Building innovation networks: the process of partner selection by young knowledge intensive firms

Cristina Sousa

Margarida Fontes

2013

WP n.º 2013/09

DOCUMENTO DE TRABALHO

WORKING PAPER

DINAMIA'CET
CENTRO DE ESTUDOS SOBRE A MUDANÇA SOCIOECONÓMICA E O TERRITÓRIO
ISCTE-IUL

FCT
Fundação para a Ciência e a Tecnologia

Building innovation networks: the process of partner selection by young knowledge intensive firms

Cristina Sousa*

Margarida Fontes**

WP n.º 2013/09

DOI: 10.7749/dinamiacet-iul.wp.2013.09

*Cristina Sousa, Instituto Universitário de Lisboa (ISCTE-IUL) and DINÂMIA'CET-IUL, Portugal (cristina.sousa@iscte.pt).
**Margarida Fontes, LNEG and DINÂMIA'CET-IUL, Portugal (margarida.fontes@lneg.pt).

Building innovation networks: the process of partner selection by young knowledge intensive firms[1]

Abstract

This paper addresses the selection of partners in innovation networks. It builds on the existing literature to develop an integrative framework that encompasses the main factors identified as influencing selection of partners by young knowledge-intensive firms. It considers that both persistence and novelty are present in the network building process, and so integrates several explanations advanced by the literature: social capital, imprinting and inertia for tie persistence; network embeddedness and proximity for new tie selection.

Using a rare event logit model, we estimate the likelihood of selecting an innovation partner using data about the partnerships established by young Portuguese biotechnology firms, purposefully collected through questionnaire-based face-to-face interviews, complemented with documentary information. The results uncover different network building strategies in terms of partner selection to access the different types of resource needed for innovation and highlight the advantages of adopting an integrated framework.

Keywords: Innovation network, partner selection, tie persistence, social capital, network embeddedness, proximity

[1] This paper draws on the research carried out within the Project ENTSOCNET - Social networks, entrepreneurs and access to knowledge: the case of biotechnology and the IT industries, funded by FCT – Fundação para a Ciência e a Tecnologia (POCI/ESC/60500/2004), Portugal. A previous version of the paper was presented at the 14th International Schumpeter Society Conference, Brisbane, 2-5 July 2012.

1. INTRODUCTION

Understanding how firms select their innovation partners is vital to grasp the evolution of inter-organizational networks. The process of partner selection has been addressed in the literature, but research tends to focus on individual factors and/or to have an exclusively theoretical approach. This paper builds on the existing literature to develop an integrative framework that encompasses the main factors identified as influencing selection of innovation partners by young knowledge-intensive firms; and assesses their combined impact on the probability of partner selection.

The selection of partners is designed (Nooteboom, 2008) and affected by search costs and uncertainty, raising adverse selection and moral hazard problems (Kirkels and Duysters, 2010). When selecting a partner, firms can rely on their past relationships or look for a new organization. In the first case, firms select organizations they know from prior partnerships (Gulati, 1995a) or with whom entrepreneurs have personal relations (Hallen, 2008) and we are in the presence of persistence and of path dependent processes (Walker et al, 1997). In the second case, new actors join the network, bringing novelty and variety that are vital for innovation (McEvily and Zaheer, 1999) and their selection is driven by evaluation mechanisms, since there is no direct knowledge of partners' capabilities (Li and Rowley, 2002).

Despite the relevant contributions of previous studies, the process of partner selection is not yet fully understood, especially in the case of new firms (Grossman et al, 2012). In this paper we argue that it is necessary to adopt an integrated perspective, considering simultaneously the several (complementary) factors identified so far; and to submit theoretical propositions to empirical testing. Thus the paper proposes a framework that: i) combines various factors identified in previous research as influencing partner selection, relating them with persistence and novelty; ii) addresses network building as a sequential process, along which decisions concerning partner selection are made, being influenced by these factors.

On the basis of this framework, we build and empirically test a (logit) model of partner selection that takes in consideration both persistence effects (associated with selection of partners known from previous relationships) and evaluation mechanisms (associated with the selection of novel, unknown partners).

2. THEORETICAL BACKGROUND

2.1 Tie persistence

Tie persistence is an important mechanism in the construction of inter-organizational networks. Previous research on alliances uncovered firms propensity to establish relations with organizations they know from prior partnerships (Gulati, 1995a), resulting in path-dependent routines on partner selection (Li and Rowley, 2002). Trust and learning effects are also described as arising from the repeated interaction in previous relationships (Gulati, 1995a; Hallen, 2008). Thus, tie persistence contributes for the reduction of search costs and uncertainty, since it allows firms to discern capable and reliable partners, based on previous experiences (Gulati and Gargiulo, 1999).

Start-up firms do not have these previous alliance-based relationships. So, entrepreneurship scholars highlight the importance of entrepreneurs' previous personal relations (Adobor, 2006), often related with their social capital (Anderson et al, 2007), in firm creation process. Since the professional and academic trajectory of the entrepreneurs is a basic element in the formation of their personal networks, it is often assumed that relationships established along this trajectory become automatically part of the early network of the new firm (Shane and Stuart, 2002). In the limit, the firm's network at start-up is equated with its entrepreneurs' social capital (Hsu, 2007). However, elsewhere we found that trajectory ties are not automatically transformed in firms' ties (Fontes et al, 2012). Rather, entrepreneurs assess the utility of their personal contacts and only select those considered as valuable for the firm.

Ties that originate from the entrepreneurs' social capital have several advantages. They are usually characterized by higher levels of trust, which facilitate communication and information exchanges (Burt, 1997). Moreover, since they are often based on shared experiences, there is a good understanding of the potential contributions they can offer (Koka and Prescott, 2002). These experiences may also have led to the development of cognitive proximity, facilitating the transmission of knowledge, particularly when it is complex or less structured (Breschi and Lissoni, 2001).

However, exactly because these ties are associated with the entrepreneurs' personal trajectory, they may be less useful when it comes to accessing resources and competences that are more distant from the entrepreneur's own experience (Ensley and Hmieleski, 2005). Scholars point to the advantages of diversity in network composition, since ties with similar actors have reduced benefits in terms of information and knowledge (Nooteboom, 1999). Therefore, establishing relations with a diverse set of actors lessens the risks of redundancy (Burt, 1992) and over-embeddedness (Adobor, 2006, Uzzi, 1997) and facilitates the access to different types of knowledge (Baum et al., 2000).

Scholars also stress the importance of decisions made at start-up in the subsequent development of the company. These are described as having an “imprinting effect” (Stinchcombe, 1965; Eisenhardt and Schoonhoven, 1990), since they shape firms’ choices regarding resource mobilization, competence development and search for partners. Regarding the latter, Milanov and Fernhaber (2009) found that initial partnerships have a long term impact on firms’ access to network resources, since the network size and centrality of the initial partners influence the subsequent size of the new venture's network.

As firms evolve, behavioural persistence at organizational level, related with the prevalence of routines and inertia, emerges (Kim et al, 2006). The development of relation-specific routines reduces the probability of alliance partner replacement based solely on economic evaluation and brings an element of rigidity into the construction of networks (Kim et al, 2006). Even when a new partner can provide better resources than the existing one, firms may maintain the old relation (Reuer et al, 2002), since it has allowed relation-specific assets to be built (Ebers, 1999). In this sense, network inertia is not a signal of poor management, but a by-product of successfully managed networks (Kim et al, 2006).

2.2 New ties

The satisfaction of resource needs also relies on the establishment of new relationships, intentionally built, which bring novel information and knowledge (Baum et al, 2000). The selection of the new network members is driven by evaluation mechanisms, since there is no direct knowledge of partners’ capabilities (Li and Rowley, 2002).

Scholars sustain that the selection of unknown organizations has to be understood in the context of existing networks. The embeddedness in inter-organizational networks enables the access to some information about the quality of potential partners and therefore reduces the uncertainty about them (Human and Provan, 2000; Glückler and Armbrüster, 2003). In this sense, an organization’s new tie opportunities are shaped by the characteristics of the network where it is embedded (Grossman et al, 2012).

