AN ANALYSIS OF THE BRAZILIAN CO-PATENTS NETWORK BASED ON INVENTORS' GAINS PERCEPTION[1]

Resumo: As invenções cada vez mais formam complexas redes de cooperação entre indivíduos. Embora a literatura tenha explorado estas redes sob diferentes óticas, sua topografia, distribuição espacial e os fatores que motivam sua formação continuam sendo um tema em aberto. O objetivo deste artigo é identificar como os inventores percebem as informações dos parceiros na rede para formar seu portfólio de laços potenciais. Para isto, utiliza-se os dados de patentes do Brasil e os microdados pessoais da RAIS, de 2000 a 2011, para investigar como os inventores criam seu sistema de percepção e, em particular, o que procuram em uma parceria. As principais contribuições são (1) a proposição de um coeficiente de dominância decomposto para examinar a percepção de cada indivíduo com base na hierarquia não-explícita do laço e (2) a identificação dos elementos que os inventores observam e como os usam para formar expectativas de ganhos das parcerias.
Palavras-Chave: parcerias inventivas; percepção de ganhos; redes de copatentes; Brasil

Abstract: Inventions are increasingly becoming a product of cooperation and forming complex networks of co-invention. Although the literature has explored the many faces of these networks, their topography and spatial distribution, the driving factors that form them remain an on-going theme. This article aims to identify how inventors perceive partners' information to build their own portfolio of potential ties. The paper combines Brazil's patent database and personal microdata with confidential information, from 2000 to 2011, to find how inventors build their perception system and in particular what they seek in a partnership. Our main contributions are (1) using a dominance coefficient decomposition to examine the perception of each individual based on the non-explicit hierarchy of the link, and (2) finding that inventors build their perception system based on the expectations of gains from a partnership.

Keywords: inventive partnerships, gains perception, co-patents network, Brazil

JEL: L14; O31; O33.

[1] Os autores agradecem o suporte financeiro do CNPq, FAPEMIG, Capes e à equipe de profissionais do Instituto Nacional de Propriedade Intelectual (INPI), sem os quais esse trabalho não seria possível.

The authors thank the financial support provided by the CNPq, FAPEMIG and Capes and the staff of the National Institute of Intellectual Property (INPI), without whom this work would not be done.

1. INTRODUCTION

The creation and diffusion of knowledge is a complex process with multiple dimensions involved. Thus, the understanding of the process is far from one dimensional but rather a combination of complementary readings. In empirical research, the introduction of new instruments has restored interest in this subject. On the one hand, empirical works throughout the literature have focused mainly on the examination of the spatial determinants of knowledge creation and diffusion. On the other hand, a more recent branch has incorporated elements and metrics of social network analysis in such a way to reflect how interactions between players influence the process.

Marshall (1920 [1890]) proposed the idea that knowledge is in the "air" to explain how co-localization in space benefit the innovation process. He stated, "if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of a further new idea" (p. 156). In fact, Marshall understood that a social space is fundamental to the spillover of new ideas. However, we also must note that the Marshallian air is a product of the co-localization of firms, whose innovation activities can inspire their cohabitants. This idea remained the core of innovation study for a long time until the introduction of new communication technologies began to change the way we create social links. Recent works tend to explain the innovation process as a complex system of social links between inventors, firms and regions (Henderson, 2007; Powell & Gianella, 2010; Fritjar & Rodriguez-Pose, 2016). Thus, knowledge spillover between inventors does not seem to occur only because co-location makes social contact easier. On the other hand, it is the social affinities that enable connections, and co-location is a process by which agents diminish their spillover costs. Therefore, while the former innovation literature was widely influenced by how location influences regional performance, the latter tries to understand why and how inventors find and cooperate with each other. Cassi and Plunket (2015) refer to the *proximity perspective* and the *relational perspective*.

In this paper we attempt to analyze how inventors connect to each other in Brazil. We are interested in two main questions. First, what do inventors seek in a partnership? Second, what are the perceptions they form before connecting? Although the first question is a more traditional objective to pursue, the second one is not very common. From a microeconomic viewpoint, inventors are optimizing players with intrinsic costs of inventing and cooperating and a bundle of types of returns. An inventor may apply him/herself in cooperation to strengthen their network, to have better reputation, to share investment costs, to access new information and, of course, for financial gain. On the other hand, an inventor might invent alone in a slow but highly profitable process. Thus, the optimization mechanism would drive the inventor to dedicate some amount of their time to cooperation.

It is important to note that the cost of cooperation has been decreasing due to the invention of low-cost communication technologies, which has made social interaction more common (Castells, 1996). This can help inventors to increase cooperation and form inventing teams with strong bonds within. However, cooperation can also involve issues of uncertainty and trustworthiness.

We address our two questions by applying the proximity perspective proposed by Torre and Gilly (1999) which tries to elaborate the reasons that bring two nodes closers. However, most inventors have information asymmetry and do not have fully knowledge the potential partner's costs, expectations and reputation. They form their own expectations based on the perception given by available information. Cowan, Jonard and Zimmermann (2006, p. 156) point out that "the repetition of interactions creates information that permits partners to reduce uncertainty and increase predictability regarding each others' behavior". Hence, as inventors tighten and select their ties, the more efficient their connections become and the more effective their perception formation.

The aim of the paper is to answer both questions above by analyzing inventors' personal microdata and network patterns within the Brazilian context. We use all Brazilian co-patenting data in all technological fields between 2000 and 2011, and only inventors with formal employment registered in the Brazilian Ministry of Employment.

Methodologically, the paper uses social networks methods combined with traditional econometric modelling. To capture a node's role in a tie, we breakdown the model to identify non-explicit hierarchy of dominance, based on the proximity analysis of Torre and Gilly (1999). We algebraically develop a way to introduce node features into interactions and call this technique "dominance coefficient decomposition". This metric, as will be shown later, is similar to the gravitational computation of characteristics, although our coefficient is better optimized for social networks rather than spatial analysis.

