

## Tipologia e caracterização dos agentes de redes de inovação no Brasil<sup>1</sup>

Raquel Coelho Reis<sup>2</sup>

Eduardo Gonçalves<sup>3</sup>

Admir Antonio Betarelli Junior<sup>4</sup>

**Resumo:** Conhecimentos localizados tornam regiões mais competitivas e especializadas tecnologicamente, porém podem gerar *lock-in* regional. Assim, o acesso a fluxo de informação externa, por agentes locais, permite renovar a fonte de conhecimento regional. Na perspectiva da análise de redes de inovação há quatro tipos de agentes: nós isolados, estrela interna, estrela externa e *gatekeeper*. O estudo analisa as ligações destes agentes na rede de inovação regional considerando as produções tecnológicas brasileiras, identificando suas características e a consistência de suas classificações. A abordagem exploratória é realizada por técnicas de análise multivariada e modelo econométrico *logit* com efeitos fixos compreendendo o período de 2000 a 2011. A acurácia das classificações dos inventores é revelada pela técnica de discriminação logística e *logit* multinomial revelando-se uma taxa de erro de classificação baixa, aproximadamente 10%, e grupos heterogêneos. Além disso, a análise de correspondência conjunta evidencia que as variáveis respostas são distintas para cada tipo de agente. Nós isolados pertencem a instituições que auxiliam sua produção inovativa e os não isolados possuem maior nível educacional, como mestrado e/ ou doutorado. E, por meio da análise econométrica, identifica-se uma maior distinção entre os tipos de agentes, destacando-se os *gatekeepers* como agentes com maior capacidade de absorção, e pertencentes a instituições públicas e de ensino.

**Palavras-chave:** Redes de inovação; *Gatekeepers*; Análise discriminante; *Logit* multinomial; análise de correspondência.

**Abstract:** Localized knowledge makes regions more competitive and technologically specialized. However, they can generate regional lock-in. Therefore, access to external information flow, by local inventors, can renew the regional knowledge source. From the perspective of the analysis of innovative networks there can be four types of agents: isolated nodes, internal star, external star and gatekeeper. This study analyzes the links among inventors in the regional innovation network considering the Brazilian technological activity to identify their characteristics and the consistency of their classifications. The exploratory analysis is carried out by multivariate techniques and econometric model based on fixed effect logit, covering the period from 2000 to 2011. The accuracy of the inventors' classification is revealed by discriminant analysis and multinomial logistic regression, the main results reveal a low classification error rate, approximately 10%, and heterogeneous groups. The analysis of correspondence identify that the response variables are distinct for each type of agent. Isolated individuals belong to institutions that aid their innovative production and non-isolated agents possess a higher educational level, such as master's and doctoral degrees. And, by means of the econometric analysis, the results reveal a greater distinction between the groups of the agents, like gatekeepers are shown as agents with greater capacity of absorption and belonging to public and teaching institutions.

---

<sup>1</sup> Os autores agradecem o apoio da FAPEMIG, CNPq e CAPES.

<sup>2</sup> Doutoranda do Programa de Pós Graduação em Economia Aplicada (PPGE) da Universidade Federal de Juiz de Fora (UFJF).

<sup>3</sup> Professor da Faculdade de Economia da Universidade Federal de Juiz de Fora (UFJF) e do PPGE, pesquisador do CNPq.

<sup>4</sup> Professor da Faculdade de Economia da Universidade Federal de Juiz de Fora (UFJF) e do PPGE.

**Key Words:** Innovative networks; Gatekeepers; Discriminant analysis; Multinomial logistic regression; Correspondence analysis.

**Área 09: Economia Industrial e da Tecnologia.**

**JEL: O31, O33, D85.**

## 1. Introdução

Atividades inovativas são geradas por meio da combinação de conhecimentos novos e distintos, o que podem ser facilitados pela proximidade geográfica entre agentes, em especial, ao promover interações e relações de colaboração entre os agentes (BOSCHMA, 2005). Sob a perspectiva de interações locais, o *local buzz* é o cenário onde ocorrem trocas de conhecimentos, limitadas geograficamente, entre empresas, instituições e demais agentes localizados em uma mesma região. O *local buzz* permite trocas de conhecimentos tácitos por haver facilidade em contatos *face a face* e colaboração mútua entre os agentes. Além disso, é capaz de gerar especialização tecnológica e, conseqüentemente, melhorar o nível de competitividade regional. Contudo, interações intensas e limitadas geograficamente devem incorrer no ônus do *lock-in* regional. Isto ocorre perante a dificuldade em acessar fluxos tecnológicos externos, originando redundância de conhecimento e dificuldade em superar barreiras locais (BATHELT *et al.*, 2004; MORRISON *et al.*, 2013).

Por outro lado, o *global pipeline* ocorre quando fluxos de conhecimentos são capazes de ultrapassar fronteiras locais, atingindo demais regiões distantes geograficamente. Isto é possível graças ao avanço da globalização incorporando o caráter transnacional e inovações abertas nas redes de inovação (CHESBROUGH, 2006). Dessa maneira, assim como ocorre em *clusters* de sucesso, inovações regionais podem ser facilitadas quando há a interação entre *local buzz* e o *global pipelines* (BATHELT *et al.*, 2004). Diante disso, acredita-se que nós centrais nas redes são capazes de exercer essa função. Tais nós são denominados *gatekeepers* e considerados agentes com o poder de intermediar ligações locais e externas, captando conhecimentos externos e difundindo-os no sistema de inovação de suas regiões (GIULIANI; BELL, 2005).

No entanto, os agentes também podem exercer outros três possíveis papéis na rede: nós isolados, estrela interna ou estrela externa. De modo que, o primeiro destes ocorre quando o inventor não se liga a nenhum outro agente na rede, seja internamente ou externamente a sua região sendo autossuficiente em gerar inovações. Já o segundo tipo, estrela interna, é aquele associado ao *local buzz* (ALLEN; COHEN, 1969). Ou melhor, é o tipo de agente que possui diversas ligações locais, o que os caracteriza como agentes potencialmente mais especializados tecnologicamente, mas propensos ao *lock-in*. E, por último, o agente do tipo estrela externa é aquele que possui variadas ligações com agentes de outras regiões, mas não são capazes ou não estão dispostos a disseminar tais conhecimentos adquiridos externamente em sua região de origem (GIULIANI; BELL, 2005; GIULIANI, 2011).

Portanto, tendo em vista as possíveis classificações dos agentes quanto ao tipo de papel que podem exercer na rede de inovação, este artigo tem como um dos objetivos analisar cada uma dessas classificações, pois há uma escassez de trabalhos que investiguem as características dos tipos de agentes nas redes de inovação. O outro propósito da pesquisa é explorar as relações existentes entre as características dos inventores. Assim, é possível destacar quais características são mais predominantes em cada tipo de agente na rede de inovação brasileira. Por conseguinte, é possível avaliar se existem características a serem mais exploradas a fim de que um agente possa se tornar central em uma rede, i.e., classifica-lo como um *gatekeeper*. Para tanto, esta pesquisa utiliza métodos estatísticos de análise multivariada, que permitem verificar probabilisticamente tanto a acurácia das classificações e

se estas incorrem em menores taxas de erro quanto a associação entre as respostas de variáveis características dos agentes. O estudo está dividido em mais quatro seções, além desta introdução. Na segunda seção é abordada a revisão de literatura a respeito de redes de inovação e a definição dos tipos de agentes nas redes, além de destacar o papel do agente central, *gatekeeper*. A terceira seção apresenta as técnicas estatísticas e a base de dados, destacando as variáveis características consideradas. A quarta seção reúne os principais resultados obtidos por meio dos métodos aplicados aos dados. Por fim, a quinta seção compreende as considerações finais da pesquisa.

## 2. Revisão de Literatura

### 2.1 Redes de Inovação e seus tipos de agentes

Agentes criam novidades tecnológicas combinando conhecimentos prévios ou adquiridos por meio de interação com demais agentes, sendo estes últimos o centro do sistema de inovação (NELSON, 1993). A falta de interação (ligações) com demais agentes pode ser uma evidência de deficiência de capacidade ou habilidades internas à região (GIULIANI; BELL, 2005).

Não obstante, em um sistema regional nem todos agentes são capazes ou incentivados a interagir com agentes locais e externos (ALLEN, 1969; GIULIANI; BELL, 2005). Espera-se que a capacidade tecnológica interna de firmas tenha relação com a habilidade destas de se relacionar com firmas de demais regiões. Ademais, considera-se que a proximidade tecnológica e de conhecimento, que podem permitir a absorção e recodificação de conhecimentos externos, mesmo a grandes distâncias geográficas (COHEN; LEVINTHAL, 1990). Nesse sentido, sugere-se que agentes capacitados a interagir com outros agentes, locais e externos, combinam o *local buzz* e *global pipeline*. Tais agentes com tais características de combinar e intermediar tais fluxos de conhecimentos, absorver e disseminar conhecimentos são os *gatekeepers*.