The structure of the whole network influences each actor’s actions, since its position in the network constrains the set of available actions (Marsden, 1981; Gulati, 1998). Some studies show that firms tend to form partnerships with organizations they know indirectly, i.e., with whom they share a partner (Gulati, 1995b; Hallen, 2008), or with organizations that occupy a central position in the network, thus signalling their quality and reliability as sources of resources (Powell et al, 1996; Gulati and Garguilo, 1999, Ahuja, 2000). Therefore, although the configuration of the whole network is influenced by the characteristics of the dyads, the whole is more than the sum of its parts and, in turn, affects the occurrence of a tie.

Another line of research departs from the embeddedness perspective and provides some insights about the selection of “socially distant” ties. Some studies stress the role of “assortative mechanisms”, i.e., of the compatibility and complementarity between partners’ attributes (Rivera et al, 2010). Thus, new ties are preferably formed with organizations with which firms share some traits, since similarity (homophily) favours trust-building and ease of communication (McPherson et al, 2001). Following this line of reasoning, some authors focus on various forms proximity (Boschma and Frenken, 2010; Nooteboom et al, 2007) as a factor that facilitates resource exchanges.

The importance of localised resource exchange has been extensively discussed in the literature, especially in the case of knowledge (Breschi and Lissoni, 2001), but also for non-technological resources (Sorenson and Stuart, 2001). Scholars stress the importance of co-localisation for learning and exchange of information and knowledge (Lorenzen, 2007).

More recently, scholars have pointed out the importance of non-geographical forms of proximity. Some degree of cognitive proximity is necessary to assess the value of the knowledge produced, fully understand it and absorb and apply it effectively (Cohen and Levinthal, 1990). Institutional/organizational proximity helps to manage resource exchange and reduces transactions costs (Boschma, 2005). However, for knowledge exchange, Boschma and Frenken (2010) identified a proximity paradox: too much proximity between organizations can reduce firms’ innovative performance. Similarly, Nooteboom (2000) found an inverted U-shape relation between cognitive distance and innovative performance, and thus evidence of the existence of an optimal distance.

2.3 Building an integrated framework

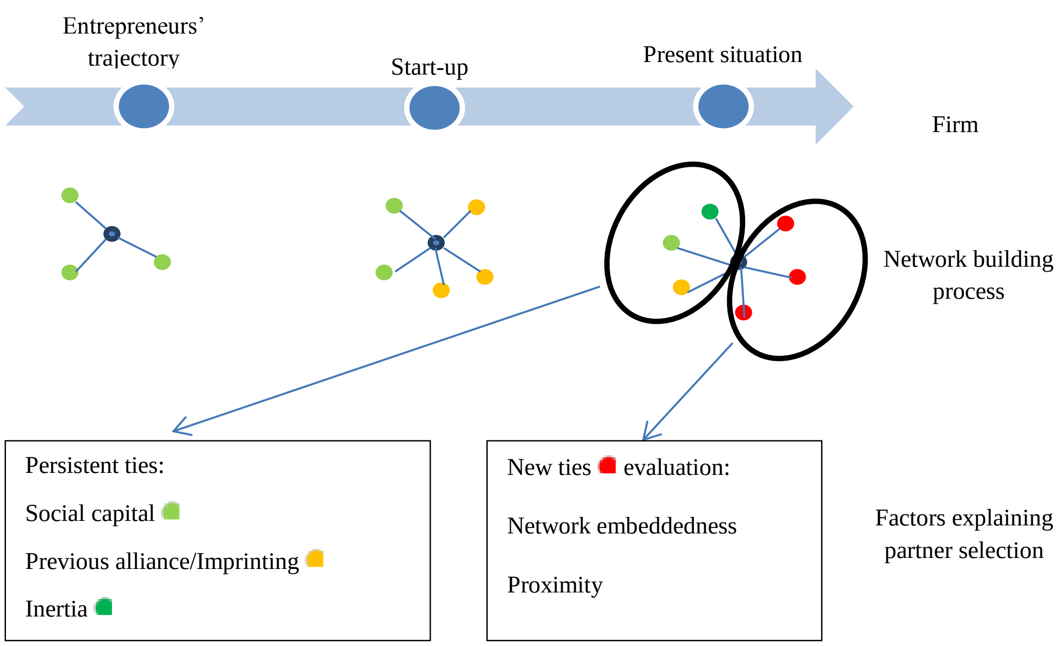
In order to pursue innovation activities, firms rely on a set of internal resources and competences which they combine with external ones, accessed via market and non-market transactions. Networks are considered essential in this process of resource gathering (Ozman, 2009), particularly in science-based sectors (Baum et al, 2000). So, in this research we consider that network partners provide resources for the innovation process.

Previous research has shown that the type of resource being accessed influences the type of networks that are established (Sammarra and Biggiero, 2008; Salavisa et al, 2012; Sousa et al, 2011). Thus, it is also likely to influence the process of partner selection. Therefore, we distinguish three types of resources - S&T knowledge, complementary assets and legitimacy/ credibility – and look at the process of partner selection in each case.

The literature has also shown that resources requirements change over time (Delmar and Shane, 2004). So, partners that are useful at a certain point of the firm's history may be useless at other points. Additionally, firms can make mistakes in selecting partners and subsequently correct them, or they may change their strategy with impact on the resource needs and thus on the type of partners required (Druilhe and Garnsey, 2004, Costa et al, 2004). These facts have implications for the dynamics of network building. Therefore, inter-organizational networks change on a continuous basis (Kim et al, 2006). To acknowledge this, we adopt a sequential approach to the process of network building in which three different phases are considered: entrepreneurs' academic and professional trajectory up to start-up, start-up (the year of formal creation and the two subsequent years of activity) and present moment (the time the information was collected).

The proposed framework (Figure 1) introduces the possibility of maintaining previous partners (or not), or selecting new ones (or not) on a continuous basis. Therefore, we consider that, at start-up, firms can mobilize entrepreneurs' pre-existing ties with organizations from their trajectory, or build new relations. Similarly, in the present moment firms can maintain the relationships with start-up partners or renew previous relationships with trajectory organizations not yet mobilized; or they can build new ones. As mentioned above, the selection of these new ties is driven by evaluation mechanisms, since there is no direct knowledge of partners' capabilities (Li and Rowley, 2002).

Figure 1 – Integrated framework



Hereby the framework enables us to consider both persistent and new ties; and to integrate the several arguments advanced in the literature to explain partner selection, namely social capital, imprinting and inertia for tie persistence, and network embeddedness and proximity for new tie selection.

3. METHOD

3.1 Empirical strategy and data sources

We model the probability that a firm i selects an organization j as a partner and thus forms a tie to access resources for innovation. Following other studies of tie formation, we use a logit model, considering all feasible dyads (Gulati, 1995b; Stuart, 1998; Gulati and Gargiulo, 1999; Roijakkers et al, 2005).

The analysis is based on the ties established by 13 young Portuguese biotechnology firms[2]. Our data base includes, for each firm, all feasible dyads, both those that have materialized and those that have not. We consider that all the 459 organizations present in the current sectoral innovation network (Figure A1 in appendix) could have been selected by each of the 13 firms. In addition we also include, for each firm, the organizations from the entrepreneurs' trajectory and the partners that were chosen at start-up, but are not present in the current sectoral innovation network, i.e. those that have decayed.

Considering all feasible dyads as a sampling procedure poses two empirical difficulties (Sorenson and Stuart, 2001). First, the observations may be interdependent because each firm appears in many dyads creating a common-actor effect. Second, the materialisation of a dyad in this sample is a rare event. Given the fact that the largest innovation network for a firm is composed of 182 organizations and that the feasible dyads for each firm exceed 500, this would imply a large number of zeros. In fact, the database includes 968 materialized dyads in a set of 6786 feasible ones. So, the ratio of materialized to non-materialized dyads is very small (14%).
For these reasons, drawing on the work of Sorenson and Stuart (2001) we have adopted a rare event logit model using the relogit stata procedure (Tomz et al, 1999) and applied a choice-based sampling procedure.

Therefore the sample used in the regressions includes all the materialized dyads (irrespective of the moment when they took place, i.e. on the entrepreneurs' trajectories, at start-up or at the present moment) and a matched sample of relations that have not occurred. These were randomly chosen from the list of organizations present in the current sectoral network. Thus, the matched sample includes 1936 dyads (both materialized and non-materialized) involving 660 partner organizations. As a result each partner enters in the sample an average of 2.9 times.

