

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

ALEXANDRE NOVAES MEJDALANI

**OS DETERMINANTES DAS REDES BRASILEIRAS DE
COPATENTEAMENTO**

JUIZ DE FORA

2017

ALEXANDRE NOVAES MEJDALANI

OS DETERMINANTES DAS REDES BRASILEIRAS DE COPATENTEAMENTO

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Faculdade de Economia da Universidade Federal de Juiz de Fora como parte das exigências para obtenção do Título de Mestre.

Orientador: Prof. Dr. Eduardo Gonçalves

Coorientador: Prof. Dr. Emilson Caputo Delfino Silva

Juiz de Fora

2017

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Mejdalani, Alexandre Novaes.

OS DETERMINANTES DAS REDES BRASILEIRAS DE COPATENTEAMENTO / Alexandre Novaes Mejdalani. -- 2017. 92 p.

Orientador: Eduardo Gonçalves

Coorientador: Emilson Caputo Delfino Silva

Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Faculdade de Economia. Programa de Pós-Graduação em Economia, 2017.

1. Redes de Inventores. 2. Copatentes. 3. Análise de Redes Sociais. I. Gonçalves, Eduardo, orient. II. Silva, Emilson Caputo Delfino, coorient. III. Título.

ALEXANDRE NOVAES MEJDALANI

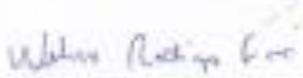
OS DETERMINANTES DAS REDES BRASILEIRAS DE
COPATENTEAMENTO

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Faculdade de Economia da Universidade Federal de Juiz de Fora como parte das exigências para obtenção do título de Mestre em Economia Aplicada.
Área de Concentração: Economia.

Aprovada em:



Prof. Dr. Eduardo Gonçalves - Orientador
Universidade Federal de Juiz de Fora



Prof. Dr. Weslem Rodrigues Faria
Universidade Federal de Juiz de Fora



Prof. Dr. Eduardo Amaral Haddad
Universidade de São Paulo

AGRADECIMENTOS

Não é sucinta a lista de quem devo agradecer pela execução deste trabalho. Tampouco esta curta seção fará jus a tamanha gratidão que devo. Tenho sorte de me pesar por essas palavras iniciais. Pego-me então refletindo sobre a importância de expressar tal gratidão em um papel que será provavelmente esquecido e empoeirado no fundo de uma biblioteca ou perdido em uma montanha digital. Faço um exercício comum: se em alguns séculos, na derrocada social, encontrarem escombros de uma biblioteca e um maço de papel perdido, que estas pessoas sejam também lembradas.

Duas mulheres merecem não apenas meu agradecimento, mas a dedicatória pessoal de todo o trabalho: minha mãe Telma *in memoriam*, meu Norte, e minha namorada Patricia, minha bússola. Agradeço a minha mãe por me dar os primeiros rabiscos de preocupação social e compaixão. Se um dia puder caminhar o mundo apenas um passo para o Norte que me ensinara, já serei realizado. Agradeço a Patricia por nunca deixar que eu me perdesse, por secar meus choros, afastar meus medos e acreditar no meu – e nosso – futuro. Nunca a distância nos parou.

Agradeço ao meu pai – o comandante dessa grande embarcação -, quem primeiro me apresentou a ciência. Agradeço ainda minha madrastra Gisele e meus incontáveis irmãos: Diego, Giovanna, Rickson e Rian. Me desculpo por cada mês que não pude ir pra casa na certeza de que os melhores votos eram enviados de vocês. À minha segunda família, agradeço por cada dia de abrigo, acolhimento e carinho dado nesse período: Dóris, Célia, Fabio, Jaqueline e Maria.

Este trabalho não seria possível sem a dedicação de dois professores que, com enorme sabedoria e conhecimento, puderam me introduzir na ciência aplicada. Agradeço ao meu orientador Prof. Eduardo Gonçalves, que como um maestro incansável orquestrou o compasso do trabalho. Ao Prof. Fernando Perobelli, que me deu outros olhos aos problemas.