O termo *gatekeeper* surgiu na década de 1960 e se refere a agentes que absorvem conhecimentos de origem externa a sua região e dissemina-os a seus parceiros locais (ALLEN; COHEN, 1969). No entanto, originalmente tais atividades eram associadas a ganhos monetários e, portanto, *gatekeepers* também eram denominados como “*brokers*”. De modo que, na ausência de confiança entre os agentes, os *gatekeepers* eram ditos capazes de mediar negociações de fluxos de conhecimentos e recursos entre agentes internos e externos em troca de quantias monetárias (GOULD; FERNANDEZ, 1989). Atualmente, em discussões sobre redes de inovação, os *gatekeepers* são considerados agentes centrais na rede, relevantes na superação de barreiras de conhecimentos locais e na renovação de conhecimentos para maior produtividade tecnológica da rede (BRESCHI; LENZI, 2015).

No entanto, mesmo sendo considerados agentes de grande importância na rede, ainda são poucos os trabalhos que tem como objetivo investigar seu comportamento e características que os tornam *gatekeepers*. Este fato tem associação com a falta de consistência de medidas na literatura que definam de maneira mais refinada e consistente desses agentes e conflite com os motivos que os levam ou capacite-os a agir como tal.

A medida mais usual tem sido a de Gould e Fernandez (1989) que considera que os *gatekeepers* são agentes que intermediam conhecimentos não redundantes entre agentes internos e externos. No entanto, existem variações dessa medida como uso de índices de centralidade, consideração de inventores com a medida usual acima da média, medidas refinadas pela intensidade e variedade das ligações, dentre outras (GIULIANI; BELL, 2005; GIULIANI, 2011; LEGALLO; PLUNKET, 2016; GRAFF, 2011; GRAF; KRUGER, 2009).

Ademais, a escassez de trabalhos sobre o tema pode se dar devido à dificuldade em acessar dados pessoais dos agentes e compatibilizá-los com os dados das produções tecnológicas inovativas. Além disso, é complexo comparar tais medidas em regiões com comportamento inovador heterogêneo e agente posicionados de maneira tão distinta na rede.

Considerando que o investimento em conhecimento tecnológico e a manutenção de relações entre indivíduos leva tempo, acredita-se que agentes de maior produtividade tecnológica inovativa, tenha maior probabilidade de se tornar *gatekeeper*. Isto ocorre devido à capacidade destes indivíduos em absorver conhecimentos externos e, conseqüentemente, serem mais procurados por agentes locais a fim de obtenção de maior conhecimento (GRAF, 2011).

Além disso, assume-se que para que um indivíduo seja capaz de possuir colaborações com agentes externos à sua região, além de ter acesso a tais conhecimentos externos, este deve possuir uma capacidade de absorção destes conhecimentos (COHEN; LENVINTHAL, 1990; GRAF, 2011). Sendo assim, não basta que o acesso aos conhecimentos externos e internos esteja disponível. É necessário que isto seja combinado com as habilidades prévias de acessar conhecimentos externos, compreende-los e recodifica-los de forma a torna-los aptos a serem disseminados e absorvidos por outros agentes locais que não possuem a mesma habilidade de absorção. Essa ideia sugere a primeira hipótese a ser investigada pelo presente estudo.

***Hipótese 1. Há uma relação positiva entre a capacidade de absorção do agente e a probabilidade de este se ligar a outros agentes na rede e potencialmente ser um gatekeeper.***

A importância relativa das ligações mediadas por *gatekeepers* varia de acordo com os conhecimentos específicos dos locais a que estes pertencem. Ligações mediadas por tais agentes, em geral, são mais importantes para regiões de conhecimentos especializados e localizados do que para as mais diversificadas. Dado que, os *gatekeepers* são capazes de recodificar conhecimentos externos, de uma linguagem e conhecimento distinto e transferi-los após tal recodificação à sua região de origem (BRESCHI; LENZI, 2015).

No entanto, nem sempre agentes que possuem características e o poder de atuar como *gatekeepers* vão estar dispostos a agir como tal. De modo que, a capacidade de absorção do indivíduo, a maneira como este se relaciona e é incentivado a se relacionar com demais agentes (locais ou externos) pode ter uma associação com os incentivos do seu ambiente de trabalho.

Acredita-se que organizações privadas e públicas possuem objetivos distintos quanto à transferência de conhecimentos e participação em atividades de colaborações com demais instituições. Acredita-se que organizações públicas são mais propensas a participar de colaborações, seja com demais instituições públicas ou privadas (GRAF, 2011). Isto ocorre devido às organizações privadas estarem mais preocupadas com a competição com demais empresas e os riscos em dividir conhecimentos com estas. Tal ideia sugere a segunda hipótese a ser investigada para o caso da rede de copatenteamento no Brasil.

***Hipótese 2. Pertencer a uma organização pública eleva as chances de os agentes agirem como gatekeeper.***

Além disso, os esforços de pesquisas universitárias e instituição de ensino estão cada vez mais inseridos em atividades colaborativas, podendo até mesmo envolver participação industrial, promovendo a inovação coletiva. Ao investigar dados de atividade inventivas dos EUA, ao avaliar as patentes universitárias e do governo, entre 1975-1979 e 2001-2005, verifica-se um aumento da participação universitária principalmente na área da biotecnologia e semicondutores. Por outro lado, os resultados encontrados para o governo é que sua

participação apenas cresceu na área da biotecnologia. Dada combinação de patentes governamentais e universitárias, houve apenas elevação nas áreas farmacêuticas e biotecnológicas, e declínio no aeroespacial, comunicação óptica e semicondutores (POWELL; GIANNELLA, 2010). Há evidências que 90% de coautores de patentes de biotecnologia são vinculados a universidades ou organizações de pesquisa (GITTELMAN, 2007).

Sendo assim, perante as características e interesses do local onde se inserem, é possível existirem outros tipos de agentes que possam exercer outros papéis perante a rede de inovação. Em alguns casos, agentes podem preferir reter os conhecimentos adquiridos externamente, não os difundindo com demais agentes que interagem ou pertençam a sua região. Isto pode ocorrer por questão de competição regional e uma consequente resistência em divulgar conhecimentos relevantes. Tais agentes são caracterizados por possuírem muitas ligações externas e nenhuma (ou rara) ligação interna, estes agentes são denominados como estrelas externas (GIULIANI; BELL, 2005; GIULIANI, 2011; MORRISON *et al.*, 2013).

Já em outros casos, os agentes podem estar vinculados a uma situação de *local buzz*, i.e., possuem diversas ligações com demais agentes localizados em sua região e nenhuma ligação com agentes externos. Nessa situação, os agentes podem ser considerados como estrela interna e serem caracterizados por possuir conhecimentos específicos e pertencerem a uma rede fechada ou com conhecimento tecnológico mais restrito e de ligações densas. Estrelas internas embora auxiliem seus contatos a se ligarem externamente são pobres em ligações externas e possuem densas ligações internas. (ALLEN; COHEN, 1969; THUSMAN; SCANLAN, 1981).

Por fim, podem existir agentes que estão associados à ausência de ligações tanto internas, quanto externas. Entretanto, são capazes de gerar produções tecnológicas de maneira autossuficiente, e por isso participam ativamente da rede de inovação, estes são denominados como nós isolados. (GIULIANI; BELL, 2005; GIULIANI, 2011). Todavia, tais inventores tendem a possuir conhecimentos mais limitados ou básicos tecnologicamente. Tais evidências sugerem a última hipótese a ser investigada nesse estudo.

***Hipótese 3. Pertencer a organizações de ensino eleva as chances de o agente se comportar como um gatekeeper.***

### **3 Metodologia e Base de dados**

As análises de redes têm como objetivo investigar a relação entre seus agentes. Dessa maneira, *gatekeepers* se destacam como nós centrais devido à relevância do seu papel na rede como intermediador de ligações internas e externas. Não obstante, demais agentes também são relevantes dado que exercem papéis distintos na rede e podem mudar de função ao longo do tempo, podendo até mesmo vir a exercer o papel de *gatekeeper*. Assim como *gatekeepers* também podem deixar de ser um, dada às variações desta medida (GIULIANI, 2011).