We can briefly summarise our results into two types. First, (i) there are high spatial boundaries in Brazilian inventive networks and (ii) the inventors and their networks tend to form clusters around central nodes that act as gatekeepers to their sub-systems. From the perception formation point of view, (iii) the partner's gender matters, (iv) individuals seek individuals with a higher educational level and salary than their own, although (v) bilateral matching tends to approximate individuals with similar levels. We have evidence to believe that individuals form their connections based on their own perceptions of what a "good tie" is, which may seem a result of boundary rationality. Moreover, the dominance coefficient fulfilled our expectations to mimicry hierarchical dominance in the model.

The paper is organized as follows. The next section introduces the basis of collaboration theory in the study of innovation. The third section presents the core of the node's gain perception theory. The fourth section is an overview of both international and Brazilian empirical works on innovation. The fifth section presents data and methodology, in which the "dominance coefficient decomposition" is algebraically demonstrated. Then, we show and discuss our empirical results. The final section presents concluding remarks.

2. THE THEORETICAL BASIS OF COLLABORATION

The motives that lead to cooperative behaviour are widely debated in innovation literature. Dachs *et al* (2008) argue that traditional firm-based views of innovation process lack an explanation of why (and how) firms engage in cooperation activities. The relational view, they point out, "focused on inter-firm relationship as a source of competitive advantages" (p. 202). De Faria et al (2010) found in the literature that firms look for knowledge beyond their own boundaries to access complementary capabilities and technological resources. In a similar way, cooperation activities can occur at varying levels rather than between firms, such as regions, scientific teams, universities or – as this paper will explore – individuals.

The polymath "garage inventor" is now a rarity. Most inventors are highly qualified workers that master a small set of specific knowledge and have to become part of a group in order to access complementary capabilities from other highly qualified inventors. In fact, scientific team size (Wuchy et al, 2007) and academic co-authorship (Hicks & Katz, 1996; Adams et al, 2005) are growing. The cooperative behaviour of inventors is thus enlightened and the question "why do they cooperate?" precedes the question "what do they seek?"

We can divide cooperation activities into three sequential steps: (i) the first contact, (ii) the cooperation itself and (iii) the result. Whether the first contact is a result of active searching or an occasional meeting (e.g. college colleagues) is difficult to capture. In general, collaboration occurs within communities that provide match opportunities (Giuri & Mariani, 2013). However, when facing new problems, inventors often search for partnerships outside their own surroundings, leading to inter-business and inter-regional collaboration.

The second step, cooperation itself, is the exchange of knowledge. In current data, it is not possible to know for sure the knowledge stock of each inventor, and what they cede and what they absorb. While also difficult to measure, it is possible to establish some premises. Giuri and Mariani (2013) identified that education, for example, enhances knowledge spillover and absorption. Since inventors cooperate to access new cognitions, and it is supposed that higher educational levels implies higher knowledge stocks, then the spillover tends to be stronger from the more educated inventor to the less educated one. The same should occur in other knowledge-stocking characteristics of the inventor. Whittington and Smith-Doerr (2008), Jung and Ejermo (2014) and Crescenzi, Nathan and Rodriguez-Pose (2016) identified that individuals use available public information to create their perception and to convert potential ties into formed ties. This information could be, for example, the partner's gender, ethnicity, language, job position or technological field. Thus, as individuals search for a group of "good" characteristics, they tend to form smaller groups of potential inventors, and convert ties inside these groups. Whittington and Smith-Doerr (2008) found that intragroup connections are much more common than inter-group connections.

The last step is the product of collaboration and this we can measure. It can be scientific papers, products placed in the market or patent applications. There are advantages of using patents as a proxy for innovation. Nagaoka, Motohashi and Goto (2010) argue that patents are highly valued inventions with economic relevance from a reliable dataset of easy access. Moreover, Ter Wal and Boschma (2008) point out that patent data is also easy to transform into a collaboration network. In fact, many recent studies of innovation networks use patent data in different aggregations as a proxy for collaboration, such as Bresci and Lissoni (2004, 2005), Ejermo and Karlsson (2006), Maggioni, Nosveli and Uberti (2007), Cassi and Plunket (2015), Morescalchi et al. (2015) and Crescenzi, Nathan and Rodriguez-Pose (2016). In this paper, we use co-patent data to identify a previous step of formed tie.

Proximity analysis as developed by Torre and Gilly (1999) and Boschma (2005) – in innovation science – is useful as an analytical resource to expand the traditional regional view. From this perspective, innovation is a combination of five different dimensions of proximity: cognitive, organizational, social, institutional and geographical. Despite the identification of this multi-dimensional perspective of innovation, this reading is more about how to understand innovation process and less about how to build a cognitive model of how inventors understand each other. In this paper, we use the concept of perception from game theory and cognitive sciences. It slightly shifts the question from "what do inventors seek in a partnership" to "what the inventors look for when they seek partnership", without major changes to our objectives, however. We present this concept in the following section.

3. THE CONCEPT OF NODE PERCEPTION

While from the regional perspective the gravitational forces of regional capabilities is one of the main causes of cooperation, from the individual inventors' perspective it is how people see each other as potential partners. Owan and Nagaoka (2011) identified two types of motives that make inventors more or less susceptible to cooperation, namely intrinsic and extrinsic. "Intrinsically motivated behaviors are behaviors which a person engages in to feel competent and self-determining" (p. 3), and "extrinsic motives such as career concerns, the desire to enhance their reputations inside and outside their organizations, and the expectation that their performance will affect their research funding and compensation" (p. 3). Thus, the inventor tends to become more cooperative if he or she perceives that the relationship will convert into personal gains. In other words, the portfolio of potential partners is built upon the inventor's perception of gains. However, what influences this perception?