Agradeço à FAPEMIG, pelo apoio financeiro ao projeto. Agradeço à equipe responsável pela base de dados proveniente do INPI, sem a qual esse trabalho não seria possível. À Juliana Gonçalves, pelo imenso auxílio na elaboração da base de dados. E a todos os funcionários da Faculdade de Economia, que prestaram suporte durante todo o período do curso.

Ao longo desses dois anos, muitas pessoas fizeram parte da minha vida juiz-forana. A estas, apenas tenho a agradecer por cada momento de acolhimento, dedicação e conversas na sala da pós. Agradeço os amigos (em ordem alfabética) Alexandre Rabelo, Ananias Vitor, Addressa Eugênio, Carlos Dias, Christian Herrera, Daniele Oliveira, Iara Alkmin, Inácio Fernandes, Izak Carlos, Raquel Coelho e Rodrigo Rodriguez.

Não poderia deixar nunca de esquecer dos amigos do Amaral Peixoto Café, que em conversas profundas sobre assuntos banais aliviaram a pressão do dia-a-dia: Adriana, Bruna, Carois, Daniel, Flaviana, Juliana, Ludmilla, Mariana, Matheus, Rafael, Roberta e Sidney. Agradeço também meus companheiros de casa, Gabriel e Ariel, por cada conversa regada a pipoca, e meu amigo de sempre, Pedro, por cada bossa.

Meu sincero obrigado.

RESUMO

Os estudos que exploram o processo e determinantes para a formação da rede se inserem no âmbito do debate acerca do papel da cooperação entre inventores para o benefício do sistema de inovação. O objetivo desta dissertação é investigar quais são os determinantes da formação de laços de copatenteamento no Brasil, utilizando a primeira década do século XX (2000-2011) nesta análise. Tradicionalmente, os trabalhos empíricos sobre o tema buscaram na geografia econômica um arcabouço para investigar o papel de espaço na cooperação. Mais recentemente, o uso de instrumentos de análise de redes sociais se mostrou útil na identificação de processos internos à topologia da rede. Nesta dissertação, ambos os arcabouços são utilizados, combinando instrumentos analíticos espaciais e relacionais. Duas óticas são dadas ao problema: como se conectam as regiões brasileiras e como se conectam os inventores brasileiros. Para isto, são utilizadas as bases de dados da Base de Dados Estatísticas de Propriedade intelectual (BADEPI) do INPI, e a base da Relação Anual de Informações Sociais (RAIS) do Ministério do Trabalho. Os principais resultados indicam que: (1) a geografia ainda possui um papel fundamental na formação de redes, (2) a rede e seus subcomponentes estão espacialmente concentrados, (3) existe um padrão de aglomeração geográfica desigual no sistema brasileiro de invenções, (4) a proximidade tecnológica e a proximidade relacional são fatores importantes na determinação de novas conexões e (5) inventores buscam similaridade socioeconômica para formar novas conexões, embora haja benefícios nas conexões com indivíduos mais ricos e bem posicionados na rede.

Palavras-chave: Redes de Invenção; Análise de Redes Sociais; Copatenteamento.

ABSTRACT

The study of processes and causes of how networks are formed emerges in the context of the debate about the role of inventors' cooperation and its benefits to innovation systems. The aim of this dissertation is to investigate the determinants of co-patents ties formation in Brazil in the first decade of the 20th century (2000-2011). Traditionally, the empirical works on the theme used the framework of Economic Geography to study the role of space in nodes cooperation. Recently, the use of social network analysis instruments proved to be useful to identify internal processes in network topology. In this dissertation, both frameworks are used, combining the analytical instruments of spaces and relations. The problem has two viewpoints: how regions connect to each other and how inventors connect to each other. The databases used are the Database on Statistics of Intellectual Property (BADEPI) from INPI, and the Annual Report of Social Information from Labour Ministry. The main results indicate that: (1) geography still plays a fundamental role in forming networks; (2) the network and its subcomponents are spatially concentrated; (3) there is an unequal geographical agglomeration patterns in the Brazilian invention system; (4) technical and relational proximity are important factors in determining new connections and (5) inventors seek for similar socioeconomic partners, although there are benefits of cooperation with players there are wealthier and better positioned in the network.