Por isso, para verificar o papel dos agentes nas redes de inovação no Brasil, este artigo procura associar as características dos agentes às quatro possíveis classificações de agentes da rede. Para tal, métodos de análise multivariada, como análise discriminante e análise de correspondência, são usados para se compreender a inter-relação dessas características dos agentes e averiguar se os grupos a que pertencem são heterogêneos entre si e consistentes com a literatura. Adicionalmente, a abordagem econométrica, por meio de dados em painel e modelo *logit*, é utilizada a fim de dar consistência aos resultados sugeridos pelas técnicas multivariadas e compreender a relação de causalidade entre as variáveis e os resultados de classificação dos agentes obtidas.

### 3.1. Base de dados e variáveis

Para investigar as características dos tipos de agentes da rede de inovação brasileira e avaliar a acurácia de suas classificações quanto ao tipo de papel que exercem na rede, serão utilizados dados do INPI. Os dados abrangem as produções tecnológicas do Brasil por municípios compreendendo o período de 2000 a 2011.

Informações sobre os inventores presentes na base do INPI têm como origem os dados da RAIS e buscam caracterizar os agentes pelo nível educacional e ambiente de trabalho. Excluem-se da amostra aqueles agentes sem informação de região de origem, tendo em vista que as definições das ligações dos tipos de agentes precisam de informações sobre o local a que pertencem.

A variável dependente *tipo\_agente* assume o formato de uma variável categórica para cada um dos quatro tipos de agentes possíveis de existir na análise de redes de inovação do Brasil. Sendo assim, quando o agente é isolado, i.e, não possui ligações internas à sua região nem ligações externas, este assume valor 1. Quando o agente apenas possui ligações internas à sua região, é denominado como estrela interna e assume valor 2 para a variável categórica. Agentes que apenas possuem ligações externas à sua região são denominados como estrelas externas e assumem valor 3 na variável. E, por último, se o agente é um possível *gatekeeper*, assume valor 4 na variável categórica *tipo\_agente*.

Como forma de investigar quais variáveis estão relacionadas com a caracterização dos agentes como um tipo específico de agente nas redes de inovação do Brasil, consideram-se as informações do nível de educação dos inventores, conhecimento tecnológico e do ambiente de trabalho que se inserem. No entanto, as informações para a caracterização dos agentes são bastante restritas, sendo a maioria dos dados obtidos na forma qualitativa. Portanto, a maior parte das variáveis investigadas constitui-se de variáveis binárias e estão resumidas junto às suas estatísticas descritivas na Tabela 1<sup>5</sup>.

A fim de verificar a relação entre um conhecimento mais elevado por parte do agente e seu papel na rede de inovação, considera-se o grau de patentes de alta tecnologia (*grau\_high*). Esta variável é dada pelo número de patentes classificadas nas classes de alta tecnologia, pela classificação IPC-4 dígitos<sup>6</sup> dividido pelo número total de patentes depositadas pelo inventor no período de análise, produzidas em parceria ou não.

Já o grau de diversificação tecnológica (*grau\_diversificação*) é dado pelo número de classes tecnológicas distintas em que o inventor possui patentes depositadas em relação ao total de patentes depositadas pelo inventor. Para isto, considera-se a classificação internacional de patentes (IPC) a quatro dígitos, que compreendem a existência de 32 grupos tecnológicos. Esta variável tem como objetivo mensurar a diversidade de conhecimento do inventor.

A variável *tamanho\_equipe* é dada pelo número de inventores distintos participantes do depósito da patente no INPI. Esta é uma medida da capacidade de absorção de conhecimento diversificado do agente, levando-se em consideração o tamanho da equipe geradora da patente, assim como em Graff (2011).

Com o objetivo de captar o grau de formação dos inventores, foram inseridas *dummies* que captam se os mesmo possuem ensino superior completo (*dummy\_esuperior*) ou se possuem título de mestrado ou doutorado (*dummy\_epós*). Neste caso, as variáveis assumem o

---

<sup>5</sup> A matriz de correlação das variáveis explicativas se encontra no Apêndice (A1), e evidencia ausência de multicolinearidade. Tal fato restringe o uso de técnicas de análises multivariadas a apenas aquelas que não pressupõem correlação alta entre as mesmas.

<sup>6</sup> As classes tecnológicas consideradas como de alta tecnologia compreendem os setores de tecnologia computacional e negócios automatizados, aviação, engenharia genética, *lasers*, semicondutores, tecnologia da telecomunicação e biotecnologia.

valor 1 caso o agente possua, respectivamente, titulação em ensino superior e título de mestre e/ou doutorado e 0, caso o contrário. Tais variáveis estão associadas ao fato de que agentes com maior grau de ensino são mais aptos a assimilar conhecimentos novos e de origens distintas. Logo, também podem ser consideradas medidas da capacidade de absorção de conhecimento específico dos mesmos.

A fim de investigar a relação do ambiente de trabalho do inventor e seu papel na rede de inovação, consideram-se *dummies* de instituição pública (*dummy\_ipub*), instituição de ensino (*dummy\_iensino*) e indústria (*dummy\_ind*). Estas assumem valor 1, caso os agentes pertençam a alguma delas e 0, caso contrário.

Além disso, considera-se uma variável sobre o tamanho da organização - ambiente de trabalho - cujo inventor faz parte, esta é denominada como *dummy\_grand<sub>org</sub>* e assume valor 1 caso o inventor pertença a uma organização com mais de 1000 (mil) funcionários e 0, caso contrário. Empresas maiores podem incorrer em maiores incentivos a colaborações externas a fim de ser um meio de superar barreiras tecnológicas e ser uma maneira mais rápida e de menor custo de obter tecnologias e conhecimentos novos.

**Tabela 1 – Descrição e estatísticas das variáveis 2000-2011**

Variável	Descrição	Média	D.P.	Min	Max
<i>Grau_high</i>	$\log\left(\frac{\text{patentes de alta tecnologia do inventor}}{\text{total de patentes do inventor}}\right)$	0,09	0,22	0	0,69
<i>Grau_diversificação</i>	$\log\left(\frac{\text{patentes em classes tec. distintas por inventor}}{\text{total de patentes do inventor}}\right)$	0,70	0,25	0	1,95
<i>Tamanho_equipe</i>	$\log(\text{n}^\circ \text{ maximo de parceiros por patente})$	0,96	0,41	0	2,89
<i>Dum_esuperior</i>	Variável binária; 1 se o agente possui curso superior e 0 se não.	0,07	0,27	0	1
<i>Dum_epos</i>	Variável binária; 1 se o agente possui título de mestrado e/ou doutorado e 0 se não.	0,02	0,13	0	1
<i>Dum_iensino</i>	Variável binária; 1 se o agente pertence a uma instituição de ensino e 0 se não.	0,05	0,22	0	1
<i>Dum_ipub</i>	Variável binária; 1 se o agente pertence à instituição pública e 0 se não.	0,05	0,22	0	1
<i>Dum_ind</i>	Variável binária; 1 se o agente pertence à indústria e 0 se não.	0,08	0,27	0	1
<i>Dum_grand<sub>org</sub></i>	Variável binária; 1 se o agente pertence a uma grande organização e 0 se não.	0,09	0,28	0	1
<i>Dum_metro-pole</i>	Variável binária; 1 se o agente pertence às regiões metropolitanas e 0 se não.	0,72	0,45	0	1

Fonte: Elaboração própria.

### 3.2. Metodologia<sup>7</sup>

Para extrair informações a respeito das características dos tipos de agentes presentes na rede de inovação no Brasil, a metodologia utilizada nesta pesquisa compreende métodos de análise multivariada e modelo *logit* com efeitos fixos. A justificativa sobre o uso de análise multivariada se dá por esta ser uma abordagem que permite analisar e medir diversas variáveis simultaneamente e suas relações com o fenômeno de interesse. Por outro lado, o modelo *logit* permite analisar por inferências estatísticas as causalidades entre as variáveis e

<sup>7</sup> Os conceitos mencionados têm por base Johnson e Dean (2007) e Wooldridge (2010).

os fenômenos de interesse, além da consistência dos resultados sugeridos pela análise multivariada.

### 3.2.1 Análise Discriminante

A análise discriminante (AD) é uma técnica de análise multivariada cujo objetivo é discriminar, i.e., separar observações distintas em grupos pré-definidos e classificar novas observações nestes grupos. Conseqüentemente, esta técnica é uma ferramenta adequada para explorar as diferenças significativas entre os grupos pré-definidos e alocar uma nova observação no grupo mais apropriado. Basicamente, o método tem como objetivo encontrar uma função ou conjunto de funções capazes de discriminar e minimizar os erros de classificação dos grupos pré-definidos. Cabe destacar, que para a eficácia do método os grupos devem ser estatisticamente distintos, levando a grupos homogêneos internamente e heterogêneo em relação aos outros grupos. Sendo assim, é adequado conhecer as características gerais dos grupos a fim de alcançar uma função discriminante capaz de alocar novas observações no grupo ideal.

