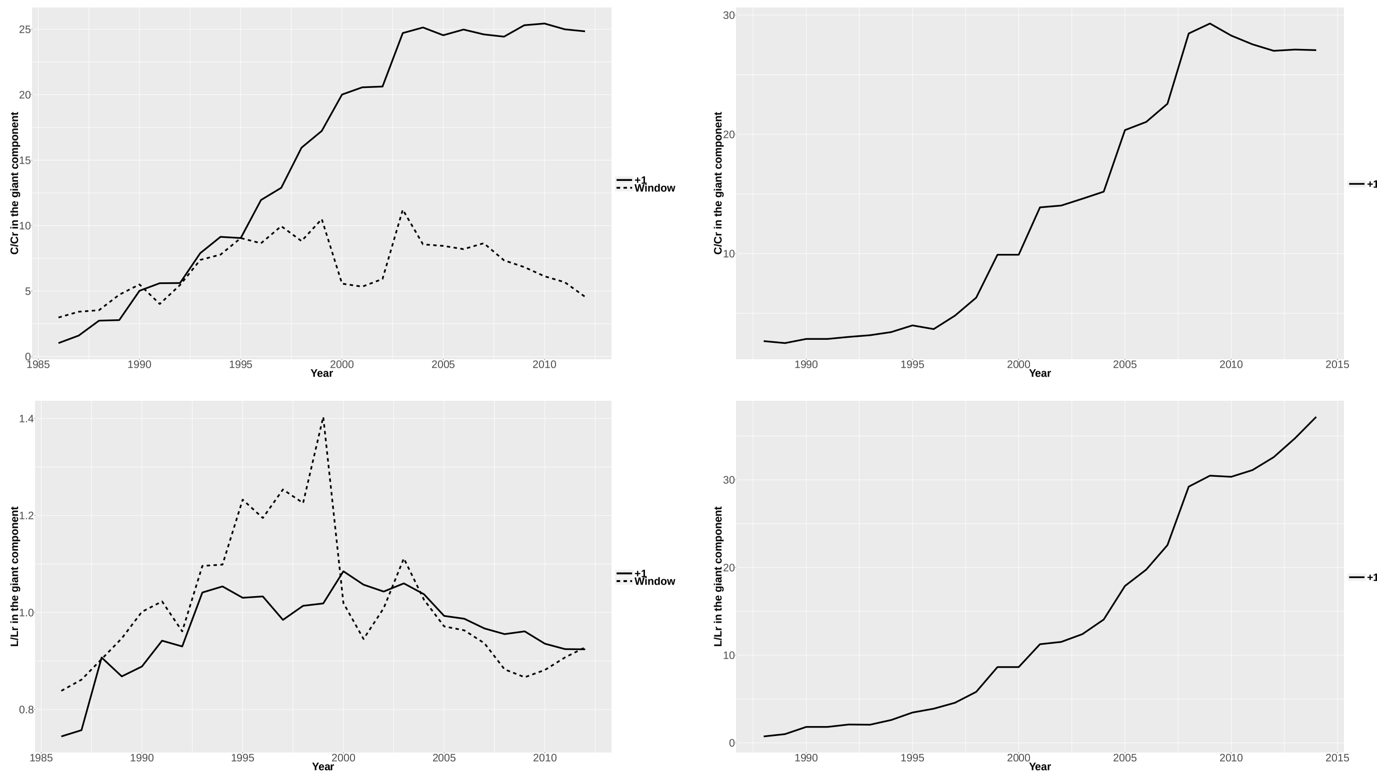

Figure 9: Adjusted clustering coefficient and adjusted average distance and its evolution over time in the French Aerospace and Biotech collaboration networks. The left column contains the graphs for the aerospace sector. The right column the data for the Biotech sector.