The Oxford Dictionary describes *perception* as "the way in which something is regarded, understood, or interpreted" and Kahnemann (2002) describes it as the process of fast, automatic, associative and slow-learning thinking. In other words, perception is a personal reading of information after filtering based on previous knowledge, experiences or prejudices.

Assuming that inventors have limited time and choose to split this time between working on their own projects and cooperating with others, there are costs of working alone (e.g. slower production) and of cooperation (e.g. attend to social meetings to find partners). If they see that, the gain of working alone is greater than when cooperating, they will not waste their time on it. However, if there is the perception that cooperation leads to greater gains, they will choose a partner in their portfolios. Hence, the inventor's perception of relationship gains translates what each inventors takes into account when choosing partners in their portfolios. This choosing process is less reasoning than the previous one, by which the inventor choose between working alone or cooperating with others. First, individuals cannot access more than a few publicly available information about a partner, such as his/her age and business field. However, personal data as salary is usually confidential. The inventor uses general information - such as educational level, social status, wealth and productivity – to create a perception based upon general information, personal beliefs and common sense. If the particular information adapts to the general perception, the probability of cooperation is expected to be high. Hence, the question "how do inventors cooperate?" can be reformulated to "*how do inventors see each other and what gains do they expect?*".

4. EMPIRICAL RESEARCHES ON INVENTOR COLLABORATION

As mentioned above, only few works use the individual level to analyze the determinants of collaboration. Firstly, because personal data is difficult to access and secondly, regional-based and firm-based studies are more common. Thirdly, the number of individuals in the network is much harder to work when compared to regions and even firms. Below, we review some concepts mentioned in empirical literature on collaboration. These differ in level of aggregation, but they have similar interpretations if we extrapolate them to the individual level.

- **Geography matters when forming new connections**: At the regional level, Morescalchi et al. (2015) and Scherngell and Barber (2008) found constraints of physical distance within European Union borders. In a study on the internationalization of inventive activity, Picci (2010) identified that a common border has a positive effect on bilateral collaboration. Moreover, in Sweden Ejermo and Karlsson (2006) found regions that the R&D resources of neighboring regions play a positive role in network affinity. At the firm level, Paier and Schengell's (2010) results indicate that geography matters in firm collaboration in the EU.

- **Proximities can be substitutable:** At the inventor ties level, Cassi and Plunket (2015) found that different forms of proximity can be substitutable. Once a connection exists, social proximity overlaps geography and technology in the probability of forming new ties.

- **Regional development level is important:** Montobbio and Sterzi (2013) studied the internationalization of co-patent activities in different countries with different development levels using GDP *per capita* and institutional maturity of the property rights rules as development measurements. They found that institution maturity and development are highly correlated, and countries with less mature institutions tend to cooperate less with other countries. In particular, we are not concerned with the internationalization of Brazilian innovation activities, but with the internal distribution of it. However, Brazilian regions have high heterogeneity in development levels, and regional features can constrain individuals within a less-developed region hampering the task of seeking others in the network.

- **Individual characteristics of inventors:** Whittington and Smith-Doerr (2008) identified that inventors that share similarities tend to form sub-groups in the network and connect within these groups. Crescenzi, Nathan and Rodriguez-Pose, for example, identified ethnic-linguist groups, while Jung and Ejermo (2007), Whittington and Smith-Doerr (2008) and Gaughan and Bozeman (2016) found that the partner's gender is an important factor in the forming of groups.

Few works focus on inventors, and even fewer on their personal features. First, the access to the microdata of patents is restricted. Second, access to complementary and indispensable data – such as salaries and residence – is even more restricted due to confidentiality. Because of this, most works cited above carried out topological research of networks rather than a study on the individual characteristics of the inventor, or used a higher aggregation level to access other characteristics of nodes.

Using patent data, Agrawal et al. (2008) – patent citation - and Nathan and Rodriguez-Pose (2016) – co-patenting – explore the inventor level in the network. Both papers access personal data in some nodes in cognitive and institutional terms. In our paper, we include a social-economic level of personal data, regarding information about educational level, wage and residence of the inventors.

The Brazilian Innovation Network

The Brazilian innovation system is immature and in process of formation (Albuquerque, 1996; Gonçalves and Almeida, 2009). According to Albuquerque (2000) in Brazil, there is a high proportion of small firms and single individual inventions, and a small participation of domestic firms. Moreover, Albuquerque *et al* (2002), Lemos *et al* (2006) and Gonçalves (2007) found that innovation activity is highly concentrated in the Southeast region, especially around innovation centers such as São Paulo, Rio de Janeiro, Campinas and Brasília. In contrast, peripheral regions in the North and Northeast regions show low numbers of patents *per capita* (Gonçalves, 2007). These findings demonstrate that geographic inequality in the Brazilian Innovation System is a mirror of the country's historic development, which has created high economic inequality between the center-southern regions and the northern regions.

The previous works focus on the regional perspective rather than innovation network. The difficulty of accessing reliable databases was a constraint to this kind of study until recently, when some institutions made efforts to build and simplify access to certain information. Sidone, Haddad and Mena-Chalco (2016a, 2016b) used information from academic résumés available on the Lattes Platform[2] to build a scientific network of Brazilian researchers. They found that distance is a key element in determining the network formation, and that less-prestigious scientific regions are becoming more important in the network.

Our paper contributes to this debate in two ways. First, we use the individually identified patent database to build the network. Second, we focus on the inventor rather than the region or the firm. However, our work analysis and conclusions are not localized, but can be understood as a contribution to the study of innovation network formation in emerging countries.