Keywords: Invention Networks; Social Network Analysis; Co-patenting

LISTA DE SIGLAS

BADEPI - Base de Dados Estatísticos de Propriedade Intelectual

CO - Região Centro-Oeste

CPF - Cadastro de Pessoa Física

EA - Painel de Efeitos Aleatórios

EF - Painel de Efeitos Fixos

EPO - European Patent Office

ERM - Modelo de Érdos-Renyi

GLS - Generalized Least Squares

IBGE - Instituto Brasileiro de Geografia e Estatística

INPI - Instituto Nacional de Propriedade Intelectual

IPC - International Patent Classification

MTE - Ministério do Trabalho e Emprego

N - Região Norte

NBRM - Modelo de Regressão Binomial Negativo

NE - Região Nordeste

P&D - Pesquisa e Desenvolvimento

PIB - Produto Interno Bruto

PNAD - Pesquisa Nacional por Amostra de Domicílios

PPML - Poisson Pseudo-Maximum Likelihood Estimation

RAIS - Registro Anual de Informações Sociais

REGIC - Região de Influência das Cidades

S - Região Sul

SE - Região Sudeste

LISTA DE FIGURAS E GRÁFICOS

Figura 1. Exemplo de grafo com 5 vértices e 5 arestas	31
Figura 2. Mapa de agrupamento do exemplo da Figura 1	35
Figura 3. Esquema de Construção da Base Regional	40
Figura 4. Primeiro procedimento de conversão de bases	41
Figura 5. Segundo procedimento para criação de base de copatentes	42
Figura 6. Mapas de Redes Trienais de Copatentes, REGICs Brasil, 2000-2011	46
Gráfico 1. Histograma da distância dos laços intermunicipais, ponderado pelo número de conexões, Brasil, 2000-2011.....	55

LISTA DE TABELAS E QUADROS

Quadro 1. Relação de sinal dos coeficientes de dominância	38
Quadro 2. Descrição das bases de dados auxiliares utilizadas na análise regional, 161 REGICS, 2000-2011	39
Quadro 3. Classificação das características individuais utilizadas, 2000-2011	49
Quadro 4. Classificação de variáveis do laço e do nó sob a ótica individual	51
Quadro 5. Descrição das variáveis do modelo sob a ótica do inventor	76
Tabela 1. Centralidades calculadas para o exemplo da Figura 1.....	34
Tabela 2. Centralidades de Vizinhança calculadas para o exemplo da Figura 1.....	35
Tabela 3. Relação das regiões (REGICs) mais conectadas a outras regiões, Brasil, 2000-2011.....	44
Tabela 4. Regiões (REGICs) com mais laços, Brasil, 2000–2011	44
Tabela 5. Distribuição macrorregional dos nós sem conexões ($y = 0$) e com conexões ($y = 1$) em todos os triênios, Brasil, 2000-2011	45
Tabela 6. Estatísticas da rede agregada por REGIC, Brasil, 2000-2011	47
Tabela 7. Correlação de características nodais das regiões, 2000-2011	48
Tabela 8. Distribuição trienal dos nós, laços e grau médio, inventores, 2000-2011	49
Tabela 9. Distribuição de Ligação entre Inventores por Gênero no Brasil, por triênio, 2000-2011	51
Tabela 10. Distribuição do número de pares formados entre indivíduos, por ano de estudo, 2000-2010	53
Tabela 11. Distribuição do número de pares formados entre indivíduos por quantil de renda (em salários mínimos), 2000-2011	53
Tabela 12. Correlação entre máximos e mínimos de educação e renda no laço	54
Tabela 13. Distribuição trienal de frequência por conexões entre empresas e universidades, Brasil, 2000-2011	54
Tabela 14. Distribuição trienal de frequência por conexões entre indivíduos do mesmo município ou de municípios diferentes, Brasil, 2000-2011	55
Tabela 15. Percentual de Laços Formados entre REGICs por Triênio, Brasil, 2000-2011	57