Data collection was conducted through face-to-face interviews, based on a semi-structured questionnaire. Data was collected on the entrepreneurs' personal network and on its role to firm start-up and early growth, permitting to obtain fine grained information about the people who

[2] This sample was obtained from a larger research project, that encompassed the universe of Portuguese molecular biology firms (23 firms) (Salavisa and Fontes, 2012), from which were selected the firms over 3 years old.

were/are important during the two periods, including the origin of the relationships and the type, nature and relevance of their respective contributions. Data was equally collected on firm innovation and technological strategies and activities, including cooperation arrangements (both formal and informal). This was complemented by documentary information that included: the Curriculum Vitae (CV) of the entrepreneurs, published data about formal collaborative projects, partnerships and patents, and a variety of information about the entrepreneurs' personal trajectories and firm formation histories.

The data obtained enabled the (re)construction of entrepreneurs´ academic and professional trajectories and of firms' innovation networks, both at start-up and at the moment of the interview (for a detailed description see Sousa, 2012). It has also permitted to distinguish between ties established to access the three types of resource defined: S&T knowledge, complementary assets and legitimacy/credibility. The concept of multiplex tie (Degenne and Forsé, 1994) is used to acknowledge the possibility that the same partner acts as a source of two (duplex) or three (triplex) different types of resources. Table 1 presents some descriptive statistics about the number of dyads for each moment and resource.

Table 1 – Number of dyads in firms' innovation networks

Moment	Resource	Mean	Std. Dev.	Min	Max
Start-up	S&T knowledge	7	7.2	1	25
	Complementary assets	7	2.9	2	14
	Legitimacy/credibility	5	4.2	1	16
	All (innovation network)	14	10.0	3	36
Present	S&T knowledge	18	23.9	1	91
	Complementary assets	45	42.5	4	119
	Legitimacy/credibility	4	4.1	0	15
	All (innovation network)	61	55.8	6	182

3.2 Variables

The dependent variable in all models, tie formation, is a dichotomous variable for the occurrence of a tie, which mirrors the selection of a partner. It assumes the value of one when a certain organization j is mobilized for innovation purposes by a firm i. We start by considering all resources and then distinguish between S&T knowledge, complementary assets and legitimacy/credibility. So, four different models are estimated.

The independent variables are organized in different groups, capturing all the dimensions referred in the extant literature already mentioned. In Table 2 we briefly present all the variables and in Table A1 (in the appendix) we report their descriptive statistics.

3.2.1 Variables capturing tie persistence

To capture the effect of the entrepreneur's *social capital* we consider a variable that indicates if the dyad derives from the entrepreneur's previous academic and professional trajectory (TRAJij). To capture the effect of *previous alliance/imprinting* we consider a variable that indicates the existence of the dyad at start-up. Dyads are distinguished according to the resource was being accessed: INNOVSUij (all resources), KNOWSUij (knowledge), CASUij (complementary assets) and LCij (legitimacy/credibility). Finally, to capture the effect of *network inertia* we consider a variable that indicates whether a relation originated from the entrepreneur's trajectory was activated to access resources for innovation at start-up (INERij).

3.2.2 Variables capturing tie evaluation

To capture the effect of *network embeddedness* the model includes two variables. To indicate the partner's positioning in the sectoral network, we use a measure of network centrality: outdegree centrality of each partner in the previously existing network (POC). This measure shows the number of ties that depart from a partner: central partners provide resources to a large number of firms and are characterized by intensive activity. To capture the *share of third partners*, and since we do not have indirect ties in the (re)constructed sectoral network, we resort to the network concept of clique. A clique is a sub-set of actors in which each one is connected to all others. Since we want to capture the existence of indirect ties, the 2-clique concept is used, i.e., a clique where the actors are connected directly or through a common neighbour and only cliques with more than three members are considered. So, our variable (NCLIQUES) considers the number of 2-cliques in which both the firm i and the partner j are present, excluding the existence of a direct tie.

To capture *geographical proximity* (PGEO) between the firm and its partners, each organization's location was considered and partners were classified in two groups: national (Portuguese) and foreign. To capture *organizational proximity*, we follow Broekel and Boschma (2012), who draw on Metcalfe's concept of organizational proximity based on the similarity of routines and incentive mechanisms, to argue that a profit and a non-profit organization have a low degree of organizational proximity, which lowers their probability to connect and collaborate. Therefore, we distinguish different types of organizations - biotechnology firms, firms from other sectors, universities and research centres, hospitals, S&T parks, financial institutions and other organizations (e.g. trade and professional associations and government agencies) - and include one variable to capture the culture of a profit organization (PFIRM) and one to capture the culture of an academic organization (PUNIV). However this distinction should be regarded with care in the case of science-based firms, which often perform an intermediate function between science and the market (Fontes, 2005; Stuart et al, 2007) and whose funders are frequently scientists. In fact, these firms tend to be close to academic culture (Ensley and Hmieleski, 2005).

Table 2– Variables definition

Variable	Explaining Factor	Description	Level	Construct
Dependent				
INNOVPij	-	The tie is present in the firm's i innovation network, indicating the selection of partner j.	Dyad	A dichotomous variable denoting whether there is a relation between i and j to access innovation resources
KNOWPij	-	The tie is present in the firm's i knowledge network, indicating the selection of partner j.	Dyad	A dichotomous variable denoting whether there is a relation between i and j to access scientific and technological knowledge
CAPij	-	The tie is present in the firm's i complementary assets network	Dyad	A dichotomous variable denoting whether there is a relation between i and j to access complementary assets
LCPij	-	The tie is present in the firm's i legitimacy/credibility access network, indicating the selection of partner j.	Dyad	A dichotomous variable denoting whether there is a relation between i and j to access legitimacy/credibility
Dependent variables capturing tie persistence				
TRAJij	Social capital	The tie is present in the academic and professional trajectory of the entrepreneurial team	Dyad	A dichotomous variable denoting whether the organization j was part of the trajectory of i's entrepreneurial team
INNOVSUij	Previous alliance/imprinting	The tie was present in the firm's i innovation network at start-up, indicating the selection of partner j at that moment	Dyad	A dichotomous variable denoting whether there was a relation between i and j to access innovation resources at start-up
KNOWSUij	Previous alliance/imprinting	The tie was present in the firm's i knowledge network at start-up, indicating the selection of partner j at that moment	Dyad	A dichotomous variable denoting whether there is a relation between i and j to access scientific and technological knowledge at start-up
CASUij	Previous alliance/imprinting	The tie was present in the firm's i complementary assets network at start-up, indicating the selection of partner j at that moment	Dyad	A dichotomous variable denoting whether there is a relation between i and j to access complementary assets at start-up
LCSUij	Previous alliance/imprinting	The tie was present in the firm's i legitimacy/credibility access network at start-up, indicating the selection of partner j at that moment	Dyad	A dichotomous variable denoting whether there is a relation between i and j to access legitimacy/credibility at start-up
INERij	Inertia	The tie is present in the academic and professional trajectory of the entrepreneurial team and in the firm's i innovation network at start-up	Dyad	A dichotomous variable denoting whether a relation from the trajectory was activated to access resources for innovation at start-up

Table 2– Variables definition (cont.)

Variable	Explaining Factor	Description	Level	Construct
Dependent variables capturing tie evaluation				
POCj	Network embeddedness	Partner centrality in the existing sectoral network	Partner	A continuous variable indicating the partner's outdegree centrality (computed with the UCINET software)
NCLIQUESj	Network embeddedness	Existence of indirect ties with the partner	Partner	A continuous variable indicating the number of 2-cliques in which both the firm i and the partner j are present (computed with the UCINET software), excluding the existence of a direct tie
PGEOj	Proximity	Geographical proximity	Partner	A dichotomous variable denoting whether the partner is located in the same country
PFIRMj	Proximity	Organizational/institutional proximity with profit partners	Partner	A dichotomous variable denoting whether the partner is a firm
PUNIVj	Proximity	Organizational/institutional proximity with academic partners	Partner	A dichotomous variable denoting whether the partner is an university/research centre
Control variables				
TMULTSUij	-	Tie intensity at start-up	Dyad	A dichotomous variable denoting whether the tie was mobilized to access more than one resource type at start-up
AGEi	-	Firm's age	Firm	A continuous variable indicating the firm's age in years
SIZEi	-	Firm's size	Firm	A continuous variable indicating the firm's size in terms of employees
FICi	-	Firm´s centrality in the existing sectoral network	Firm	A continuous variable indicating the firm´s outdegree centrality (computed with the UCINET software)

3.2.3 Control variables

Our model controls for the characteristics of the previous dyads, since they may affect the development of relation-specific assets (Kim et al, 2006). Therefore we consider the intensity of the dyad at start-up in terms of its multiplexity (TMULTSU). At start-up, entrepreneurs will tend to choose organizations perceived to offer access to several resources, in the absence of a precise understanding of which resources are best suited for the new company and its growth (Grossman et al, 2012). Thus fewer partners can give access to a variety of resources. This can influence the longevity of the relationship.