A análise discriminante necessita de uma variável dependente categórica, sendo que o número de categorias representa o número de grupos pré-definidos. Portanto, na presente análise a variável dependente *tipo\_agente* se mostra adequada para investigar a qualidade da classificação dos inventores nos quatro grupos de agentes possíveis de existir na rede de inovação no Brasil. No entanto, como além de variáveis contínuas há presença de variáveis binárias entre os grupos de variáveis a serem consideradas na análise, o que sugere-se o uso do método discriminante logístico junto ao método *logit* multinomial. Contudo, o *logit* multinomial é a base para o método discriminante logístico e utilizado para calcular sua taxa de probabilidade e probabilidades posteriores. Os métodos se assemelham por possuir variável dependente categórica e serem utilizados em regressões contendo variáveis binárias e contínuas.

O modelo logístico prescinde da normalidade<sup>8</sup> das variáveis, bem como permite lidar com possíveis problemas causados por variáveis binárias como variância não constante, e, conseqüentemente probabilidades  $0 > p > 1$ . O método logístico utiliza a razão de chance (RC) no lugar da probabilidade de pertencimento para atribuir probabilidade da observação pertencer a um determinado grupo, de modo que  $\beta$  é estimado por máxima verossimilhança.

$$RC = \frac{\hat{p}(X)}{1 - \hat{p}(X)} = \exp(\hat{\beta}_0 X_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_p X_p) > 1$$

sendo  $X$  o vetor de variáveis,  $\hat{p}(X)$  é a probabilidade de pertencer e  $1 - \hat{p}(X)$  é a probabilidade de não pertencimento. Logo, o *logit* é dado por:

$$\ln\left(\frac{\hat{p}(X)}{1 - \hat{p}(X)}\right) = \hat{\beta}_0 X_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_p X_p$$

Destarte, é calculado o escore de cada observação objetivando que esta seja classificada no grupo em que obtêm a maior probabilidade estimada de pertencer em comparação às demais.

Vale mencionar que além de permitir a investigação de variáveis quantitativas e qualitativas ao mesmo tempo, o método também possui flexibilidade quanto à ausência de

---

<sup>8</sup> Os testes de normalidade univariada e multivariada foram realizados e evidenciaram que nenhuma das variáveis assume normalidade.

normalidade dos dados. No entanto, o modelo logístico é sensível a problemas de multicolinearidade entre as variáveis explicativas, que podem afetar as estimativas dos parâmetros de interesse. Contudo, como a matriz de correlação (A1) exibe resultados de correlações baixas entre as variáveis a técnica se mostra como adequada aos dados.

### 3.2.2 Análise de Correspondência Múltipla e Conjunta

A análise de correspondência múltipla e conjunta (ACM) é um método de análise multivariada apropriada para mais de duas variáveis categóricas e não exige a normalidade das mesmas. O método tem como objetivo reduzir os dados, por meio de redução da dimensão das variáveis originais, porém mantendo-se a estrutura original dos dados; e dependendo da quantidade de variáveis utilizadas os escores gerados pelo método podem ser utilizados como variáveis contínuas.

O método ACM utiliza uma matriz indicadora, no caso de poucas observações, e uma matriz Burt que é uma tabulação cruzada. Ambas são responsáveis por explorar e analisar as relações entre um conjunto de variáveis categóricas. No entanto, estas matrizes são sujeitas a inflar as distâncias qui-quadrado entre os perfis colunas – vetor de proporção das frequências em relação ao total de cada coluna – e a inércia total – verifica se a associação entre as variáveis são estatisticamente significativas. Deste modo, sugere-se a utilização do método de análise de correspondência conjunta (ACJ) que tem como princípio corrigir este problema por meio de uma matriz que considera apenas os elementos fora da diagonal da matriz Burt e executa m interações substituindo os elementos de fora por elementos da matriz de aproximação até que as variações das interações sejam inexpressivas.

Este é um método apropriado quando o interesse é reduzir as correspondências entre os dados das variáveis categóricas em dimensões menores e que estas possam representar de forma satisfatória a estrutura dos dados originais. Sendo assim, na presente pesquisa o método permite analisar a relação de correspondência entre as variáveis *dummies*, que representam as características de formação e ambiente de trabalho de cada tipo de agente. Ademais, podem-se comparar os resultados obtidos sobre as correspondências entre as variáveis e os tipos de agentes, assim como se é possível reduzir tais características em dimensões menores. Todavia, sugere-se como medidas de qualidade que as dimensões possam estar associadas a 90% da estrutura dos dados originais e que as análises considerem dimensões com autovalores relacionados com a inércia total e que atinjam o valor mínimo de 0,2. A inércia total principal é uma forma de verificar se a associação entre as variáveis são estatisticamente significativas.

### 3.2.3. Logit com Efeito Fixo

O modelo *logit* é apropriado para o caso em que as variáveis dependentes assumem o formato de variáveis binárias. No presente estudo o interesse se pauta na caracterização dos quatro tipos de agentes na rede. Sendo assim, ao analisar os quatro tipos de agentes separadamente, o método se mostra como adequado já que para cada tipo de agente a variável assume valor igual a 1 para a ocorrência do fenômeno, e 0 caso contrário.

No caso de dados em painel, os mesmos podem assumir a existência de efeitos fixos ou efeitos aleatórios nos dados. Os efeitos aleatórios pressupõem que existam diferentes interceptos para cada observação de maneira aleatória. Por outro lado, os efeitos fixos sugerem que os regressores possuem correlação ao nível do indivíduo. Ou seja, embora o intercepto seja distinto entre os indivíduos, ele não se altera para o mesmo indivíduo ao longo do tempo, deste modo, é necessário o controle ou eliminação dos efeitos fixos.

A fim de detectar qual o efeito existente na amostra, utiliza-se o teste de *Hausman* que tem como hipótese nula a presença de efeitos aleatórios como eficientes. Realizou-se o teste

para todos os quatro casos de agentes considerados na amostra, havendo a rejeição da hipótese nula em todos os casos. Assim, sugere-se a utilização do modelo *logit* considerando a presença de efeitos fixos, sendo estes mais consistentes.

Além disso, cabe destacar que como os inventores considerados na amostra não inventam todos os anos, este é um painel desbalanceado. No entanto, como a origem dos dados faltantes não são decorrentes de erros, estas não configuram *viés* de seleção.

#### 4. Resultados

Uma vez que o principal objetivo deste estudo é analisar quais características colaboram para que os agentes sejam classificados em cada um dos tipos possíveis de agentes em redes de inovação, foi definida uma estratégia baseada em uso de técnicas de análises multivariadas. Para isto, a primeira abordagem foi investigar a acurácia das classificações. Os agentes foram, primeiramente, classificados como passíveis de exercer os papéis de nós isolados, estrela interna, estrela externa ou *gatekeeper*. Logo, a fim de verificar se as probabilidades de pertencimento dos inventores em um determinado grupo possuem taxas de erro de classificação baixa objetivando-se alcançar o menor custo de erro de classificação, utilizou-se o método de discriminante logística.

A Tabela 2 apresenta os resultados para a taxa de erro de classificação de agentes nos grupos pré-definidos, considerando-se probabilidades de pertencimento proporcionais *a priori*. A fim de buscar consistência e comparabilidade dos resultados, a amostra foi dividida em duas compreendendo os períodos 2000-2005 e 2006-2011.

Além disso, teve-se o cuidado de analisar as probabilidades de pertencimento dos agentes cujas probabilidades de pertencimento eram próximas a de pertencer a outro grupo. Como a finalidade é compreender as variáveis que levam os agentes a se classificarem ou não se classificarem nos grupos pré-determinados, evita-se reclassificá-los. Deste modo, os resultados da taxa de erro de classificação foram recalculados considerando-se como regra de desempate, corretas as observações que obtiveram probabilidade de classificação maior do que 25% do que o grupo de probabilidade próxima. Esta nova taxa é denominada na Tabela 2 como taxa de erro de classificação (real) e objetiva ser mais parcimoniosa quanto aos resultados obtidos da reclassificação ao ponderar a probabilidade proporcional *a priori* pela probabilidade igual de pertencimento em cada um dos grupos. Adicionalmente, a Tabela 2 também reporta os resultados para os testes estatísticos de médias dos grupos a fim de identificar se os mesmos são heterogêneos entre si e se as variáveis são independentes. A Tabela 3 tem como objetivo reportar as médias das variáveis por grupo.