Tabela 16. Percentual de Laços Formados entre REGICs por Triênio, regiões não isoladas, Brasil, 2000-2011	57
Tabela 17. Média e Desvio Padrão das ligações por Triênio, Brasil, 2000-2011	58
Tabela 18. Resultado das Estimações do Painei Inter-regional, 2000-2011	64
Tabela 19. Resultado das Estimações Trienais Inter-regionais, Brasil, 2000-2011	65
Tabela 20. Estimação dos coeficientes de dominância, ótica regional	66
Tabela 21. Autocorrelação das ligações entre os períodos, Brasil, 2000-2011	69
Tabela 22. Distribuição trienal dos laços de controle criados no ERM, Brasil, 2000-2011	72
Tabela 23. Resultado das estimações PPML e BNRM (y: CONT), Brasil, 2000-2011	76
Tabela 24. Cálculo de Coeficientes de Dominação	78
Tabela 25. Resultado das estimações trienais por PPML, ótica individual, 2000-2011	80

Sumário

1 INTRODUÇÃO	12
2 REVISÃO DE LITERATURA	14
2.1 INOVAÇÃO, TERRITÓRIO E DESENVOLVIMENTO	14
2.2 A IMPORTÂNCIA DO ESTUDO DAS REDES NO PROCESSO DE DIFUSÃO DE CONHECIMENTO TECNOLÓGICO	16
2.3 FORMAÇÃO DE REDES DE COLABORAÇÃO.....	17
2.3.1 Incentivos à cooperação	17
2.3.2 Espaço de lugares <i>versus</i> espaço de fluxos	18
2.3.3 Cooperação entre indivíduos e entre regiões: dois níveis de análise.....	20
2.4 EVIDÊNCIAS EMPÍRICAS DO PAPEL DAS REDES NA DIFUSÃO E SEUS DETERMINANTES	24
2.5 INOVAÇÃO E REDES DE COLABORAÇÃO NO BRASIL	27
2.6 COPATENTES COMO INDICADOR DE DIFUSÃO POR COLABORAÇÃO	29
3 METODOLOGIA GERAL, BASE DE DADOS E DESCRIÇÃO DA REDE	30
3.1 DEFINIÇÕES FORMAIS DE GRAFOS E REDES SOCIAIS	30
3.2 TRATAMENTO DE CARACTERÍSTICAS NODAIS.....	36
3.3 BASE DE DADOS E CONSTRUÇÃO DA REDE.....	38
3.3.1 Construção da rede regional.....	40
3.3.2. Construção da rede individual.....	41
3.4 DESCRIÇÃO DA REDE BRASILEIRA DE COPATENTEAMENTO	42
3.4.1 Análise exploratória da rede ao nível regional.....	43
3.4.2 Análise exploratória da rede ao nível individual.....	48
4 DETERMINANTES DA FORMAÇÃO DE LAÇOS INTER-REGIONAIS	56
4.1 MÉTODO DE ESTIMAÇÃO	56
4.2 DESCRIÇÃO DAS VARIÁVEIS.....	58
4.3 MODELO EMPÍRICO.....	61
4.4 RESULTADOS DA ESTIMAÇÃO.....	62
5 DETERMINANTES DA REDE BRASILEIRA DE INVENTORES	69
5.1 MÉTODO DE ESTIMAÇÃO	69
5.1.1 Construção do contrafactual.....	70
5.1.2 Estimador PPML para eventos raros	73
5.2 VARIÁVEIS E MODELO EMPÍRICO.....	74
5.3 RESULTADOS E INTERPRETAÇÕES.....	75
5.4 INTERPRETAÇÕES GERAIS, LIMITAÇÕES E CONCLUSÕES	81
CONCLUSÕES	82
REFERÊNCIAS BIBLIOGRÁFICAS	86

1 INTRODUÇÃO

A figura do inventor como um peculiar polímata que desenvolve sozinho suas ideias está longe de ser uma realidade na atual configuração da atividade inventiva. Este dito “inventor de garagem” é um espécime cada vez mais raro desde a revolução nas comunicações descrita por Castells (1996). Fato é que a invenção fruto da cooperação entre inventores é cada vez mais comum, assim como se amplia o tamanho das equipes de pesquisa e desenvolvimento (P&D) (WUCHTY *et al.*, 2007). Por outro lado, atributos tradicionalmente associados à atividade inventiva (JAFFE, 1989), como o papel do espaço físico para a difusão de informação e conhecimento, também atuam na atividade de invenção cooperativa (SINGH, 2005). Assim, a compreensão da inovação como um fenômeno sociodemográfico encontra espaço no debate sobre a forma como a inovação ocorre, o porquê ocorre e quais os efeitos deste fenômeno.