Firms' age (AGE) and size (SIZE) are equally included, since they may influence structural inertia (Kim et al, 2006) and also the tendency to activate entrepreneurs' social capital (Hite and Hesterly, 2001).

Finally, the centrality of the firm in the whole network can influence the ability to identify and gain access to partners (Bae and Gargiulo, 2004), as well as lead to the development and accumulation of network capabilities (Foss, 1999) affecting the choice of partners and the survival of the relationship. Therefore, the indegree centrality of the firm in the previously existing network (FIC) is considered in the model. The indegree centrality measures the total number of ties directed towards the firm. Thus a central firm receives resources from several different organizations, being characterized as very attractive.

4. RESULTS

Table 3 reports the results of the rare events logit models for partner selection in the various networks. Model 1 provides estimates of the probability of partner selection, in order to obtain the resources required for innovation. Models 2 to 4 provide estimates of the probability of partner selection in order to access scientific and technological knowledge, complementary assets and credibility/reputation, respectively.

All models provide a good fit to the data. The chi-squared goodness-of-fit test for the change in the –2Loglikelihood value is statistically significant (Model 1: $\chi2(12)= 238.25$, $p < .001$; Model 2: $\chi2(12) = 183.02$, $p < .001$; Model 3: $\chi2(12) = 351.45$, $p < .001$; Model 4: $\chi2(12) = 394.73$, $p < .001$) providing support for acceptance of the models as significant logistic regressions. Furthermore, the overall rate of correct classification is very satisfactory: above 80% for all models. Additionally, observed sensitivity (i.e. the probability of predicting selection when it occurs) and specificity (i.e. the probability of predicting no selection when it does not occur) are high (see Tables A2 in the appendix). Also the sensitivity/specificity analysis performed through the ROC curve reveals the high predictive power of these models (see Figure A2 in the Appendix).

The presence of multicollinearity was verified in two ways: i) by inspection of the correlation matrix and ii) running the corresponding multiple regression models and requesting the collinearity diagnostics. There is no evidence of strong linear relationships between independent variables, and the variance inflation factor (VIF) never exceeds 4, far below the often recommended threshold of 10 (see Tables A3 and A4 in the Appendix).

Results for model 1 show that both persistence and evaluation mechanisms affect the likelihood of tie formation. Regarding persistence, the existence of a prior relation at start-up (INNOVSU) and inertia (INER) increase the probability of selecting a specific partner, while social capital (TRAJ) reduces it. Regarding evaluation mechanisms, network embeddedness, both in terms of partner centrality (POC) and of share of third partners (NCLIQUES), increases the probability of selecting a specific partner, while geographical proximity and the fact that the partner has an academic organizational culture reduce it. Regarding control variables, intensity of the tie at start-up and firm centrality affect positively the probability of tie formation, while firm size reduces it.

Table 3 - Rare event logit models of partner selection

Variable	Model 1 Innovp	Model 2 Knowp	Model 3 cap	Model 4 Pip
TRAJ	-1.457*** (0.490)	-1.363** (0.540)	-0.885* (0.508)	-12.919*** (0661)
INNOVSU	1.164*** (0.257)	-	-	-
KNOWSU	-	1.357*** (0.334)	-	-
CASU	-	-	1.821*** (0.538)	-
LCSU	-	-	-	6.960*** (0.875)
INER	1.350* (0.721)	2.380*** (0.728)	0.475 (0.768)	13.089*** (0.895)
POC	0.321*** (0.069)	0.538*** (0.078)	0.101 (0.092)	0.240 (0.218)
NCLIQUES	0.259*** (0.034)	-0.111*** (0.019)	0.355*** (0.038)	-0.013 (0.057)
PGEO	-0.440*** (0.146)	-1.507*** (0.261)	1.182*** (0.261)	0.056 (0.545)
PACADEMIC	-0.318* (0.190)	0.511** (0.250)	-1.283*** (0.327)	-0.017 (0.711)
PFIRM	0.069 (0.184)	0.202 (0.248)	0.207 (0.235)	0.352 (0.619)
TMULTSU	1.137** (0.461)	0.913* (0.502)	0.034 (0.656)	-0.946 (1.039)
AGE	-0.010 (0.0265)	0.069** (0.031)	-0.165*** (0.052)	-0.242*** (0.094)
SIZE	-0.025*** (0.008)	-0.070*** (0.014)	0.057*** (0.013)	0.008 (0.028)
FIC	0.002* (0.001)	0.011*** (0.002)	-0.017*** (0.003)	0.001 (0.006)
Intercept	-1.501*** (0.259)	-2.911*** (0.344)	-2.231*** (0.432)	-4.920*** (0.995)
N	1936	1936	1936	1936
Log likelihood	-694.10	-551.127	-334.015	-89.447
$\chi^2_{(12)}$	238.25	183.02	351.45	394.73
Pseudo R^2	0.4693	0.2164	0.718	0.652
Correct classification (%)	82.46	88.26	93.69	98.81

Note: numbers in brackets are the robust standard errors; *** $p < 0.01$: ** $p < 0.05$; * $p < 0.1$

Results for model 2 also reveal the relevance of persistence and evaluation mechanisms in the selection of knowledge partners. Comparing with Model 1, and in addition to differences in the magnitude of the coefficients, it is noteworthy the change of sign of the NCLIQUES and of the PACADEMIC variables. For knowledge access purposes, these firms tend to select partners which have an academic culture and with which they share few other partners in the sectoral network.

Results for model 3 indicate the existence of a smaller number of significant explanatory variables for the selection of partners to access complementary assets, although both factors – persistence and evaluation - appear as relevant. Inertia and partner centrality have no effect in partner selection. Comparing with Model 1, we find that geographical proximity (PGEO) now increases the probability of partner selection, indicating the relevance of this factor in the access to complementary assets.

Results for Model 4 show that in the access to legitimacy/credibility neither network embeddedness nor proximity affect partner selection, which is solely driven by persistence. However, the results for the control variables suggest that persistence seems to be less relevant as firm ages, in line with previous research (Lechner et al, 2006).

5. DISCUSSION AND CONCLUSION

This research provides evidence that contributes to on-going debates about the evolution of innovation networks, allowing a more in-depth understanding of the process of partner selection by young knowledge-intensive firms.

Previous research has shown that network building through partner selection involves both the persistence of previous partners and the inclusion of new ones. So, to understand processes of partner selection we have to consider the complementarity between persistence effects and evaluation mechanisms (Li and Rowley, 2002). Therefore, an integrated framework that considers elements of persistence and novelty was developed and tested.

Regarding persistence, three different explanatory factors, suggested by the extant literature, were considered: entrepreneurs' social capital, previous alliance/imprinting and network inertia. Results indicate that firms tend to select organizations they know from previous relations, to access all types of resources necessary for innovation. This result is in line with previous research on alliances (Gulati and Gargiulo, 1999), and with the imprinting literature (Milanov and Fernhaber, 2009). Previous ties seem to help firms in the choice of partners to include in their innovation networks.

Contrary to the arguments of the social capital literature, entrepreneurs' social capital decreases the likelihood of tie formation. This result may be related with the fact that we are considering partner selection at the firms' early growth phase and not at start-up. In fact, previous research has shown that the relevance of entrepreneurs' social capital decays during the process of firm development (Hite and Hesterly, 2001).