Verifica-se por meio da Tabela 2 que as taxas de erro de classificação foram moderadamente baixas considerando-se que esta é uma amostra grande. No entanto, verifica-se que para o período mais recente a taxa de erro elevou-se em 5% se comparada ao período anterior. Isto pode ser explicado pelo aumento em aproximadamente 7% do número de observações. Identifica-se também que o grupo 3 obteve 100% de erro de classificação. Ademais, também é possível notar que o número de agentes previamente classificados como *gatekeepers* para o período mais recente de análise quase que dobrou em relação ao período anterior.

Todavia, encontrou-se que as observações obtiveram probabilidades muito próximas de pertencimento com outros grupos, principalmente com o grupo 4. Isto pode ser justificado pelo fato de que inventores considerados como estrela externa podem futuramente se enquadrar como *gatekeepers* por possuir características semelhantes. Contudo, considerando-se os resultados para a taxa de erro de classificação (real) obteve-se uma queda significativa na taxa de erro de classificação de 30% no primeiro período e aproximadamente 40% no segundo.

**Tabela 2 – Porcentagem de taxa de erro de classificação dos agentes na rede de inovação brasileira nos períodos 2000-2005 e 2006-2011**

	2000-2005 (N = 23.065)	2006-2011 (N = 24.662)
G1 - Nó isolado (N = 14.646)	0,5%	G1 - (N = 13.410) 1%
G2 - Estrela interna (N = 4.769)	8%	G2 - (N = 5.016) 17%
G3 - Estrela externa (N = 1.698)	100%	G3 - (N = 2.519) 100%
G4 - <i>Gatekeeper</i> (N = 1.952)	49%	G4 - (N = 3.717) 28%
<i>Lambda de Wilk</i>	0,2332***	0,2197***
<i>Traço de Pillai</i>	0,7826***	0,7907***
<i>Traço de Lawley-Hotelling</i>	3,2213***	3,5044***
<i>Raiz de Roy</i>	3,2003***	3,4909***
<i>LR ajustado chi</i>	16092,53****	17615,56****
<b>Taxa de erro de classificação</b>	<b>13%</b>	<b>18%</b>
<b>Taxa de erro de class. (real)</b>	<b>10%</b>	<b>11%</b>

Nota<sup>1</sup>: N representa o número de observações que foram classificadas dentro de cada grupo.

Nota<sup>2</sup>: \*\*\* Representa significância de 1%

Fonte: Elaboração Própria de acordo com as saídas do STATA.

Além disso, evidenciam-se por meio dos testes de média (*lambda Wilk*) que os grupos são heterogêneos entre si, i.e, possuem médias distintas. Este resultado valida o objetivo de gerar grupos homogêneos internamente e heterogêneos entre si. Em geral, um coeficiente pequeno e a rejeição do teste sugere que não há linearidade entre as variáveis e, conseqüentemente, sua rejeição leva à rejeição dos demais testes, o que de fato, ocorre. Ademais, o teste LR também indica que as variáveis são independentes.

**Tabela 3 – Média das variáveis por grupo de agentes nos períodos 2000-2005 e 2006-2011**

Variável	2000-2005				2006-2011			
	G1	G2	G3	G4	G1	G2	G3	G4
<i>Grau_high</i>	0,76	0,77	0,77	0,80	0,79	0,81	0,81	0,84
<i>Grau_diversificação</i>	0,07	0,07	0,08	0,07	0,08	0,08	0,08	0,09
<i>Tamanho_equipe</i>	0,69	1,29	1,32	1,80	0,70	1,34	1,41	1,87
<i>Dum_esuperior</i>	0,01	0,04	0,05	0,14	0,06	0,16	0,19	0,31
<i>Dum_epos</i>	0,00	0,01	0,01	0,04	0,00	0,03	0,05	0,10
<i>Dum_iensino</i>	0,02	0,06	0,08	0,19	0,01	0,07	0,08	0,17
<i>Dum_ipub</i>	0,05	0,07	0,08	0,17	0,03	0,05	0,04	0,09
<i>Dum_ind</i>	0,07	0,12	0,17	0,16	0,06	0,09	0,11	0,09
<i>Dum_grand<sub>org</sub></i>	0,03	0,08	0,11	0,22	0,05	0,15	0,16	0,31
<i>Dum_metropole</i>	0,73	0,75	0,71	0,76	0,70	0,73	0,68	0,77

Fonte: Elaboração própria de acordo com as saídas do STATA.

Tendo em vista que os grupos possuem médias distintas, cabe investigar o comportamento das médias das variáveis por grupo de agentes, resultados que são reportados por meio da Tabela 3. Verifica-se que o grupo dos *gatekeepers* apresenta as maiores médias do grau de conhecimento em tecnologias de alta tecnologia (*grau\_high*), assim como grau de conhecimentos diversificados (*grau\_diversificação*), tamanho de equipe (*tamanho\_equipe*), pertencimento à região metropolitana (*dum\_metropole*). Além disso, os *gatekeepers* também possuem aproximadamente o dobro da média em relação aos demais grupos com inventores que possuem ensino superior completo (*dum\_esuperior*), título de mestrado e/ doutorado (*dum\_epos*), que fazem parte de uma grande organização (*dum\_grand<sub>org</sub>*), que pertencem a instituição pública (*dum\_ipub*) e à instituição de ensino (*dum\_iensino*). Já considerando

inventores que pertencem ao setor industrial (*dum\_ind*), o grupo dos inventores considerados como estrelas externas é o que apresenta a maior média em comparação aos demais grupos.

Em relação à associação das variáveis com o resultado de pertencer a cada um dos grupos, a Tabela 4 apresenta os resultados obtidos pela utilização da extensão do método de discriminante logística, o modelo *logit* multinomial. Além das significâncias das variáveis, a Tabela 4 apresenta os principais testes estatísticos e a taxa de erro de classificação assim como no resultado da discriminante logística, calculada a partir das probabilidades de pertencimento positiva obtidas junto ao modelo. Como grupo base do modelo *logit* multinomial, optou-se por considerar o grupo de nós isolados, dado que o maior interesse está em compreender o que leva os agentes a se ligarem aos demais e potencialmente virem a se comportar como agentes centrais, seja no âmbito local, externo ou em ambos.

Verifica-se que o conhecimento de alta tecnologia tem associação positiva com os agentes pertencerem ao grupo de *gatekeepers*, considerando-se ambos os períodos de análise. Além disso, pertencer a uma região considerada metrópole tem associação negativa com o agente pertencer ao grupo de inventores classificados como estrelas externas.

**Tabela 4 – Resultados modelo logit multinomial nos períodos 2000-2005 e 2006-2011**

Variáveis	2000-2005			2006-2011		
	G2	G3	G4	G2	G3	G4
<i>Grau_high</i>	0,47	0,58	1,39**	0,44	0,42	0,94**
<i>Grau_diversificação</i>	0,57	0,62	0,48	0,61**	0,67**	0,67***
<i>Tamanho_equipe</i>	28,63***	28,87***	32,17***	24,07***	24,77***	27,63***
<i>Dum_esuperior</i>	1,28***	1,36***	1,69***	-0,58**	-0,33	-0,08
<i>Dum_epos</i>	-0,78	-0,66	-0,33	0,10	0,22	0,77
<i>Dum_iensino</i>	-0,55	-0,35	0,11	-0,11	0,06	0,51
<i>Dum_ipub</i>	-0,03	0,08	0,30	-0,27	-0,40	-0,03
<i>Dum_ind</i>	-0,09	0,25	-0,01	0,91***	1,08***	1,23***
<i>Dum_grand<sub>org</sub></i>	0,55	0,74	0,78	-0,11	-0,27	-0,15
<i>Dum_metropole</i>	-0,24	-0,47*	-0,38	-0,16	-0,38*	-0,03
<i>Const.</i>	-28,07***	-29,43***	-35,09***	-23,87***	-25,40***	-30,58***
Testes	$R^2$ 0,68; <i>LogLike.</i> -7418,6; <i>Wald</i> 6048,8			$R^2$ 0,63; <i>LogLike.</i> -10651,1; <i>Wald</i> 5518,9		
Taxa de erro de classificação	13%			18%		
<b>Taxa de erro de classificação (real)</b>	10%			11%		

Nota<sup>1</sup>: O G1 foi suprimido por ser o grupo base do modelo multinomial.

Nota<sup>2</sup>: \*, \*\*, \*\*\* denotam respectivamente os níveis de significância de 10%, 5% e 1%.

Fonte: Elaboração própria de acordo com as saídas do STATA.