Os trabalhos em inovação, e mais recentemente em redes de inovação, buscaram ampliar o entendimento teórico da atividade e incorporar novos instrumentos na análise empírica do problema. O clássico “ar” marshalliano (MARSHALL, 1920) que propiciava o desenvolvimento regional foi gradualmente sendo adaptado a fim de se entender o espaço econômico como um espaço social que, por um lado, modifica o espaço físico, e, por outro lado, adapta-se a este (SAXENIAN, 1994). Trabalhos seminais no entendimento de redes de inovação, como Rogers (1962), estabeleceram metodologias analíticas que foram aprimoradas com a melhora da capacidade computacional e das bases de dados publicamente disponíveis. Já outro eixo da literatura, reforçada pela Nova Geografia Econômica, buscou aprofundar o estudo do papel do espaço físico na atividade inventiva (JAFFE, 1989; KRUGMAN, 1991; GLAESER *et al.*, 1992; JAFFE, TRAJTENBERG e HENDERSON, 1993).

O ressurgimento do interesse na literatura de redes aplicado aos estudos em inovação se deu a partir da década de 1980 e, com maior destaque, a partir dos anos 2000 (Cowan, 2005). Embora ainda muito relacionados ao papel da geografia na rede, estes novos estudos absorveram a literatura de análise de redes sociais, das ciências cognitivas e da economia institucional. Esta abordagem, baseada no conceito de “proximidades” de Torre e Gilly (1999) e Boschma (2005), fomentou estudos multidimensionais sobre a natureza inventiva das firmas, regiões e – ainda mais recentemente – dos indivíduos. Contudo, embora tenha sido explorada a topografia da rede e sua distribuição espacial, pouco se explorou acerca das motivações endógenas à cooperação.

No Brasil, estudos de inovação pela ótica geográfica foram bastante explorados, sobretudo por Albuquerque *et al.* (2002), Lemos *et al.* (2006) e Gonçalves e Almeida (2009). Nestas análises, características fundamentais de distribuição e desigualdade regional no desempenho da atividade inovadora e da difusão de conhecimento foram identificadas. Por outro lado, o estudo de redes permanece restrito, sobretudo devido à dificuldade no acesso a bases de dados completas. Os principais trabalhos em redes de pesquisa e inovação são as contribuições de Sidone, Haddad, e Mena-Chalco (2016a, 2016b), em cientometria, e Oliveira (2016) sobre o papel das redes na capacidade de invenção.

O objetivo geral desta dissertação é contribuir para a identificação dos motivos endógenos – ou determinantes – que levam à cooperação na atividade inventiva. Para isto, dois níveis de análise são propostos com objetivos particulares. Primeiro, propõe-se a analisar a rede formada entre regiões a fim de responder quais características regionais tornam uma região mais atrativa à cooperação e facilitam a troca informacional. Segundo, serão investigados os determinantes pessoais que levam o inventor a cooperar. Nesta abordagem, busca-se compreender como a percepção de ganhos do indivíduo afeta sua decisão de cooperar ou não.

Duas bases de dados serão utilizadas: (1) Banco de Dados de Propriedade Intelectual (BADEPI), do Instituto Nacional de Propriedade Intelectual (INPI) e (2) Relação Anual de Informações Sociais (RAIS), do Ministério do Trabalho e Emprego (MTE). Bases auxiliares serão utilizadas e apresentadas ao seu tempo ao longo da dissertação. Será utilizado o período de 2000 a 2011, dividido em quatro triênios. Para a abordagem regional, utilizar-se-á o conceito de Regiões de Influência das Cidades (REGIC) proposto pelo IBGE (2013). Esta divisão geográfica possibilita capturar regiões centrais e seu entorno polarizado. Para a ótica individual, são utilizados inventores formalmente empregados no mesmo ano de registro da patente.