However, the positive and significant coefficients for the inertia variable, in line with the findings of previous research (Li and Rowley, 2002); indicate that the combination of social capital with previous alliance has a positive effect on the likelihood of tie formation. This repeated contact allows the development of relation-specific assets and routines that facilitate network building and management processes. Hence, only social capital that was already activated at start-up seems to have a positive role on the probability of a given organization to be selected, namely to provide knowledge and legitimacy/credibility.

Entrepreneurs' social capital seems to have no effect on the likelihood of tie formation in access to complementary assets. This is possibly linked with the more arm's length nature of the relations established to access this type of resource and also to the biotechnology entrepreneurs' predominantly scientific/academic trajectory, which is less useful in accessing non-technological resources (Ensley and Hmieleski, 2005). Conversely, their social capital is particularly useful to access scientific and technological knowledge, as well to provide legitimacy, since the association with reputed research organizations or scientists can have a quality signalling and credibilization effect (Luo et al, 2009).

But, as the construction of networks is not solely based on already known organizations, our framework also considers evaluation mechanisms linked with the choice of new members, namely network embeddedness and proximity between the firm and the partner.

The results show that the existing sectoral network influences the selection of innovation partners. Considering the aggregated innovation network, more central organizations, or organizations with which firms share a partner in the existing sectoral network, have a higher probability of being selected by firms. Therefore, the selection of partners is influenced by information about partners' quality collected through the network, either due to their positioning or to indirect ties.

However the breakdown by resource reveals significant differences in the signal and significance of network embeddedness variables. Centrality has no significant effect in the choice of partners granting access to complementary assets or to legitimacy/credibility. The share of third partners exerts opposite effects on the selection of partners in the case of knowledge (negative) and complementary assets networks (positive). So, results suggest the existence of different mechanisms of selection of partners to access different resources, in terms of network embeddedness, which are not captured in an aggregate analysis.

In the choice of knowledge sources firms prefer central partners with which they share few other partners. This suggests a need to connect to the "best" knowledge sources and to avoid the risks of over-embeddedness (Owen-Smith and Powell, 2004), but also to protect the knowledge from potential leakages (Hurmelinna-Laukkanen and Puumalainen, 2007). Previous research has concluded that these firms access knowledge thought communities (cliques) with strong inner connections and, usually, a single connection to the rest of the network performed by an academic partner (Salavisa et al, 2012).

On the contrary, in choosing partners for accessing complementary assets, companies prefer organizations with which they share a large number of partners. The signal given by the network positioning of the partner is not relevant. Therefore, firms prefer to gather information about these partners through organizations with which they have a direct tie. Thus, clique membership is central in selecting partners to access complementary assets.

The effect of proximity in partner selection also differs between resources. To access knowledge firms prefer foreign academic partners; to access complementary assets they prefer national non-academic (although not necessarily for-profit) organizations. This result confirms that biotechnology firms' access to international academic knowledge is vital to their innovation processes, especially in more peripheral economies (Gilding, 2008, Fontes et al, 2012). It also confirms that the local context is important to provide the complementary assets for the opportunity exploitation (Cooke, 2002).

The selection of partners to achieve legitimacy/credibility is not affected by the network embeddedness variables neither by the proximity variables. This is consistent with the endorsement function played by these partners, which requires previous close interactions and the development of some trust (Shane and Stuart, 2002).

Summing up, the results highlight the relevance of considering an integrated framework that encompasses several explanations for persistence and novelty, which so far were addressed separately. They also uncover different network building strategies in terms of partner selection to access the different types of resources needed for innovation.

The results of this research are globally relevant and increase our understanding of the process of innovation partner selection. Further research will enable us to mitigate some limitations in the specification of the logit model, namely: to account for common-actor effect; to consider the interaction between the variables, since the several mechanisms are closely interwoven; to introduce other forms of proximity described in the literature, namely cognitive proximity; to refine the geographical proximity, considering the actual distance (in Km or travel hours) between the company and each of the partners.

6. REFERENCES

ADOBOR, H. (2006), The role of personal relationships in inter-firm alliances: benefits, dysfunctions, and some suggestions, *Business Horizons*, 49(6): 473–86.

AHUJA, G. (2000), Collaboration networks, structural holes, and innovation: a longitudinal study, *Administrative Science Quarterly*, 45: 425–55.

ANDERSON, A.R., Park, J. and Jack, S. (2007), Entrepreneurial social capital: conceptualizing social capital in new high-tech firms, *International Small Business Journal*, 25: 245–72.

BAE, J. and Gargiulo, M. (2004), Partner substitutability, alliance network structure, and firm profitability in the telecommunication industry, *Academy of Management Journal*, 47: 843-859.

BAUM, J.A.C., Calabrese, T. and Silverman, B.S. (2000), Don't go it alone: alliance network composition and start-ups' performance in Canadian biotechnology, *Strategic Management Journal*, 2: 267–94.

BOSCHMA, R. (2005), Proximity and Innovation: A Critical Assessment, *Regional Studies*, 39:61-74.

BOSCHMA, R. and Frenken, K. (2010), The spatial evolution of innovation networks: a proximity prespective, in: Boschma, R. and Martin, R. (Eds.), *Evolutionary Economic Geography*, Edward Elgar: Cheltenham.

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

BROEKEL, T. and Boschma, R. (2012), Knowledge networks in the Dutch aviation industry: the proximity paradox, *Journal of Economic Geography*, 12(2): 409-433.

BURT, R.S. (1992), *Structural Holes: the social structure of competition*, Harvard University Press: Cambridge, MA.

BURT, R.S. (1997), The contingent value of social capital, *Administrative Science Quarterly*, 42: 339–65.

COHEN, W. and Levinthal D. (1990), Absorptive capacity: a new perspective on learning and innovation, *Administrative Science Quarterly*, 35, 128-152.

COOKE, P. (2002), Regional Innovation Systems: General Findings and Some New Evidence from Biotechnology Clusters, *Journal of Technology Transfer*, 27: 133-145.

COSTA, C., Fontes, M. and Heitor, M.V. (2004), A methodological approach to the marketing process in the biotechnology-based companies, *Industrial Marketing Management*, 33: 403-418.

DEGENNE, A. and Forsé, M. (1994), *Les Réseaux Sociaux: une analyse structurale en sociologie*, Paris: Armand Colin.

DELMAR, F. and Shane, S. (2004), Legitimating first: Organizing activities and the survival of new ventures, *Journal of Business Venturing*, 19, 385-410.

DRUILHE C. and Garnsey E. (2004). Do academic spin-outs differ and does it matter? Journal of Technology Transfer, 29(3-4), 269-285.

EBERS, M. (1999), The dynamics of inter-organizational relationships, *Research in the Sociology of Organizations*, 16: 31-56.

EISENHARDT, K. and K. Schoonhoven (1990), Organizational Growth: Linking Founding Team, Strategy, Environment, and Growth among U.S.Semiconductor Ventures, 1978–1988, *Administrative Science Quarterly*, 40: 84–110.

ENSLEY, M.D. and Hmieleski, K.M. (2005), A comparative study of new venture top management team composition, dynamics and performance between university-based and independent start-ups, *Research Policy*, 34(7): 1091–105.

FONTES, M. (2005), The process of transformation of scientific and technological knowledge into economic value, conducted by biotechnology spin-offs, *Technovation*, 25: 339-347.

FONTES, M., Sousa, C. and Videira, P. (2012), Social networks and the entrepreneurial process in molecular biotechnology in Portugal: from science to industry, in: Salavisa, I. and Fontes (Ed.), *Social Networks*, *Innovation and the Knowledge Economy*, Routledge.

FOSS, N.J. (1999), Networks, capabilities, and competitive advantage, Scandinavian *Journal of Management*, 15(1): 1–15

GILDING, M. (2008), 'The tyranny of distance': Biotechnology networks and clusters in the antipodes, *Research Policy*, 37: 1132–1144.

GLÜCKLER, J. and Armbrüster, T. (2003), Bridging Uncertainty in Management Consulting: The Mechanisms of Trust and Networked Reputation, Organization Studies, 24(2): 269-297.

GROSSMAN et al, (2010), Resource Search, Interpersonal Similarity, and Network Tie Valuation in Nascent Entrepreneurs' Emerging Networks doi: 10.1177/0149206310383693

GULATI, R. (1995a), Does familiarity bredd trust? The implications of repeated ties for contractual choice in alliances, *Academy of Management Journal*, 38(1):85-112.