A variável de tamanho da equipe apresenta-se com uma associação positiva com todos os tipos de agentes considerados, e em ambos os períodos. Pertencer ao setor industrial também apresentou resultados positivos com todos os grupos para o período mais recente (2006-2011). Por fim, a variável que consiste na aquisição de título de ensino superior completo apresenta uma associação positiva com todos os grupos no primeiro período de análise (2000-2005), mas com sinal negativo para o grupo de estrela interna no período mais recente, sendo não significativa para os demais assim como as demais variáveis. Sugere-se que esta mudança de sinal pode ser devido ao nível de ensino superior estar perdendo o destaque para títulos em pós-graduação, mestrado e doutorado. Verifica-se que estes últimos coeficientes, em anos recentes, passaram a ter sinal positivo, porém foram não significativos para confirmar tal evidência.

A fim de investigar a similaridade entre as respostas das variáveis binárias utilizadas no estudo e procurar reduzir tais variáveis em uma dimensão menor, utiliza-se o método de análise de correspondência conjunta. Para isto, considera-se apenas o período mais recente dos dados (2006-2011) e investiga-se cada um dos grupos de agentes separadamente, considerando-se a classificação inicial já que esta possui uma taxa de erro de classificação

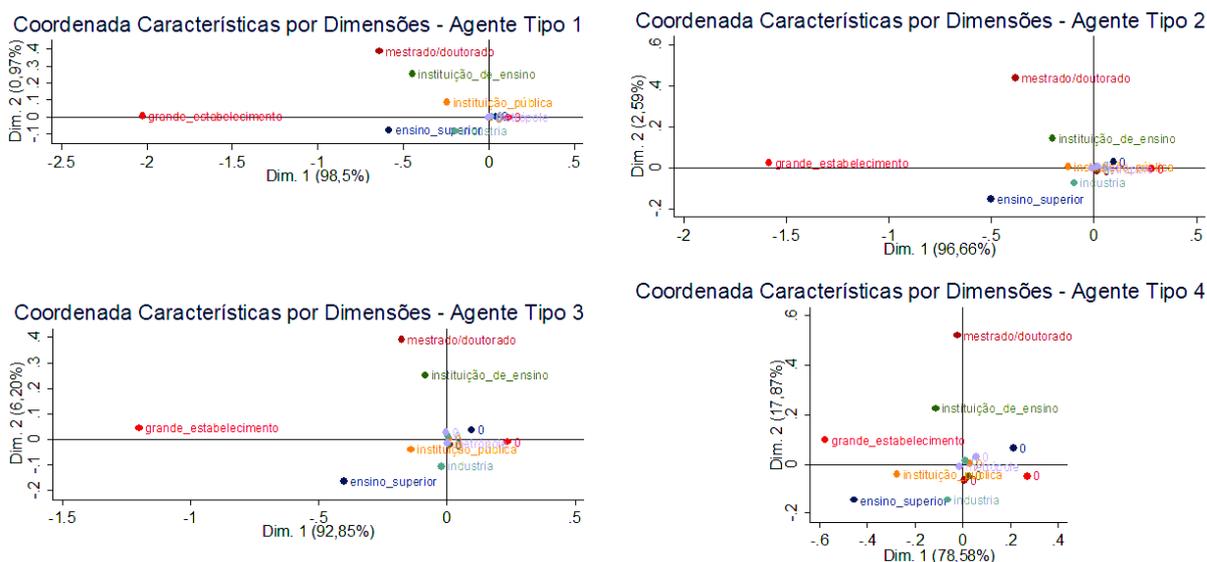
baixa. Os resultados obtidos a respeito das similaridades das respostas e composição das dimensões de variáveis podem ser evidenciados pela Figura 1 e Tabela 5.

É possível identificar por meio da Figura 1 que para o tipo de agente 1 (nó isolado), assim como agente tipo 2 (estrela interna) e agente tipo 3 (estrela externa) apenas o primeiro componente resume 90% das informações. Já para o tipo de agente tipo 4 é necessário considerar as duas dimensões para atingir um nível mínimo exigido de 90%.

Além disso, pode-se identificar que respostas sobre inserção em grande organização, setor industrial e aquisição de diploma em ensino superior são mais similares e são associadas ao agente ser capaz de produzir atividades inovativas sozinho. Isso é evidenciado por estas serem variáveis respostas presentes na primeira dimensão de variáveis para o grupo de agentes tipo 1, representando 98% das informações relevantes.

No segundo tipo de agente, as variáveis mais similares e associadas com a primeira dimensão representam aproximadamente 97% das informações características do agente tipo 2. Logo, variáveis respostas associadas ao agente estrela interna são: possuir ensino superior, pertencer ao setor de indústria e instituição pública, além de não estar inserido em uma grande organização.

**Figura 1 – Gráfico de coordenadas das características dos tipos de agentes no período 2006-2011**



Fonte: Elaboração Própria de acordo com resultados obtidos pelo STATA.

Já o agente tipo 3 (estrela externa) parece estar associado com mesmas variáveis respostas do agente tipo 2. Contudo, a sua primeira dimensão representa 92% das informações presentes na base de dados. Podendo ser estas características relacionadas a estes agentes estar associada ao fato de serem apenas agentes centrais ou no ambiente local (estrela interna) ou no ambiente externo (estrela externa).

E, finalmente, ao analisar a disposição das variáveis quanto a similares de variáveis para o tipo de agente *gatekeeper*, tem-se que as duas dimensões passam a ser relevantes para resumir as informações características dos mesmos. De modo que, apenas considerando-se 2 dimensões atinge-se o patamar de 90% de representação das informações do agente. A dimensão 1 é composta pelas mesmas variáveis destacadas para o tipo de agente 2 e 3: ensino superior completo, presença no setor industrial e presença em instituição pública. Enquanto, a dimensão 2 é composta das variáveis resposta: estar inserido em uma grande organização, pertencer a uma instituição de ensino e possuir título de mestrado e/ ou doutorado. Isso pode

ser uma evidência de que os agentes considerados *gatekeepers* possuem maior dinâmica de inserção no mercado de trabalho e maior nível de estudo, o que pode ser um fator relevante para que estes consigam absorver conhecimentos externos e disseminá-los ao seu local de origem. A Tabela 5 permite confirmar algumas dessas evidências.

**Tabela 5 – Estatísticas de Análise de Correlação Conjunta (2006-2011)**

Variável	Qualidade	% inércia	Dimensão 1 (sqcorr)	Dimensão 2 (sqrcorr)
<i>Agente tipo 1 – nó isolado</i>				
<i>Dum_esuperior</i>	0,99	0,08	0,98	0,02
<i>Dum_epos</i>	0,85	0,01	0,62	0,23
<i>Dum_iensino</i>	0,97	0,01	0,74	0,24
<i>Dum_ipub</i>	0,86	0,01	0,76	0,10
<i>Dum_ind</i>	0,90	0,01	0,77	0,13
<i>Dum_grand<sub>org</sub></i>	0,99	0,82	0,99	0,00
<i>Dum_metropole</i>	0,09	0,00	0,06	0,04
<i>Agente tipo 2 – estrela interna</i>				
<i>Dum_esuperior</i>	0,99	0,08	0,91	0,08
<i>Dum_epos</i>	0,99	0,02	0,43	0,56
<i>Dum_iensino</i>	0,84	0,01	0,56	0,28
<i>Dum_ipub</i>	0,45	0,00	0,45	0,00
<i>Dum_ind</i>	0,49	0,00	0,32	0,17
<i>Dum_grand<sub>org</sub></i>	0,99	0,73	0,99	0,00
<i>Dum_metropole</i>	0,80	0,00	0,68	0,12
<i>Agente tipo 3 – estrela externa</i>				
<i>Dum_esuperior</i>	0,99	0,10	0,90	0,14
<i>Dum_epos</i>	0,96	0,03	0,16	0,79
<i>Dum_iensino</i>	0,95	0,02	0,10	0,80
<i>Dum_ipub</i>	0,49	0,01	0,45	0,03
<i>Dum_ind</i>	0,55	0,01	0,02	0,53
<i>Dum_grand<sub>org</sub></i>	0,99	0,68	0,99	0,00
<i>Dum_metropole</i>	0,65	0,00	0,01	0,64
<i>Agente tipo 4 – gatekeeper</i>				
<i>Dum_esuperior</i>	0,99	0,22	0,91	0,09
<i>Dum_epos</i>	0,97	0,09	0,00	0,97
<i>Dum_iensino</i>	0,84	0,04	0,18	0,67
<i>Dum_ipub</i>	0,67	0,03	0,66	0,01
<i>Dum_ind</i>	0,42	0,01	0,07	0,35
<i>Dum_grand<sub>org</sub></i>	0,99	0,32	0,97	0,03
<i>Dum_metropole</i>	0,67	0,00	0,50	0,17

Nota: Os resultados são referentes a *dummy* = 1.

Fonte: Elaboração Própria de acordo com resultados obtidos pelo STATA.