5. DATA AND METHODOLOGY

Data

In this paper, two databases are used. First, the database thankfully ceded by the Brazilian National Institution of Intellectual Property (INPI), which contains information on patent application, applicants and inventors. Second, the Annual Report on Social Information (in Brazilian Portuguese: RAIS), available from the Brazilian Labour and Employment Ministry (MTS), which contains the personal data of formal employees. Both databases uses the Natural Persons Register (in Brazilian Portuguese: CPF), an individual taxpayer registry identification number as an impersonal identity of inventors and employees. The time period used is from 2000 to 2011, divided into 4 equal 3 year periods.

The advantage of using this is the access it provides the social-economic classified data of each inventor, such as educational level, salary, residential address, and also employers' data, such as firm size and business type. This type of data – as far as we know – has not previously been used in a less developed country study. Mainly due to the fact that it is difficult to obtain reliable database sources with a workable time-series. On the other hand, the use of a formal employment database excludes "garage inventors", those who are not formally employed in companies, government organizations, research facilities or universities. Moreover, international cooperation with Brazilian inventors is underestimated if foreign partners are not formally employed in Brazilian firms/institutions. Neither pose a significant problem though. First, there are few garage inventors within the network and second, the main target of this paper is to investigate how Brazilians cooperate.

The untreated INPI's database relates patents to inventors in a long shape in which each line correspond to a pair patent x inventor. With this information, it is possible to reshape the database to create a matrix with size *n x p*, where *n* is the number of unique inventors and *p* is the number of unique patents. The matrix multiplication of $M_{nxp}M_{pxn}^{T}$ results in a new matrix W_{nxn}, where W is a matrix that describes partnerships among the network. For each $w_{ij} \in W$, if w_{ij} is zero, then there is no formal evidence of cooperation between i and j. Otherwise, if $w_{ij} > 0$, then *i* and *j* cooperated. The value of each *w* counts the number of times that two nodes cooperated in the period.

W_{nxn} can be rearranged into a new workable databases of three variables: (i) identification of inventor, (ii) identification of co-inventor, and (iii) number of cooperation. To avoid double counting, the identification order was sorted using inventor's CPF number. Thus, we called the greater inventor's CPF number as inventor 1, and the lower as inventor 2 (the co-inventor). This classification does not intend to represent any hierarchy though.

[2] Lattes is a public provided platform maintained by CNPq (National Council for Scientific and Technological Development, in English) to manage information related to research (http://lattes.cnpq.br/).

From the RAIS database, three sets of characteristics were selected: (i) personal, (ii) occupational and (iii) geographic. Table 1 classifies the variables chosen. Inventors with an incomplete set of data were excluded.

Table 1. Classification of data from RAIS' database

Type	Data
Personal	Gender
	Educational Level
Occupational	Salary
	Business Field
Geographic	City

The variables Gender and Business Field were included as a *dummy* of equality. If two nodes in a tie are both from the same gender, then the variable has a value of 1, otherwise 0. The Business Field data uses the National Classification of Economic Activities (in Brazilian Portuguese: CNAE), based on the International Standard Industrial Classification of Economic Activities – ISIC. If the 3-digit code were equal between nodes, then the *business* dummy variable has a value of 1, otherwise 0. The *City* data used Global Positioning System – GPS – coordinates to find the centroid of residence city. The variable *distance* is calculated using Haversine formula of geo-distance. Educational Level and Salary uses the methodology presented in the next section.

Hierarchical dominance coefficient decomposition

Nagaoka (2011) points out that literature treats a patent-generating partnership as an invention with higher commercial value. Hence, a tie – or a partnership cooperation – exists if two inventors co-participate in a patent. The network formed is bilateral, weighted and non-hierarchical. Although non-hierarchical, information is thought to have a common direction in the network. For example, inventors with a higher educational level transmit knowledge to their partner who has a lower level. Therefore, the model is based on the premise that there is information dominance – or, at least, equality – in the relationship, even if there is no observed hierarchy.

In a tie *(i, j)* information may be split into two parts: who gives and who gains. Similar to gravitational models, the relational forces between actors define attraction forces that make connections stronger or weaker. These forces can be described as a combination of differences and sums or means. Assume a simple model where variable *y* is determined by the variable *x* of *i* and *j*. Then:

$$y_{ij} = \beta_0 + \beta_1 |x_i - x_j| + \beta_2 (x_i + x_j).\frac{1}{2}$$

Where the first term, multiplying β_1, is the absolute difference of the observations of *x* in both nodes. The second term, multiplying β_2, is the mean of *x* in the tie. By the triangular inequality:

$$|z_i| - |z_j| \leq |z_i - z_j| \leq |z_i| + |z_j|$$

If $x_i, x_j \geq 0 \ \forall i, j \in V$ and $z_i = \max(x_i, x_j)$ and $z_j = \min(x_i, x_j)$, then

$$0 \leq \max(x_i, x_j) - \min(x_i, x_j) = |z_i - z_j| \leq \max(x_i, x_j) + \min(x_i, x_j)$$
$$|z_i - z_j| = \max(x_i, x_j) - \min(x_i, x_j) \geq 0$$

Therefore

$$y_{ij} = \beta_0 + \beta_1(\max(x_i, x_j) - \min(x_i, x_j)) + \beta_2(\max(x_i, x_j) + \min(x_i, x_j)).\frac{1}{2}$$

After some manipulation

$$y_{ij} = \beta_0 + \left(\frac{\beta_2}{2} + \beta_1\right)\max(x_i, x_j) + \left(\frac{\beta_2}{2} - \beta_1\right)\min(x_i, x_j)$$

Call $\gamma_1 = \left(\frac{\beta_2}{2} + \beta_1\right)$ and $\gamma_2 = \left(\frac{\beta_2}{2} - \beta_1\right)$. The coefficients γ are individual values for the node in the dominant position – the max – and the one in dominated position – the min. Table 2 analyzes the influence of β direction on the γ sign.