GULATI, R. (1995b), Social Structure and Alliance Formation Patterns: A Longitudinal Analysis, *Administrative Science Quarterly*, 40: 619-652.

GULATI, R. (1998) Alliances and networks, *Strategic Management Journal*, 19: 293–317.

GULATI, R. and Garguilo, M. (1999), Where do interorganizational networks come from?, *American Journal of Sociology*, 104(5):1439–93.

HALLEN, B.L. (2008), The Causes and Consequences of the Initial Network Positions of New Organizations: From Whom Do Entrepreneurs Receive Investments?, *Administrative Science Quarterly*, 53:685-718.

HITE, J. and Hesterly W. (2001), The Evolution of Firm Networks: From Emergence to Early Growth of the Firm, *Strategic Management Journal*, 22: 275-286.

HSU, D. (2007), Experienced entrepreneurial founders, organizational capital, and venture capital funding, *Research Policy*, 36: 722–41.

HUMAN, S. and Provan, K. (2000), Legitimacy building in the evolution of small-firm multilateral networks: a comparative study of success and demise, *Administrative Science Quarterly*, 45(2): 327–65.

HURMELINNA-LAUKKANEN, P., Puumalainen, K. (2007), The nature and dynamics of appropriability – Strategies for appropriating returns on innovation, *R&D Management*, 37, 95-112.

KIM, T.Y., Oh, H. and Swaminathan, A. (2006), Framing Interorganizational Network Change: a Network Inertia Perspective, *Academy of Management Review*, 31(3):704-720.

KIRKELS, Y. and Duysters, G. (2010), Brokerage in SME networks, *Research Policy*, 39(3):375-385.

KOKA, B.R. and Prescott, J.E. (2002), Strategic alliances as social capital: a multidimensional view, *Strategic Management Journal*, 23: 795–816.

LECHNER, C., Dowling, M. and Welpe, I. (2006), Firm networks and firm development: The role of the relational mix, *Journal of Business Venturing*, 21: 514– 540.

LI, S.X. and Rowley, T.J. (2002), Inertia and evaluation mechanisms in interorganizational partner selection: syndicate formations among U.S. Investment Banks, *Academy of Management Journal*, 45(6): 1104-1119.

LORENZEN, M. (2007), Localised Learning and Social Capital: The Geography Effect in Technological and Institutional Dynamics, *Urban Studies*, 44: 799-817.

LUO, X.R., Koput, W. K. and Powell, W.W. (2009). Intellectual capital or signal? The effects of scientists on alliance formation in knowledge-intensive industries. *Research Policy*, 38(8): 1313-1325.

MARSDEN, P.V. (1981), Introducing influence processes into a system of collective decisions, *American Journal of Sociology*, 86:1203-1235.

MCEVILY, B. and Zaheer, A. (1999), Bridging ties: a source of firm heterogeneity in competitive capabilities, *Strategic Management Journal*, 20:1133-56.

MCPHERSON, M., Smith-Lovin, L. and Cook, J.M. (2001), Birds of a feather: homophily in social networks, *Annual Review of Sociology*, 27:415–44.

MILANOV, H., Fernhaber, S.A. (2009), The impact of early imprinting on the evolution of new venture networks, *Journal of Business Venturing*, 24(1):46-61.

NOOTEBOOM, B. (1999) *Inter-firm Alliances: Analysis and Design*, London: Routledge.

NOOTEBOOM, B. (2000), Learning by Interaction: Absorptive Capacity, Cognitive Distance and Governance, *Journal of Management and Governance*, 4(1-2): 69-92.

NOOTEBOOM, B. (2008) In what sense do firms evolve? *Papers on Economics and Evolution*, No.0812, Max Planck Institute of Economics.

NOOTEBOOM, B., Van Haverbeke, W.P.M., Duijsters, G.M., Gilsing, V.A. and Oord, A. V.d. (2007), Optimal cognitive distance and absorptive capacity, *Research Policy*, 36: 1016-1034.

OWEN-SMITH J. and Powell W. (2004), Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology, *Organization Science* 15: 6-21.

OZMAN, M. (2009), Inter-firm networks and innovation: a survey of literature, *Economics of Innovation and New Technology*, 18(1): 39–67.

POWELL, W., Koput, K. and Smith-Doerr, L. (1996) 'Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology', Administrative Science Quarterly, 41: 116–45.

REUER, J.J., Zollo, M. and Singh, H: (2002), Post-formation dynamics in strategic alliances. Strategic Management Journal, 23: 135-151.

RIVERA, M.T., Soderstrom, S.B. Uzzi, B. (2010), Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms, *Annual Review of Sociology*, 36: 91-115.

ROIJAKKERS, N., Hagedoorn, J. and van Kranenburg, H. (2005), Dual market structures and the likelihood of repeated ties – evidence from pharmaceutical biotechnology, *Research Policy*, 34: 235-245.

SALAVISA, I. and Fontes (Ed.) (2012), *Social Networks, Innovation and the Knowledge Economy*, Routledge.

SALAVISA, I., Sousa, C. and 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: 380-399.

SAMMARRA, A. and Biggiero, L. (2008), Heterogeneity and Specificity of Inter-Firm Knowledge Flows in Innovation Networks, *Journal of Management Studies*, 45: 800–28.

SHANE, S. and Stuart, T. (2002), Organizational endowments and the performance, *Management Science*, 48: 154–70.

SORENSON, T.E. and Stuart, T.E. (2001), Syndication networks and the spatial distribution of venture capital investments, *American Journal of Sociology*, 106(6): 1546-1588.

SOUSA, C. (2012), Using social network analysis to study entrepreneurship: Methodological issues", in: Salavisa, I. e Fontes, M. (eds), *Social networks, Innovation and the Knowledge Economy*, London and NY, Routledge.

SOUSA, C., Videira, P. and Fontes, M. (2011), The role of entrepreneurs' social networks in the creation and early development of biotechnology companies, *International Journal of Entrepreneurship and Small Business*, 12(2): 227–44.

STINCHCOMBE, A.L. (1965), Social structure and organizations, in: March, J. (Ed.), *Handbook of Organizations*, Rand-McNally: Chicago.

STUART, T.E., Ozdemir, S.Z. and Ding, W.W. (2007), Vertical alliance networks: The case of university–biotechnology–pharmaceutical alliance chains, *Research Policy*, 36(4): 477–498.

STUART, T.E. (1998), Network positions and propensities to collaborate: An investigation of strategic alliances formation in a high-technology industry, *Administrative Science Quarterly*, 43: 668-698.

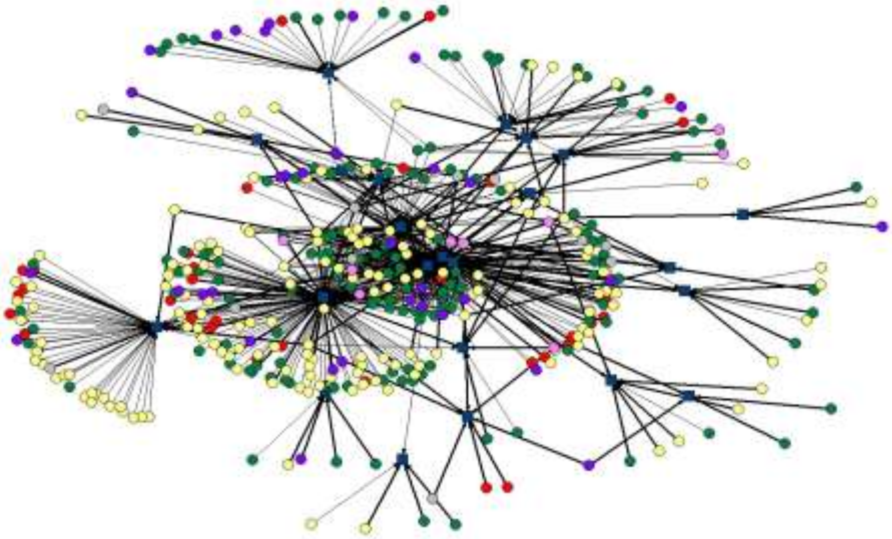
TOMZ, M, King, G. and Zeng, L. (1999), RELOGIT: Rare Events Logistic Regression, Version 1.1 Cambridge, MA: Harvard University, October 1, http://gking.harvard.edu/

UZZI, B. (1997), Social structure and competition in interfirm networks: the paradox of embeddedness, *Administrative Science Quarterly*, 42: 35–67.