network structure, in other words, the manner in which the different communities are interconnected. The value chain structure of the aerospace sector creates a network with a few firms with a high level of collaborators at its core. These firms ensure the efficient flow of knowledge coming from the different communities inside the network. This core is absent from the biotech collaboration network. The structure of a network provides information about R&D strategies of firms, even if the strategies are the same at the firm level they still can result in very different network structures as shown by the results here. If one is interested in understanding the manner in which new knowledge is created, analysing the different levels is a requirement. The structure of the network shows that knowledge is created in clusters that are sparsely interconnected through a small number of large firms. Since there is barely any inflow of new knowledge in these clusters, a risk of diminishing diversity exists. In addition we know that firms prefer working with historical partners and choose to work with partners of partners inducing an even greater risk of loss of diverse knowledge. The latter is the case in both sectors. The main difference resides in the presence of a central firm interconnecting the different clusters in the biotech sector compared to the aerospace sector. From a knowledge flow perspective this reduces the speed at which knowledge could flow through the network. The role of this firm is vital in the aerospace sector since this central firm centralised all knowledge and assembles the airplane, there is no such need in the biotech sector. Some firms work on some market tiers but none are present on all even though there might be a potential for learning from firms in all these tiers.

References

M. Algamdi, S. McDonald, and B. Pailthorpe. *The Emergence of a Small World in a Network of Research Joint Ventures*. University of Queensland, School of Economics, 2012.

A-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

B. R. Barringer and J. S. Harrison. Walking a tightrope: Creating value through interorganizational relationships. *Journal of Management*, 36(3):367–403, 2000.

V. D. Blondel, J-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, (10), 2008.

J. Brennecke and Olaf N Rank. The firm's knowledge network and the transfer of advice among corporate inventors—a multilevel network study. *Research Policy*, 46(4):768–783, 2017.

S. Breschi and F. Lissoni. Knowledge spillovers and local innovation systems: a critical survey. *Industrial and Corporate Change*, 10(4):975–1005, 2001.

A. Caimo and N. Friel. Bayesian inference for exponential random graph models. *Social Networks*, 33(1):41–55, 2011.

B. Cassiman and R. Veugelers. R&d cooperation and spillovers: some empirical evidence from belgium. *The American Economic Review*, 92(4):1169–1184, 2002.

M. G. Colombo. Firm size and cooperation: the determinants of cooperative agreements in information technology industries. *International Journal of the Economics of Business*, 2(1):3–30, 1995.

P. Erdös and A. Rényi. On random graphs, i. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.

E. Estrada. *The structure of complex networks: theory and applications*. Oxford University Press, 2012.

Pyka S. Gilbert, N. and P. Ahrweiler. Innovation networks - a simulation approach. *Journal of Artificial Societies and Social Simulation*, 4(3), 2001.

R. Gulati, D. Lavie, and R. R. Madhavan. How do networks matter? the performance effects of interorganizational networks. *Research in Organizational Behavior*, 31:207–224, 2011.

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

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

J. Hagedoorn and R. Narula. Choosing organizational modes of strategic technology partnering: international and sectoral differences. *Journal of International Business Studies*, 27(2):265–284, 1996.

J. K. Harris. *An introduction to exponential random graph modeling*, volume 173. Sage Publications, 2013.

A. B. Jaffe. Technological opportunity and spillovers of r & d: Evidence from firms' patents, profits, and market value. *The American Economic Review*, 76(5):984–1001, 1986.

B. Kogut and U. Zander. Knowledge of the firm and the evolutionary theory of the multinational corporation. *Journal of International Business Studies*, 34: 526–529, 2003.

M. König, C. J. Tessone, and Y. Zenou. A dynamic model of network formation with strategic interactions. *CEPR Discussion Paper No. DP7521*, 2009.

A. Lomi and F. Pallotti. Relational collaboration among spatial multipoint competitors. *Social Networks*, 34(1):101–111, 2012.

D. Lusher. Masculinity, educational achievement and social status: A social network analysis. *Gender and Education*, 23(6):655–675, 2011.

D. Lusher, J. Koskinen, and G. Robins. *Exponential random graph models for social networks: Theory, methods, and applications.* Cambridge University Press, 2012.

B. McEvily and A. Marcus. Embedded ties and the acquisition of competitive capabilities. *Strategic Management Journal*, 26(11):1033–1055, 2005.

D. C. Mowery, J. E. Oxley, and B. S Silverman. Strategic alliances and interfirm knowledge transfer. *Strategic management journal*, 17(S2):77–91, 1996.