Table 2. Relationship of signs of Beta and Gamma

Coefficient	$\beta_1 > 0$	$\beta_1 < 0$
$\beta_2 > 0$	$\gamma_1 \geq 0$ $\gamma_2 \lesseqqgtr 0$	$\gamma_1 \lesseqqgtr 0$ $\gamma_2 \geq 0$
$\beta_2 < 0$	$\gamma_1 \lesseqqgtr 0$ $\gamma_2 \lesseqqgtr 0$	$\gamma_1 \leq 0$ $\gamma_2 \lesseqqgtr 0$

Quantitative variables such as salary and educational level will use this method. It enables not only the analysis of variable influence on tie connectivity but also how the position of the node influences it.

Contrafactual Building

The probability of two inventors becoming partners is directly related to previous contacts – such as co-working, co-studying, or having mutual friends. In a network with 7,733 inventors, it is highly improbable that a single node has a potential partnership with other 7,732 inventors. If that were the case, the network would have 29,895,778 potential collaborations. However, the amount of time an inventor has to dedicate to collaboration is limited, and the time available for social meeting and other personal contacts is limited. It is reasonable to limit the potential connections one can have. A second network with all potential ties – where $w_{ij} = 0$ – will work fine as pseudo-counterfactual for the real connections.

The literature suggests methods of network counterfactual building (see King & Zeng, 2001; Sorenson, Rivkin & Fleming, 2006; Chandrasekhar, 2015). In common, the strategy adopted is to select case-controlled ties randomly with $w_{ij} = 0$.

Some works, such as that of Cassi and Plunket (2014), use a fixed number of case-controlled ties. This paper adopts a fixed proportion in case-controlled ties selection based on the Erdös-Renyi Model (Erdös & Renyi, 1959). This method has some advantages due to its good behavior (Chandresekhar, 2015), since (i) the number of ties expecting from a node is $(n-1)p$, (ii) the probability that two neighbors of i , j and k forms a tie is p, and (iii) the probability that i, j and k are mutually connected is p^3.

In this paper, the probability of tie connection in ERM is 0.05. Table 3 describes the resulting network after combining ERM and *de facto* network. On average, 2.2% of ties were formed. The network seems to be a case of a rare event due its low rate of formed connections. King and Zeng (2001) define *rare event* as the cases in which the number of occurred events is much higher than the number of non-occurred, and the conditional variance is much higher than the conditional mean. These cases require models that correct the high proportion of zero values. Santos Silva and Tenreyo (2011a) suggest the Poisson Pseudo-Maximum Likelihood (PPML) estimator.

Table 3. Distribution of formed ties *versus* counterfactual ERM ties, by triennials

		2000-2002	2003-2005	2006-2008	2009-2011	2000-2011
(1)	Formed Ties	2,764	1,999	4,766	5,150	14,679
(2)	ERM Ties	68,198	109,644	209,144	266,150	653,136
(1)+(2)	Total Ties	70,962	111643	213,910	271,300	667,815
	% Formed	3.90%	1.79%	2.23%	1.90%	2.20%

Estimation Method

Santos Silva and Tenreyo (2011a) and Fally (2015) consider PPML estimator to be consistent to treat rare event problem, Cameron and Trivedi (2013) argues that estimators based on a negative binomial distribution are enough. This paper does not intend to enter the debate of what estimator is desirable in a rare event case, thus both PPML and Negative Binomial estimators are used. The PPML estimator is based on the Poisson regression with a density distribution function (Cameron & Trivedi, 2013)

$$f(y_i|x_i) = \frac{e^{-\mu}\mu_i^{y_i}}{y_i!}, y_i = 0,1,2, \ldots$$

On the other hand, PPML estimators use an interaction mechanism based on the Generalized Linear Model, which can lead to convergence issues (Santos Silva & Tenreyo, 2006, 2011a). Consequently, Santos Silva and Tenreyo (2010, 2011b) suggest using subsamples of the issued variable. In a network estimation the use of subsamples can exclude important nodes and ties, however. Therefore, we treat any convergence issue excluding the entire variable from the model.

Empirical Model

The proposed model aims to find variables that explain connections in Brazilian inventor networks by combining network and individual characteristics. The results will help to understand how inventors' gain perception works in a situation in which there are no explicit hierarchies. We divide variables into four types: (i) absolute difference, (ii) mean value, (iii) control dummies, (iv) linear. The first two types describe the implicit hierarchy of knowledge flowing within the tie. The third type is an explicit equality variable that assumes the value of 1 if the two nodes of a tie have the same value, and 0 otherwise. The fourth type is used to measure the distance between the nodes. The dependent variable is the number of observed connections. We describe independent variables in Table 4.

Table 4. Independent variables, expected signs and description

Variable	Expected Sign	Description
Distance	-	Physical distance between two inventors, in kilometers, using the centroid of the residence city.
Centrality Difference	-	Absolute difference of proximity centrality between the inventors. The proximity centrality of a tie in a network measures how easily a node can be accessed due to its connections
Educational Level Difference	-	Absolute difference of educational level observed between the inventors.
Wage Difference	-	Absolute salary difference observed among the inventors measured in the number of Brazilian National Minimum Wages.
Educational Level Average	+	Average of educational level of the nodes within the tie.
Wage Average	+	Average salary of the nodes within the tie measured in Brazilian National Minimum Wage.
Same Gender	+	Tie is formed between inventors of the same gender.
Same City	+	Both inventors live in the same city.
Same Business	+	Both inventors work in companies in the same business field.
Connection between University and Company	?	Tie is formed between academic inventor (university) and a company inventor.
Triennials	?	Dummies to indicate triennial of connection: T1, T2 or T3. T4 is time-control variable.

ESTIMATION AND RESULTS

We show results of estimation in Table 5 by the model used in columns, PPML or BNMR. There are two estimated models for each, one including triennial dummies. Overall, the results of all the models are very similar in sign, magnitude and significance.