Investigando-se os resultados tem-se que para o agente 1 as variáveis destacadas em termos de qualidade da representação são as mesmas destacadas pela Figura 1, a saber: pertencer a uma grande organização, instituição de ensino, setor industrial e possuir ensino superior completo, sendo as com maior escore (>0,90) possuir ensino superior e estar presente em uma grande organização.

No caso do agente do tipo 2, as variáveis resposta que contribuem para a qualidade da representação dos dados são as *dummies* de ensino superior, titulação em mestrado e/ou doutorado e presença em grande organização. A presença em grande organização e possuir ensino superior completo são as variáveis de maior escore na dimensão 1.

Para o agente do tipo 3, destacam-se como responsáveis pela maior qualidade as variáveis resposta pertencer a instituição de ensino, possuir ensino superior completo, possuir título em mestrado e doutorado e pertencer a uma grande organização. Contudo, as variáveis que mais colaboram para a primeira dimensão é pertencer a uma grande organização e possuir

ensino superior completo. As outras duas variáveis são as de maior escore apenas considerando-se a segunda dimensão.

E, considerando o grupo dos *gatekeepers*, tem-se que as variáveis respostas responsáveis pela qualidade ( $>0,90$ ) são as *dummies* de ensino superior, título em mestrado e doutorado e pertencer a uma grande organização. Segundo, destaca-se que as variáveis de maior escore para a primeira dimensão são as variáveis *dummies* de ensino superior e grande organização, sendo a *dummy* de pertencimento à instituição de ensino a variável de maior escore na segunda dimensão.

Ademais, considerando-se a participação das variáveis quanto à inércia total, i.e, que visa obter informação se a associação entre as variáveis são estatisticamente significativas, toma-se como referência um coeficiente acima de 0,20. Tem-se que para o caso do agente tipo 1 (nó isolado) que estar inserido em uma grande organização já representa 82% da inércia total. Para o agente tipo 2 (estrela interna), esse coeficiente sobe para 83%. Já para o agente tipo 3 (estrela externa) o coeficiente perde sua participação sobre a inércia total, atingindo um valor de 86% desta. Todavia, no caso dos *gatekeepers* tem-se que possuir ensino superior tem participação de 22% sobre a inércia enquanto estar presente em uma grande organização perde ainda mais sua participação sobre o parâmetro, em relação aos demais agentes, atingindo um valor de 32%. Resultados a respeito da relação entre título de ensino superior completo e os tipos de agentes na rede também foi evidenciado na literatura, principalmente para o caso destes serem *gatekeeper*, encontrando-se uma menor relação entre nível de conhecimentos e os nós isolados (GIULIANI; BELL, 2011). Outras evidências também encontradas pela literatura são de que haja associação entre *gatekeepers* e pertencimento destes às instituições públicas (GRAFF, 2011) e que o tamanho da equipe seja relevante para o agente *gatekeeper* ou estrela externa (LEGALLO; PLUNKET, 2016). No entanto, a presente pesquisa ainda apontou associações entre características e tipos de agentes, além das já investigadas pela literatura.

A fim de dar explorar melhor os resultados obtidos pelas técnicas de análise multivariada, a Tabela 6 reúne os resultados para cada tipo de agente por meio de dados em painel e uso do modelo *logit* com efeitos fixos. Por se considerar a existência de efeitos fixos nos dados, optou-se por excluir da amostra as variáveis que se referem a características fixas dos agentes ao longo do tempo, como a *dummy* de região metropolitana e de nível educacional de ensino superior. As demais variáveis *dummies* foram mantidas por serem consideradas como mais passíveis de alteração ao longo dos anos seja por aquisição de título educacional mais elevado ou por mudança de organização de trabalho, dado que se considera 11 anos de análise.

É possível identificar por meio da Tabela 6 que a probabilidade de pertencer a cada um dos grupos de agentes possíveis de se existir nas redes de inovação no Brasil está atrelada a características distintas dos agentes. Para o G1 – nós isolados, há evidências de que conhecimentos em patentes de alta tecnologia contribuem negativamente para que o agente inove de maneira isolada. O que pode estar atrelado à geração de inovações de alta tecnologia estar associada a maior troca de conhecimento entre os agentes, enquanto a produção independente ser associada a um conhecimento tecnológico mais básico. Além disso, encontra-se evidências de que o número de parceiros distintos (*tamanho\_equipe*) tem uma associação negativa com a probabilidade do agente se comportar como um inventor isolado. Outra evidência encontrada para o G1 é que tais agentes tem uma relação negativa com o pertencimento a uma instituição pública (*dum\_ipub*) e grande organização (*dum\_grand\_org*), e uma associação positiva em pertencer a uma organização industrial (*dum\_ind*), sendo as demais variáveis não significativas.

**Tabela 6 – Determinantes do tipo de agente inovador no Brasil. Modelo *logit* de efeitos fixos (2000-2011)**

Variáveis	G1	G2	G3	G4
<i>Grau_high</i>	-0,71**	0,16	0,32*	0,13
<i>Grau_diversificação</i>	0,12	0,11	-0,22	-0,39**
<i>Tamanho_equipe</i>	-12,04***	0,73***	1,11***	2,82***
<i>Dum_esuperior</i>	-	-	-	-
<i>Dum_epos</i>	-0,82	-0,53***	0,16	0,59***
<i>Dum_iensino</i>	-0,54	-0,18**	-0,22	0,49***
<i>Dum_ipub</i>	-0,71**	-0,14	-0,09	0,40***
<i>Dum_ind</i>	0,52**	0,05	-0,02	-0,19
<i>Dum_grand<sub>org</sub></i>	-0,45*	-0,26***	-0,09	0,41***
<i>Dum_metropole</i>	-	-	-	-
<i>LR chi</i>	3995,09***	144,97***	128,17***	1075,01***
<i>Log like.</i>	-559,80	-3185,47	-1576,19	-1683,42

Nota<sup>1</sup>: \*, \*\*, \*\*\* denotam respectivamente os níveis de significância de 10%, 5% e 1%.

Fonte: Elaboração própria de acordo com as saídas do STATA.

Para o tipo de agente estrela interna (G2), encontram-se evidências de que possuir ensino de maior qualidade como pós ou mestrado (*dum\_epos*), assim como pertencer a uma instituição de ensino (*dum\_iensino*) ou grande organização (*dum\_grand<sub>org</sub>*), diminuem as chances de o agente vir a se tornar uma estrela interna. Os resultados também sugerem que um grande número de parceiros distintos (*tamanho\_equipe*) tem uma associação positiva com o seu pertencimento ao G2.

Como resultados para o tipo de agente (G3), tem-se que a probabilidade de os agentes se comportarem como estrela externa tem uma relação positiva com estes possuírem conhecimentos em patentes de alta tecnologia (*grau\_high*) e possuir parceiros maior número de parceiros distintos (*tamanho\_equipe*). Ademais, ressalta-se que apenas estas variáveis foram significativas para explicar as chances de pertencimentos dos agentes ao G3.

Como evidência da probabilidade de pertencimento do agente ao grupo de *gatekeepers* (G4), destaca-se a associação positiva com o número de parceiros distintos (*tamanho\_equipe*). Tal resultado corrobora a hipótese 1 de que exista relação entre a capacidade de absorção do agente e a probabilidade de este se ligar a demais agentes na rede e potencialmente ser um *gatekeeper*.

Outro resultado encontrado para o G4 é de que haja uma relação negativa entre o agente possuir conhecimento diversificado (*dum\_diversificação*) e ser um *gatekeeper* na rede brasileira. Fato que pode estar associado aos inventores brasileiros serem mais especializados e possuírem conhecimentos mais restritos em poucas tecnologias, em geral. Ademais, também sugere-se que os agentes que pertençam a grandes organizações (*dum\_grand<sub>org</sub>*), instituições públicas (*dum\_ipub*) e instituições de ensino (*dum\_iensino*) são mais propensos a se comportarem como *gatekeepers*. Os dois últimos resultados, de pertencer a instituição pública e de ensino vão de acordo com a segunda e terceira hipótese da presente pesquisa, que pressupõe que haja elevação nas chances de agente pertencentes a tais organizações serem *gatekeepers*. E por fim, encontra-se que possuir grau de ensino mais elevado com mestrado e/ou doutorado (*dum\_epos*) também tem uma associação positiva com o agente ser um *gatekeeper*.

## 5. Conclusões

As técnicas utilizadas neste trabalho possuem objetivos distintos. No entanto, juntas possibilitaram a exploração e um melhor entendimento da disposição das variáveis de acordo com os possíveis tipos de agentes na rede de inovação do Brasil.