Além desta introdução, a dissertação segue por mais quatro capítulos e uma conclusão geral. O capítulo seguinte (Cap. 2) apresenta a revisão teórica e empírica da literatura, indicando os principais debates sobre o papel da rede na atividade de inovação na literatura internacional e do Brasil. O terceiro capítulo possui três partes: metodologia geral, base de dados e descrição da rede. Na primeira parte são apresentados conceitos de redes sociais e se desenvolve uma metodologia de tratamento de características nodais como contribuição metodológica à literatura. Na sequência mostra-se como a base de

dados é utilizada para se construir a rede nos dois níveis de análise, e a rede construída é investigada, a fim de se intuir características importantes para a construção do modelo. O quarto e quinto capítulos compreendem as estimações dos modelos regional e individual, respectivamente. Nestes capítulos se apresenta métodos de estimação, resultados e interpretações. A dissertação se encerra com as conclusões.

2 REVISÃO DE LITERATURA

O processo de criação e difusão de informação e conhecimento é um importante elemento para compreender a dinâmica do desenvolvimento econômico e objeto de estudo de economistas, geógrafos, historiadores e sociólogos. Embora se tenha o conhecimento de que estes processos são essenciais para a evolução e reprodução do sistema, diferentes leituras podem ser feitas para analisar a relação de causa-efeito. Neste capítulo é feita uma revisão da literatura onde se explora as diferentes interpretações teóricas e os principais resultados empíricos nas experiências internacionais e brasileiras.

2.1 INOVAÇÃO, TERRITÓRIO E DESENVOLVIMENTO

Os esforços teóricos para compreender como o processo de difusão do conhecimento ocorre e como este afeta o desenvolvimento econômico remonta a Marshall (1920), Von Thünen (1826), Christaller (1933), Isard (1956) e Schumpeter (1934). Em alguma medida, estes autores fundamentam o arcabouço da Nova Geografia Econômica (KRUGMAN, 1991) e da Economia Evolucionária (NELSON e WINTER, 1982), que estabeleceram eixos teóricos para compreender o desenvolvimento geográfico da inovação.

A leitura marshalliana (MARSHALL, 1920) do processo pode ser interpretada de duas formas. Primeiro, sob a ótica da localização da firma e, segundo, sob a ótica dos fluxos de informações, serviços e trabalhadores. Estas compreensões surgem a partir do livro *Principle of Economics*, originalmente publicado em 1980, onde Marshall (1920) afirma que

When an industry has thus chosen a locality for itself, it is likely to stay there long [...]. *The mysteries of the trade become no mysteries; but are as it were in the air* [...] (MARSHALL, 1920, p. 156, grifo nosso).

Neste trecho, Marshall identifica que as firmas tomam decisões de longo prazo quanto a sua localização baseando-se na perspectiva de que o “ar” no qual se insere será propício ao seu desenvolvimento. Contudo, a composição espaço é *ante hoc* pouco relacionado às decisões de localização das futuras indústrias. Ademais, embora Marshall esboce as condições que criam o ambiente industrial (o “ar”)¹, este não desenvolve qual o meio pelo qual as firmas e trabalhadores no ambiente possuem para respirá-lo, como destacam Fitjar e Rodríguez-Pose (2016). Se for possível descrever o espaço de inovação como aquele no qual as características do espaço geográfico facilitam os transbordamentos de informação pela velocidade de dispersão de conhecimento e recursos e a formação de contatos não-casuais, então o espaço marshalliano de inovação pode ser compreendido como um insumo da inovação. Neste mesmo sentido, na literatura de distritos industriais, Becattini (1989 *apud* Zaratiegui, 2004) definem a unidade básica de análise marshalliana como uma entidade socioterritorial.

Posteriormente, Schumpeter (1934, 1943) buscou compreender o processo inovador como o meio pelo qual os agentes econômicos participam da mudança econômica, que leva ao desenvolvimento. Sob esta ótica, o desenvolvimento pode ser compreendido como um meio no qual as decisões organizacionais das firmas e dos agentes interagem com o ambiente institucional (NORTH, 1990), o que inclui as regras sociais a serem seguidas no meio. Sendo um meio e não o insumo básico do desenvolvimento, a inovação é, portanto, um subproduto dos insumos fornecidos pela estrutura socioeconômica do espaço onde ocorre, e emana de uma decisão estratégica dos agentes.