The coefficient of geographic distance between the nodes is negative and significant, as expected. This means that two close nodes are more likely to connect than two distant ones. Literature calls this phenomenon "spatial decay" and its presence and effects are well documented in Jaffe (1989), Audretsch and Feldman (1996) and Henderson (2007). In particular, Gonçalves and Almeida (2007) present evidence of spatial decay in the distribution of patents in Brazil. Likewise, Cassi and Plunket (2015) show that geographical distance matters in tie formation, although it is more relevant for new connections.

Table 5. Results of PPML and BNMR estimators

Variables	BNRM	PPML	BNMR¹	PPML¹
Geographical Distance	-0.0009***	-0.0009***	-0.000919***	-0.0009***
	(2.97e-05)	(4.80e-05)	(2.97e-05)	(4.77e-05)
(D) Centrality	-2.194***	-2.195***	-2.225***	-2.215***
	(0.0415)	(0.0579)	(0.0424)	(0.0603)
(D) Education	-0.143***	-0.140***	-0.138***	-0.137***
	(0.00506)	(0.0066)	(0.00512)	(0.0064)
(D) Salary	-0.0118***	-0.0119***	-0.0130***	-0.0133***
	(0.000795)	(0.0012)	(0.000812)	(0.0012)
(A) Salary	-0.0087***	-0.0079**	-0.00655***	-0.00531
	(0.000809)	(0.00399)	(0.000861)	(0.0041)
(A) Education	0.0232***	0.0195**	0.0267***	0.0192**
	(0.0058)	(0.00774)	(0.00610)	(0.0086)
Same Gender	0.199***	0.168***	0.181***	0.149***
	(0.0162)	(0.0172)	(0.0164)	(0.0172)
Same City	2.263***	2.255***	2.287***	2.275***
	(0.0230)	(0.0289)	(0.0230)	(0.0287)
Same Business	1.601***	1.607***	1.603***	1.606***
	(0.0192)	(0.0258)	(0.0192)	(0.0252)
University-Company	0.479***	0.484***	0.513***	0.521***
	(0.0185)	(0.0209)	(0.0188)	(0.0218)
Constant	-4.394***	-4.334***	-4.527***	-4.386***
	(0.0995)	(0.114)	(0.106)	(0.117)
Observations	667.815	667.815	667.815	667.815
R^2		0.262		0.264

(1) Controlled by 3-year-period dummies. Results omitted.

Note: (D) denotes "Difference" treatment, and (A) denotes "Average" treatment.

The variable *same* city has a positive coefficient, evidence of agglomeration of social structures within boundaries. Saxenian (1994) and Granovetter (1985) identify an important feature in social clusters in the development of creative regions. In Brazil the network is clustered in the central-southeast-south polygon, while peripheral regions of the North and Northeast (the poorest and least productive states) have little integration with the main network.

The variable *proximity centrality difference* has a negative and significant coefficient. If two inventors are similar in proximity centrality – or have similar ease in connecting – the tie connectivity is higher than the one formed between nodes with a great centrality gap. In perception terms, a node can see as the popularity of a potential partner as an entry point to new partnership. On the other hand, inventors with high centrality do not have many incentives to connect to nodes with much lower centrality. Consequently, inventors seek partners with similar popularity to access new potential partners. This interpretation is close to Cassi and Plunket's (2015) "number of partners in common" variable for France, and Crescenzi, Nathan and Rodríguez-Pose's (2016) "social proximity" variable for the United Kingdom.

The wage and educational level variables are separated in two terms: absolute difference and tie average. For the variable wage, both difference and average is negative and significant. The sign for average is different from that expected. Brazilian workers with higher salaries occupy better hierarchical positions. These are typically managers and are less inventive. Therefore, there is less propensity for inventors to connect with such people. On the other hand, educational level coefficients have expected signs, negative for difference and positive for average. Inventors tend to connect to others with a similar educational level, and there is a connectivity gain if the combined educational level is high. This result confirms for Brazil what Giuri and Mariani (2013) verified in EU.

Table 6 shows the calculated dominance coefficient. Both educational level and salary coefficients show a negative impact for the dominant position. This indicates that nodes with higher values on these attributes perceive losses from connecting to others with lower values. On the other hand, a dominated position has a positive impact on education and salary. We can understand that nodes with lower levels perceive higher gains when connecting with more educated and richer nodes. This result shows that an inventor with lower levels of education and salary – the dominated position – is more likely to connect with individuals of a higher status. From the gains perception viewpoint, inventors seek partners with higher status, since the return on the partnership is also higher. On the other hand, inventors with a higher status have a negative return when connecting with lower status inventors. The combination of these effects shows that even though inventors seek potential partners of a higher status, the connection is most likely to occur between inventors with similar status.

Table 6 – Hierarchical Dominance Coefficient Decomposition of the Negative Binomial Regression Model Results

	$\beta > 0$	$\beta < 0$
$\alpha > 0$	-	$\gamma_{1_{BNMR}} = -0.1314$ $\gamma_{1_{PPML}} = -0.1303$ $\gamma_{2_{BNMR}} = 0.1546$ $\gamma_{2_{PPML}} = 0.14975$
$\alpha < 0$	-	$\omega_{1_{BNMR}} = -0.0162$ $\omega_{1_{PPML}} = -0.0159$ $\omega_{2_{BNMR}} = 0.0118$ $\omega_{2_{PPML}} = -0.0119$

In general, these results show that there is evidence of gender, educational and social status intra-group proximities in the Brazilian network. In the international literature, our results lead to the same conclusions as Whittington and Smith-Doerr (2008) and Jung and Ejermo (2007) for gender. On educational and wage levels, our results are a novelty to the empirical research on the theory of proximities.