WALKER, G., Kogut, B., Shan, W. (1997), Social Capital Structural Holes and the Formation of an Industry Network, *Organization Science*, 8(2):109-125.

7. Appendix

Figure A1 – Portuguese molecular biology sectoral innovation network



Legend: Blue squares – interviewed firms; red circles – biotech firms; green circles – other firms; yellow circles – universities & research centres; pink circles – S&T parks; grey circles – financial institutions; purple circles – other organizations.

Table A1 – Variables descriptive statistics

Variable	N	Mean	Std. Dev.	Min	Max
Innovp	1936	.4085744	.4916973	0	1
Knowp	1936	.1182851	.3230289	0	1
Cap	1936	.3016529	.459094	0	1
Pip	1936	.0294421	.169086	0	1
Traj	1936	.0852273	.2792917	0	1
Innovsu	1936	.0909091	.2875541	0	1
Knowsu	1936	.0470041	.2117024	0	1
Casu	1936	.0480372	.2139001	0	1
Pisu	1936	.0315083	.174732	0	1
Inerinnov	1936	.0206612	.142284	0	1
Poc	1933	2.010347	2.200578	0	10
Ncliques	1936	5.746901	9.504412	0	53
Pgeo	1936	.5779959	.4971353	0	3
Pacademic	1936	.3941116	.4887853	0	1
Pfirm	1936	.3946281	.4888969	0	1
Multsu	1936	.0268595	.1617145	0	1
Age	1936	5.555785	2.742523	3	12
Size	1936	16.92252	9.758701	1	35
Fic	1936	109.3574	56.21035	5	194

Table A2 – Classification tables for logistic models

a) Model 1 - INNOVP

Classified	True D	True ~D	Total
+	513	61	574
-	278	1081	1359
Total	791	1142	1933

Classified + if predicted Pr(D) >= .5
True D defined as innovp != 0

Sensitivity	Pr(+| D)	64.85%
Specificity	Pr(-|~D)	94.66%
Positive predictive value	Pr(D| +)	89.37%
Negative predictive value	Pr(~D| -)	79.54%
False + rate for true ~D	Pr(+|~D)	5.34%
False - rate for true D	Pr(-| D)	35.15%
False + rate for classified +	Pr(~D| +)	10.63%
False - rate for classified -	Pr(D| -)	20.46%
Correctly classified		82.46%

b) Model 2 – KNOWP

Classified	True D	True ~D	Total
+	28	26	54
-	201	1678	1879
Total	229	1704	1933

Classified + if predicted Pr(D) >= .5
True D defined as knowp != 0

Sensitivity	Pr(+| D)	12.23%
Specificity	Pr(-|~D)	98.47%
Positive predictive value	Pr(D| +)	51.85%
Negative predictive value	Pr(~D| -)	89.30%
False + rate for true ~D	Pr(+|~D)	1.53%
False - rate for true D	Pr(-| D)	87.77%
False + rate for classified +	Pr(~D| +)	48.15%
False - rate for classified -	Pr(D| -)	10.70%
Correctly classified		88.26%

c) Model 3 – CAP

Classified	True D	True ~D	Total
+	501	39	540
-	83	1310	1393
Total	584	1349	1933

Classified + if predicted Pr(D) >= .5
True D defined as cap != 0

Sensitivity	Pr(+| D)	85.79%
Specificity	Pr(-|~D)	97.11%
Positive predictive value	Pr(D| +)	92.78%
Negative predictive value	Pr(~D| -)	94.04%
False + rate for true ~D	Pr(+|~D)	2.89%
False - rate for true D	Pr(-| D)	14.21%
False + rate for classified +	Pr(~D| +)	7.22%
False - rate for classified -	Pr(D| -)	5.96%
Correctly classified		93.69%

.

d) Model 4 - LCP

Classified	True D	True ~D	Total
+	47	13	60
-	10	1863	1873
Total	57	1876	1933

Classified + if predicted Pr(D) >= .5
True D defined as pip != 0

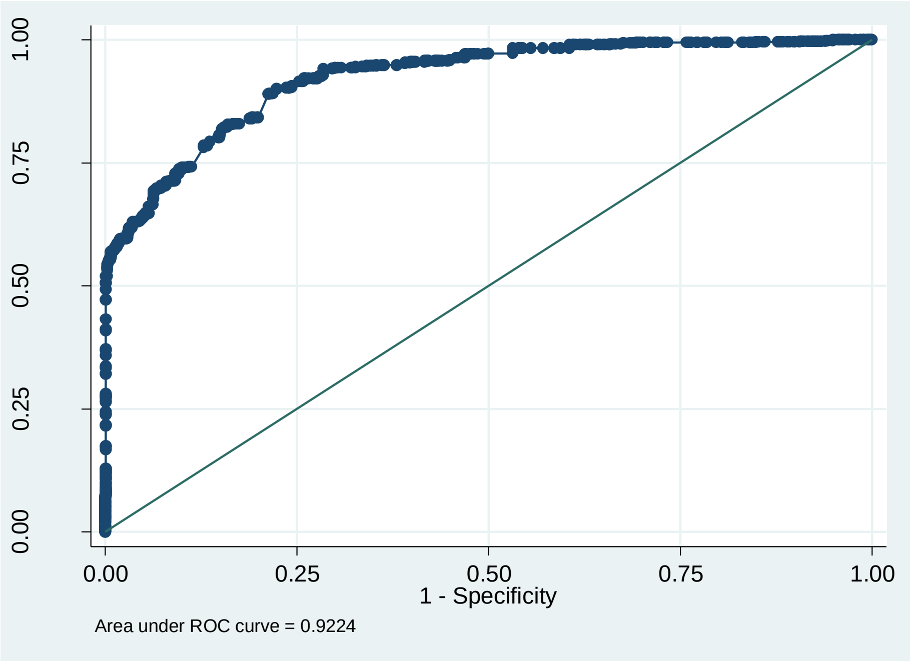
Sensitivity	Pr(+| D)	82.46%
Specificity	Pr(-|~D)	99.31%
Positive predictive value	Pr(D| +)	78.33%
Negative predictive value	Pr(~D| -)	99.47%
False + rate for true ~D	Pr(+|~D)	0.69%
False - rate for true D	Pr(-| D)	17.54%
False + rate for classified +	Pr(~D| +)	21.67%
False - rate for classified -	Pr(D| -)	0.53%
Correctly classified		98.81%

Figure A2 – ROC curves

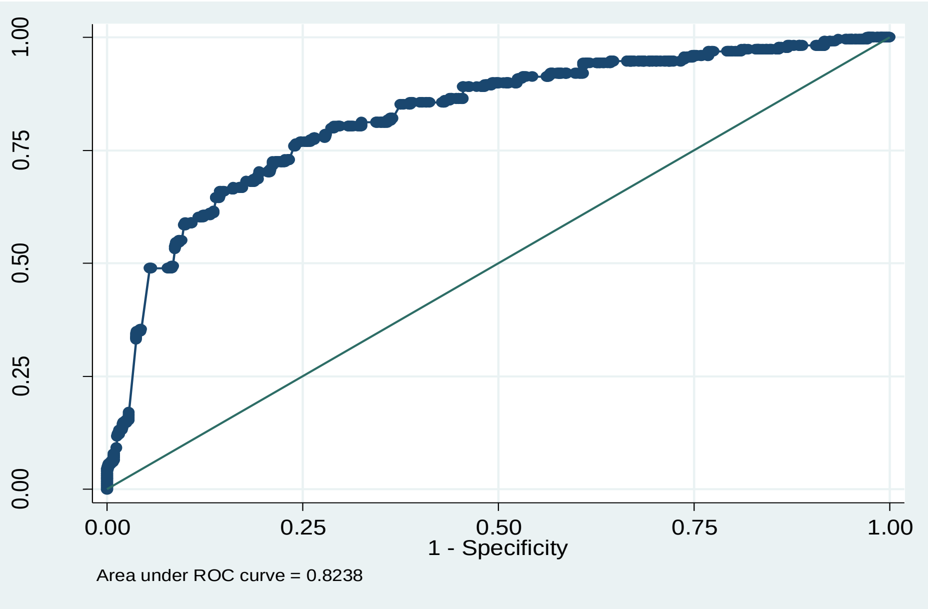
The ROC curve is a graph of sensitivity versus one minus specificity as the cutoff c is varied. Sensitivity is the fraction of observed positive-outcome cases that are correctly classified; specificity is the fraction of observed negative-outcome cases that are correctly classified. When the purpose of the analysis is classification, you must choose a cutoff.