Mohammad Masrurul Mowla. An overview of strategic alliance: Competitive advantages in alliance constellations. *Journal of Business Management and Corporate Affairs*, 1(1):1–10, 2012.

M. E. J. Newman. *Networks, an introduction.* Oxford Univertsity Press, 2011.

W. W. Powell and S. Grodal. Networks of innovators. *The Oxford Handbook of Innovation*, pages 56–85, 2005.

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

A. Pyka. Innovation networks in economics from the incentive-based to the knowledge-based approaches. 1999.

E. M. Rogers. *Diffusion of innovations*. Simon and Sschuster, 1995.

Jasjit Singh. Collaborative networks as determinants of knowledge diffusion patterns. *Management science*, 51(5):756–770, 2005.

A. L. J. Ter Wal. The dynamics of the inventor network in german biotechnology: geographic proximity versus triadic closure. *Journal of Economic Geography*, 2013.

W. Tsai. 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, 2001.

B. Verspagen and G. Duysters. The small worlds of strategic technology alliances. *Technovation*, 24(7):563–571, 2004.

J. Watson. Modeling the relationship between networking and firm performance. *Journal of Business Venturing*, 22(6):852–874, 2007.

D. J. Watts. Networks, dynamics, and the small-world phenomenon 1. *American Journal of Sociology*, 105(2):493–527, 1999.

W. Zidorn and M. Wagner. The effect of alliances on innovation patterns: an analysis of the biotechnology industry. *Industrial and Corporate Change*, 22 (6):1497–1524, 2012.

Cahiers du GREThA
Working papers of GREThA

GREThA UMR CNRS 5113

Université de Bordeaux

Avenue Léon Duguit
33608 PESSAC - FRANCE
Tel : +33 (0)5.56.84.25.75
Fax : +33 (0)5.56.84.86.47

http://gretha.u-bordeaux.fr/

Cahiers du GREThA (derniers numéros – last issues)

2017-13 : BONIN Hubert, Le quasi-krach de la Société générale en 1913 : une crise de son modèle économique ?

2017-14 : ROUGIER Eric, LECTARD Pauline, Can developing countries gain from defying comparative advantage? Distance to comparative advantage, export diversification and sophistication, and the dynamics of specialization

2017-15: MOYES Patrick, EBERT Udo, The Impact of Talents and Preferences on Income Inequality

2017-16: RAZAFIMANDIMBY Andrianjaka Riana, ROUGIER Eric: What difference does it make (when a middle-income country is caught in the trap)? An evidence-based survey analysis of the determinants of Middle-Income Traps

2017-17: LECHEVALIER Sébastien, DEBANES Pauline, SHIN Wonkyu: Financialization and industrial policies in Japan and Korea: Evolving institutional complementarities and loss of state capabilities

2017-18: NAVARRO Noemí, VESZTEG Róbert: On the empirical validity of axioms in unconstrained bargaining

2017-19: LACOUR Claude, GAUSSIER Nathalie: Un écosystème sur la vague ? L'arrivée de la LGV à Bordeaux et l'écosystème start-up

2017-20 : FRIGANT Vincent, MIOLLAN Stéphane, PRESSE Maëlise, VIRAPIN David : Quelles frontières géographiques pour le Technological Innovation System du véhicule à pile à combustible ? Une analyse du portefeuille des co-brevets des constructeurs automobiles

2017-21 : LEVASSEUR Pierre, ORTIZ-HERNANDEZ Luis : Comment l'obésité infantile affecte la réussite scolaire ? Contributions d'une analyse qualitative mise en place à Mexico

2017-22 : A. LAGARDE, A. AHAD-CISSÉ, S. GOURGUET, O. LE PAPE, O. THÉBAUD, N. CAILL-MILLY, G. MORANDEAU, C. MACHER, L. DOYEN: How MMEY mitigates bio-economic impacts of climate change on mixed fisheries

2018-01: BALLET Jérôme, POUCHAIN Delphine: Fair Trade and the Fetishization of Levinasian Ethics

---

La coordination scientifique des Cahiers du GREThA est assurée par Valerio STERZI. La mise en page et la diffusion sont assurées par Julie VISSAGUET