Moreover, the Brazilian network follows the historical development of Brazilian regions, rich and high-density coastal cities in the southeast, and a poor and low-density area in central and northern areas. Adding to this, the fact that Brazil covers an area of 8.5 million km², the clustering of highly qualified workers in a small area reinforces the role of geographical proximity. The geographical distribution of the collaboration network reinforces the geographical agglomeration pattern of patents in Brazil, as identified by Gonçalves and Almeida (2007) and Oliveira, Gonçalves and Almeida (2016).

6. CONCLUSIONS

This paper investigates the formation of perceptions within the Brazilian inventor network over the first decade of the 21st century. We identify the motivation factors that make ties stronger or weaker. We can divide our contributions as follows: (a) the use of new database with personal data, (b) the proposition of a dominance coefficient as a perception measurement, (c) the case study of the proximity theory applied to an emerging country, (d) the model result itself.

First, we merged two datasets to create a new database: Brazilian Patent applications at the National Institution of Intellectual Property (INPI) and the workers registries from the Ministry of Labor and Employment of Brazil. The data merging allows us to link inventors and their personal data, such as educational level, salary and address.

Second, we developed a coefficient to extract implicit hierarchy from ties which knowledge flows both ways between nodes. Since this technique attempts to simulate a node's weight in the link based on its relative status, we called it *dominance hierarchy coefficient*. We broke it down into (a) average and (b) difference coefficients, which can be manipulated to indicate the role of each node in the tie: the dominant node and the dominated node. We used these two roles as a perception proxy to illustrate how inventors form their perception system based on their role. We believe this coefficient is useful to deal with networks without explicit hierarchies and future empirical works would benefit from it.

Third, our paper is an addition to the scarce literature on networks applied to Brazil. Moreover, while many papers explore innovation networks in developed countries, our work is a contribution to the understanding of the innovation cooperation in an emerging country, where intellectual property institutions are still immature.

Finally, we summarize the results below:

- **Physical distance and border constraints matter:** Individuals have a higher propensity to connect with close partners. This results is similar to the geographic agglomeration pattern of patents in Brazil (Gonçalves and Almeida, 2007). Moreover, compared to international findings, this result is similar to what Morescalchi et al (2015) and Scherngell and Barber (2008) found within European borders.

- **Network positioning is important:** Central nodes have a higher probability of connecting with other central nodes. A highly connected inventor tends to form groups with other highly connected inventors more easily than with an occasional one. This result demonstrates that in Brazil the social proximity of Torres and Gilly (1999) is similar to what Cassi and Plunket (2015) found in France.

- **There is gender inequality in the network**: The gender of inventors matters for tie formation and inventor grouping, analogous to what Jung and Ejermo (2007) and Whittington and Smith-Doerr (2008) found. In perception perspective, male inventors (the majority) connect to other male inventors expecting higher gains from these links.

- **Inventors seek partners with better levels of education and wealth, although it is not always possible**: Inventors form their perception based on what they expect from partnership. Thus, if they believe that better levels of education and wealth is a good proxy for better returns, they will look for partners with these characteristics. Nonetheless, they also avoid partners of a much lower status than their own, which leads links forming between similar inventors.

In conclusion, there are many paths to explore in the study of networks, especially in Brazil and other emerging countries. Our paper is, as far as we know, a first attempt to understand how inventors form their connections based on their perceptions in Brazil. Moreover, the use of personal information is an under-explored field, which requires some efforts to access and disclose data.

REFERENCES

Adams, J.D., Black, G.C., Clemmons, J.R., Stephan, P.E., 2005. Scientific teams and institutional collaborations: Evidence from US universities, 1981--1999. Res. Policy 34, 259–285.

Agrawal, A., Kapur, D., McHale, J., 2008. How do spatial and social proximity influence knowledge flows? Evidence from patent data. J. Urban Econ. 64, 258–269.

Albuquerque, E. da M., 2000. Domestic patents and developing countries: arguments for their study and data from Brazil (1980--1995). Res. Policy 29, 1047–1060.

Albuquerque, E. da M., 1996. Sistema nacional de inovação no Brasil: uma análise introdutória a partir de dados disponíveis sobre a ciência e a tecnologia. Rev. Econ. Política 16, 56–72.

Audretsch, D.B., Feldman, M.P., 1996. R&D spillovers and the geography of innovation and production. Am. Econ. Rev. 86, 630–640.

Balconi, M., Breschi, S., Lissoni, F., 2004. Networks of inventors and the role of academia: an exploration of Italian patent data. Res. Policy 33, 127–145.

Boschma, R., 2005. Proximity and innovation: a critical assessment. Reg. Stud. 39, 61–74.

Breschi, S., Lissoni, F., 2005. "Cross-Firm" Inventors and Social Networks: Localized Knowledge Spillovers Revisited. Ann. Econ. Stat. 189–209.

Cassi, L., Plunket, A., 2015. Research collaboration in co-inventor networks: combining closure, bridging and proximities. Reg. Stud. 49, 936–954.

Cassi, L., Plunket, A., 2014. Proximity, network formation and inventive performance: in search of the proximity paradox. Ann. Reg. Sci. 53, 395–422.

Castells, M., 1996. The network society. Oxford: Blackwell.

Chandrasekhar, A., 2015. Econometrics of network formation, in: The Oxford Handbook of the Economics of Networks. pp. 303–357.

Crescenzi, R., Nathan, M., Rodríiguez-Pose, A., 2016. Do inventors talk to strangers? On proximity and collaborative knowledge creation. Res. Policy 45, 177–194.

Dachs, B., Ebersberger, B., Lööf, H., 2008. The innovative performance of foreign-owned enterprises in small open economies. J. Technol. Transf. 33, 393–406.

De Faria, P., Lima, F., Santos, R., 2010. Cooperation in innovation activities: The importance of partners. Res. Policy 39, 1082–1092.

Ejermo, O., Karlsson, C., 2006. Interregional inventor networks as studied by patent coinventorships. Res. Policy 35, 412–430.

Erdös, P., Rényi, A., 1959. On random graphs, I. Publ. Math. 6, 290–297.