The curve starts at (0; 0), corresponding to c = 1, and continues to (1; 1), corresponding to c = 0. A model with no predictive power would be a 45º line. The greater the predictive power, the more bowed the curve. Hence the area beneath the curve is often used as a measure of the predictive power: a model with no predictive power has area 0.5; a perfect model has area 1.

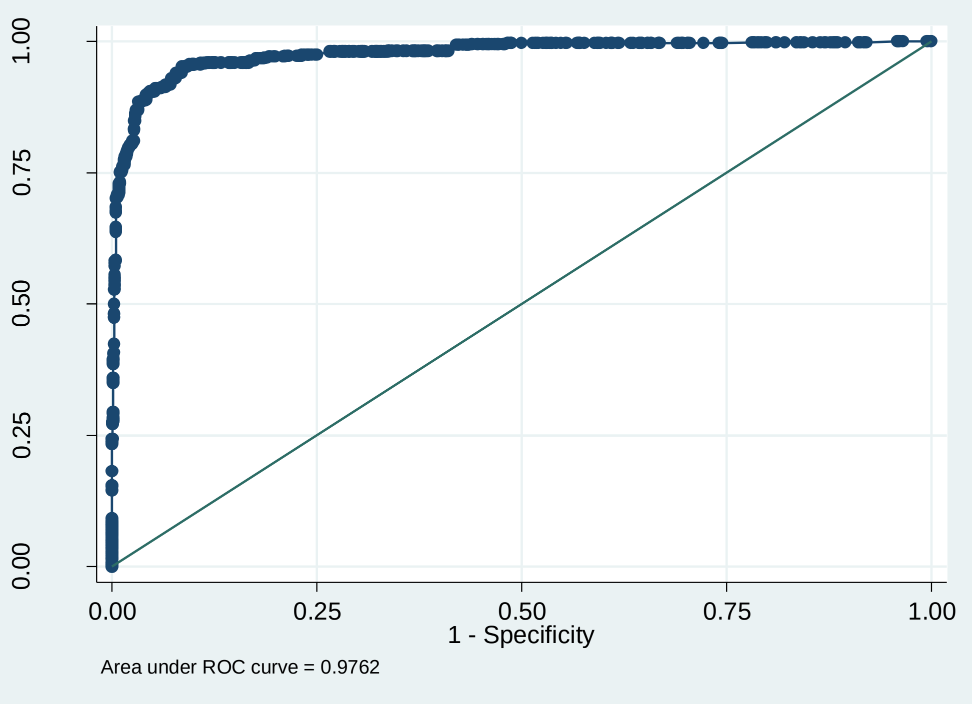
a) Model 1 - INNOVP



b) Model 2 - KNOWP



c) Model 3 - CAP



d) Model 4 – LCP

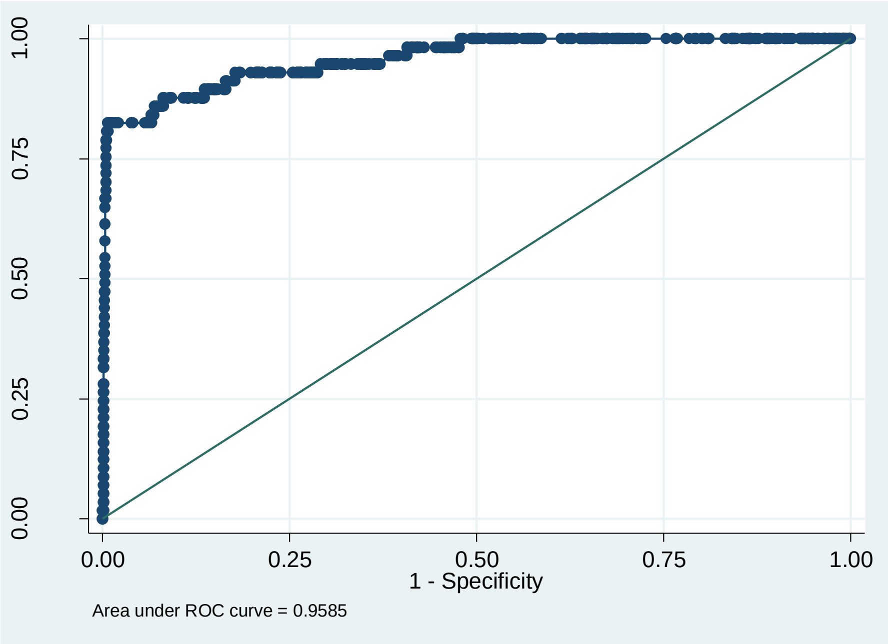


Table A3 Correlations for the independent and dependent variables

innovp	1.00																		
knowp	0.44	1.00																	
cap	0.79	-0.11	1.00																
pip	0.21	0.19	0.07	1.00															
traj	- 0.10	0.03	-0.10	0.14	1.00														
innovsu	0.16	0.21	0.05	0.49	0.16	1.00													
knowsu	0.10	0.26	-0.01	0.25	0.13	0.70	1.00												
casu	0.13	0.07	0.15	0.38	0.18	0.71	0.26	1.00											
pisu	0.16	0.17	0.05	0.79	0.19	0.57	0.28	0.44	1.00										
inerinnov	0.09	0.18	0.05	0.36	0.48	0.46	0.36	0.46	0.45	1.00									
multsu	0.14	0.16	0.10	0.56	0.20	0.53	0.46	0.64	0.74	0.47	1.00								
poc	0.65	-0.00	0.76	0.08	-0.05	0.06	0.05	0.09	0.06	0.11	0.09	1.00							
ncliques	0.64	-0.08	0.78	0.02	-0.09	0.01	0.01	0.03	0.00	0.04	0.03	0.83	1.00						
pgeo	0.32	-0.18	0.51	0.04	0.01	0.03	-0.05	0.12	0.05	0.08	0.08	0.51	0.45	1.00					
pacademic	- 0.14	0.13	-0.22	0.02	0.18	0.04	0.14	-0.04	0.02	0.11	0.02	-0.10	-0.15	-0.27	1.00				
pfirm	0.11	-0.06	0.15	-0.03	-0.16	-0.06	-0.07	-0.03	-0.04	-0.08	-0.03	0.05	0.13	0.09	-0.65	1.00			
age	0.04	0.05	0.00	-0.09	-0.12	-0.06	-0.01	-0.05	-0.08	-0.03	-0.05	0.05	0.07	-0.02	0.00	0.02	1.00		
size	0.09	-0.13	0.18	-0.03	-0.13	-0.13	-0.10	-0.10	-0.03	-0.07	-0.04	0.11	0.22	0.07	-0.07	0.06	0.16	1.00	
fic	0.09	0.09	0.02	-0.05	-0.13	-0.11	-0.06	-0.14	-0.06	-0.05	-0.09	0.03	0.17	-0.08	-0.02	0.04	0.22	0.41	1.00

Table A4 – VIF

	Model			
Independent variable	1 INNOVP	2 KNOWP	3 CAP	4 LCP
Traj	1.39	1.38	1.38	1.38
Innovsu	1.54	-	-	-
Knowsu	-	1.36	-	-
Casu	-	-	1.84	-
Pisu	-	-	-	2.29
Inerinnov	1.78	1.68	1.73	1.68
Poc	3.65	3.65	3.65	3.65
Ncliques	3.58	3.58	3.58	3.58
Pgeo	1.52	1.53	1.52	1.52
Pacademic	1.93	1.95	1.96	1.93
Pfirm	1.80	1.79	1.80	1.80
Multsu	1.53	1.49	1.81	2.34
Age	1.07	1.07	1.07	1.08
Size	1.27	1.27	1.26	1.26
Fic	1.31	1.31	1.32	1.31
Mean VIF	1.86	1.84	1.91	1.98