Como resultado, pode-se concluir que os grupos de classificação pré-definidos pela literatura são consistentes com o caso brasileiro. Tal conclusão está atrelada à qualidade da discriminante em atingir taxas de erros de classificação consideradas baixas em uma amostra de tamanho elevado e obter estatísticas significativas de que tais grupos possuem médias distintas entre si. O método possibilitou evidenciar que as médias entre as variáveis considerando os grupos distintos são diferentes. Além disso, o grau de conhecimento em alta tecnologia e conhecimentos diversificados, assim como o tamanho de equipe e pertencimento à região metropolitana, foram moderadamente maiores no caso dos *gatekeepers*. Também foram constatados resultados acima da média para o número de inventores que possuem ensino superior completo e título de mestrado e/ doutorado, além de pertencerem a grandes organizações, instituição pública e/ou à instituição de ensino. Todavia, constatou-se que inventores pertencentes ao setor industrial são mais propensos a serem classificados no grupo dos agentes estrelas externas, já que obtiveram maior média neste grupo em comparação aos demais.

Além disso, método *logit* multinomial sustenta tais resultados ao evidenciar significância entre o conhecimento em alta tecnologia e o pertencimento de agentes ao grupo de *gatekeepers*. Variáveis, como pertencer ao setor industrial, o tamanho da equipe e ter ensino superior, apresentaram resultados positivos com todos os grupos. No entanto, para a primeira variável mencionada, isso só é válido considerando-se o período mais recente e, para a última, o período anterior. Além disso, pertencer a uma região considerada metrópole tem associação negativa com o agente ser estrela externa.

O modelo de análise de correspondência conjunta atinge os referenciais de qualidade da representação da dimensão das variáveis categóricas e supõe que os nós isolados são aqueles que pertencem a uma grande organização, instituição de ensino, setor industrial e possuem ensino superior completo. Já para estrelas internas, as variáveis que contribuem para a representação dos dados dos agentes evidenciam que os mesmos possuem ensino superior completo, titulação em mestrado e/ou doutorado e presença em grande organização. Estrelas externas supostamente fazem parte de instituição de ensino e grande organização, e possuem ensino superior completo e/ou título em mestrado e doutorado. E, finalmente, considerando o grupo dos *gatekeepers*, tem-se que estes possuem ensino superior completo e título em mestrado e doutorado e pertencem a uma grande organização. Isso pode ser uma evidência de que agentes centrais em rede, i.e. centrais em sua região (estrela interna), centrais externamente (estrela externa) ou com centralidade de intermediadora (*gatekeepers*) possuem maior nível de conhecimento do que aqueles que não se ligam, que são nós isolados.

Contudo, o uso de técnica de análise multivariada não pressupõe relações causais entre as variáveis e fenômenos observados e apenas sugere uma associação entre as mesmas. Portanto, o uso do modelo *logit* com efeitos fixos é utilizado como uma ferramenta para dar suporte a hipóteses e controlar os termos de erros não observados. Há indícios de que os nós isolados são agentes que pertencem à indústria e que as estrelas externas possuem maior conhecimento em patentes de alta tecnologia. Já os *gatekeepers* são agentes com associações positivas com a capacidade de absorção de conhecimento, pertencimento a instituições públicas e organizações de ensino, confirmando as três hipóteses investigadas pelo presente estudo.

Como futuras investigações, considerando que o maior foco das análises de redes é sobre os agentes que se comportam como *gatekeepers*, e com pauta na pequena taxa de erro de classificação, pode-se investigar isoladamente o grupo de agentes pré-classificados como *gatekeepers*. Isso permite refinamentos da medida de *gatekeeper* existente na literatura e a consistência da classificação e caracterização dos mesmos diante as variações das medidas pré-existentes. Além disso, pode-se investigar quais agentes mudaram de grupo ao longo dos anos e quais variáveis afetaram tal comportamento. Dado que, os resultados obtidos por esta

pesquisa apontam que não somente as ligações internas e externas são capazes de explicar por si só o motivo de os agentes exercerem cada um dos papéis destacados na rede. É possível pesquisar se os níveis educacionais maiores e o pertencimento a instituições que motivem seus agentes a colaborar mais são de fato promotoras de tais funções. Por último, outra extensão de análise viável seria utilizar outras abrangências geográficas buscando identificar os agentes que possuem maior alcance de ligações em termos de ligações externas e captação de conhecimentos distintos. E, além disso, podendo-se explorar mais as características do local onde se inserem e uma consequente facilitação de interação com demais agentes.

## Referências

ALLEN, T. J. **Roles in technical communication networks**. Massachusetts, MIT, 1969.

ALLEN, T. J.; COHEN, S. I. Information flow in research and development laboratories, **Administrative Science Quarterly**, n. 14, p. 12–19, 1969.

BATHELT, H.; MALMBERG, A.; MASKELL, P. Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation, **Progress in Human Geography**, v. 28, n. 1, p. 31–56, 2004.

BOSCHMA, R. Proximity and innovation: a critical assessment. **Regional Studies**, v. 39, n. 1, p. 61-74, 2005.

BRESCHI, S.; LENZI, C. The Role of External Linkages and Gatekeepers for the Renewal and Expansion of US Cities' Knowledge Base, 1990 – 2004, **Regional Studies**, v.49, n.5, p. 782-797, 2015.

CHESBROUGH, H. W. The era of open innovation. **Managing innovation and change**, v. 127, n. 3, p. 34-41, 2006.

COHEN, W.; LEVINTHAL, D. Absorptive capacity: a new perspective on learning and Innovation. **Administrative Science Quarterly**, v. 35, n. 1, p. 128–52, 1990.

GITTELMAN, M. Does geography matter for science-based firms? Epistemic communities and the geography of research and patenting in biotechnology. **Organization Science**. n. 18, p. 724–741, 2007

GIULIANI, E. Role of technological gatekeepers in the growth of industrial clusters: Evidence from Chile. **Regional Studies**, p. 1329-1348, v.45, n.10, 2011.

GIULIANI, E.; BELL, M. The micro-determinants of meso-level learning and innovation: evidence from a Chilean wine cluster, **Research Policy**, v. 34, n. 1, p. 47–68, 2005.

GOULD, R. V.; FERNANDEZ, R. M. Structures of mediation: a formal approach to brokerage in transaction networks. **Sociological Methodology**, v. 19, p. 89-126, 1989.

GRAF, H. Gatekeepers in regional networks of innovators. **Cambridge Journal of Economics**, n. 35, p. 173–198, 2011.

GRAF H., KRUGER, J. J. The performance of gatekeepers in innovator networks. **Jena economic research papers**, n. 58, 2009.

JOHNSON, R. A.; DEAN, W. W. **Applied multivariate statistical analysis**. London: Prentice Hall, 2007.

LEGALLO, J.; PLUNKET, A. Technological gatekeepers, regional inventor networks and inventive performance. **HAL**, 2016.

MORRISON, A.; RABELLOTI, R.; ZIRULIA L. When do global pipelines enhance the diffusion of knowledge in clusters? **Economic Geography**, v.89 n.1, p. 77-96, 2013.

NELSON, R. R., **National innovation systems: a comparative analysis**. Oxford University Press on Demand, 1993.

POWELL, W. W.; GIANNELLA, E. Collective invention and inventor networks. **Handbook of the Economics of Innovation**, v. 1, p. 575-605, 2010.

TUSHMAN, M. L.; SCANLAN, T. J. Boundary spanning individuals: Their role in information transfer and their antecedents. **Academy of management journal** v. 24, n.2, p. 289-305, 1981.

WOOLDRIDGE, J. M. **Econometric analysis of cross section and panel data**. MIT press, 2010.

## APÊNDICE

**A1 – Matriz de Correlação entre as variáveis (2000-2011)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) <i>Grau_high</i>	1									
(2) <i>Grau_diversificação</i>	0,20	1								
(3) <i>Tamanho_equipe</i>	0,23	0,01	1							
(4) <i>Dum_esuperior</i>	0,09	0,01	0,24	1						
(5) <i>Dum_epos</i>	0,11	0,03	0,18	-0,07	1					
(6) <i>Dum_iensino</i>	0,15	0,06	0,20	-0,06	-0,03	1				
(7) <i>Dum_ipub</i>	0,07	0,03	0,10	-0,06	-0,03	0,02	1			
(8) <i>Dum_ind</i>	-0,04	-0,02	0,08	-0,08	-0,04	-0,04	-0,06	1		
(9) <i>Dum_grand<sub>org</sub></i>	0,10	-0,01	0,26	0,50	0,23	0,08	0,01	0,06	1	
(10) <i>Dum_metropole</i>	0,04	0,02	0,04	0,01	-0,01	0,01	-0,01	-0,02	0,01	1