Fally, T., 2015. Structural gravity and fixed effects. J. Int. Econ. 97, 76–85.

Fitjar, R.D., Rodríguez-Pose, A., 2016. Nothing is in the air. Growth Change.

Gaughan, M., Bozeman, B., 2016. Using the prisms of gender and rank to interpret research collaboration power dynamics. Soc. Stud. Sci. 46, 536–558.

Giuri, P., Mariani, M., 2013. When distance disappears: Inventors, education, and the locus of knowledge spillovers. Rev. Econ. Stat. 95, 449–463.

Gonçalves, E., 2007. O padrão espacial da atividade inovadora brasileira: uma análise exploratória. Estudos Econômicos (São Paulo) 37, 405–433.

Goncalves, E., Almeida, E., 2009. Innovation and spatial knowledge spillovers: evidence from Brazilian patent data. Reg. Stud. 43, 513–528.

Granovetter, M., 1985. Economic action and social structure: The problem of embeddedness. Am. J. Sociol. 91, 481–510.

Henderson, J.V., 2007. Understanding knowledge spillovers. Reg. Sci. Urban Econ. 37, 497–508.

Hicks, D.M., Katz, J.S., 1996. Where is science going? Sci. Technol. Hum. Values 21, 379–406.

Jaffe, A.B., 1989. Real effects of academic research. Am. Econ. Rev. 957–970.

Jung, T., Ejermo, O., 2014. Demographic patterns and trends in patenting: Gender, age, and education of inventors. Technol. Forecast. Soc. Change 86, 110–124.

Kahneman, D., 2002. Maps of bounded rationality: A perspective on intuitive judgment and choice. Nobel Prize Lect. 8, 351–401.

King, G., Zeng, L., 2001. Logistic regression in rare events data. Polit. Anal. 137–163.

Lemos, M.B., Campos, B., Biazi, E., Santos, F., 2006. Capacitação Tecnológica e Catching Up: o caso das regiões metropolitanas emergentes brasileiras. Rev. Econ. Política 26, 95–118.

Maggioni, M.A., Nosvelli, M., Uberti, T.E., 2007. Space versus networks in the geography of innovation: A European analysis. Pap. Reg. Sci. 86, 471–493.

Marshall, A., 1890. Principle of Economics, 8 th. Ed. 1920).

Montobbio, F., Sterzi, V., 2013. The globalization of technology in emerging markets: a gravity model on the determinants of international patent collaborations. World Dev. 44, 281–299.

Morescalchi, A., Pammolli, F., Penner, O., Petersen, A.M., Riccaboni, M., 2015. The evolution of networks of innovators within and across borders: Evidence from patent data. Res. Policy 44, 651–668.

Nagaoka, S., Motohashi, K., Goto, A., 2010. Patent statistics as an innovation indicator. Handb. Econ. Innov. 2, 1083–1127.

Nagaoka, S., Owan, H., 2011. Exploring the Sources of Firm-level Scale Economies in R&D: Complementary assets, internal and external knowledge inflows, and inventor team size.

Ozman, M., 2009. Inter-firm networks and innovation: a survey of literature. Econ. Innov. New Technol. 18, 39–67.

Paier, M., Scherngell, T., 2011. Determinants of collaboration in European R&D networks: empirical evidence from a discrete choice model. Ind. Innov. 18, 89–104.

Picci, L., 2010. The internationalization of inventive activity: A gravity model using patent data. Res. Policy 39, 1070–1081.

Powell, W.W., Giannella, E., 2010. Collective invention and inventor networks. Handb. Econ. Innov. 1, 575–605.

Saxenian, A., 1994. Regional networks: industrial adaptation in Silicon Valley and route 128.

Scherngell, T., Barber, M.J., 2009. Spatial interaction modelling of cross-region R&D collaborations: empirical evidence from the 5th EU framework programme. Pap. Reg. Sci. 88, 531–546.

Shaw, A.T., Gilly, J.-P., 2000. On the analytical dimension of proximity dynamics. Reg. Stud. 34, 169–180.

Sidone, O.J.G., Haddad, E.A., Mena-Chalco, J.P., 2016. A ciência nas regiões brasileiras: evolução da produção e das redes de colaboração científica. Transinformação 28.

Sidone, O.J.G., Haddad, E.A., Mena-Chalco, J.P., 2016. Scholarly publication and collaboration in Brazil: The role of geography. J. Assoc. Inf. Sci. Technol.

Silva, J.M.C.S., Tenreyro, S., 2006. The log of gravity. Rev. Econ. Stat. 88, 641–658.

Silva, J.M.C.S., Tenreyro, S., 2010. On the existence of the maximum likelihood estimates in Poisson regression. Econ. Lett. 107, 310–312.

Silva, J.M.C.S., Tenreyro, S., 2011. Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. Econ. Lett. 112, 220–222.

Silva, J., Tenreyro, S., 2011. Poisson: some convergence issues. Stata J. 11, 207–212.

Simões, R., Baessa, A., Campolina, B., Silva, L., 2002. A distribuição espacial da produção científica e tecnológica brasileira: uma descrição de estatísticas de produção local de patentes e artigos científicos. Rev. Bras. Inovação 1, 225–251.

Sorenson, O., Rivkin, J.W., Fleming, L., 2006. Complexity, networks and knowledge flow. Res. Policy 35, 994–1017.

Ter Wal, A.L.J., Boschma, R.A., 2009. Applying social network analysis in economic geography: framing some key analytic issues. Ann. Reg. Sci. 43, 739–756.

Whittington, K.B., Smith-Doerr, L., 2008. Women inventors in context: Disparities in patenting across academia and industry. Gend. Soc. 22, 194–218.

Wuchty, S., Jones, B.F., Uzzi, B., 2007. The increasing dominance of teams in production of knowledge. Science (80-.). 316, 1036–1039.