Tem-se, então, a seguinte cadeia: estrutura socioterritorial – inovação – desenvolvimento econômico. Embora a relação não seja exata e tampouco se pode crer ser linear, esta permite realizar uma abstração teórica que leva a interpretação da inovação como um meio pelo qual as interações sociais ocorrem no espaço. Assim como a inovação, a opção pela interação de um indivíduo ou firma com as demais também é um processo decisório que emana das estruturas socioterritoriais e da perspectiva de retorno dos agentes. Assim, a lacuna entre a estrutura socioterritorial é um desafio presente tanto

¹ Marshall (1920) aponta duas fontes de aglomeração: por um lado, as firmas se localizam em espaços onde as vantagens emanam do compartilhamento de trocas especializadas entre os agentes; por outro lado, é a proximidade entre as firmas do mesmo setor industrial que cria o ambiente propício ao rápido intercâmbio de conhecimento que irá promover o desenvolvimento.

sob a perspectiva teórica que busca relacionar a estrutura social e a estrutura territorial no processo inovador, quanto sob a perspectiva empírica, que investiga as características mais relevantes no processo de cooperação entre os agentes.

Estes estudos seminais de Marshall e Schumpeter fundamentaram os recentes trabalhos que buscaram atualizar as leituras às recentes tecnologias e instrumentos empíricos. Tanto os estudos de ótica mais marshalliana, voltados à compreensão do espaço e da região no processo, quanto os estudos mais schumpeterianos, que estudam o ambiente de decisão e interação das firmas, se integraram aos estudos sobre inovação e, mais recentemente, sobre redes. A seção seguinte desenvolverá como as redes podem ser mais bem compreendidas no processo inovador.

2.2 A IMPORTÂNCIA DO ESTUDO DAS REDES NO PROCESSO DE DIFUSÃO DE CONHECIMENTO TECNOLÓGICO

A revolução tecnológica ocorrida com o aprimoramento da tecnologia de telecomunicações possibilitou a reconfiguração das relações de criação e difusão de conhecimento e informação em redes de inovação interligadas, que ampliam a capacidade inovadora (CASTELLS, 1996). Neste contexto, é notável que a formação de redes reduza o custo de cooperação entre inventores geograficamente separados e facilite a difusão de conhecimentos mais complexos.

Polanyi (1967) distingue informação, ou conhecimento codificado, e conhecimento tácito, de caráter mais complexo. Na interpretação de Ancori, Bureth e Cohendet (2000), informação é o conhecimento reduzido a mensagens que podem ser transmitidas entre os agentes, enquanto o conhecimento tácito dificilmente pode ser codificado sem que o custo para absorção deste conhecimento codificado seja elevado. Para Griliches (1992), o transbordamento de conhecimento (*knowledge spillover*) é fruto da divulgação do conhecimento gerado pela pesquisa, e dá condições às demais firmas de fazer sua própria pesquisa. Quanto mais codificado é o conhecimento e quanto maior é a capacidade de absorção de outras empresas ou inventores, mais transbordamentos de conhecimento ocorrerão. Por outro lado, conhecimentos complexos requerem tanto maior capacidade de absorção dos inventores quanto a formação de relacionamentos sociais. Owen-Smith e Powell (2004) destacam ainda que redes são componentes fundamentais do mercado que canalizam fluxos de informação entre agentes de uma estrutura social.

Este conceito é fundamental para a compreensão dos processos de criação e difusão de conhecimento no interior das redes. Desta forma, redes de inovação podem ser definidas como os espaços onde os inventores realizam trocas de informação e conhecimento, restritas ou não a um espaço físico delimitado.

Rogers (1962) chama de difusão² o processo de comunicação ao longo do tempo entre os membros de um sistema social, por meio da qual a informação converge (ou diverge) em uma, outra ou ambas as direções da relação. Desta forma, o processo de difusão na relação social das redes é uma forma de estabelecer uma comunicação sistematizada. Além da comunicação, Buschmann e Pyka (2015) destacam que estas redes são meios pelos quais as empresas e inventores dividem custos de pesquisa e desenvolvimento (P&D), acessam recursos escassos e, principalmente, administram processos de inovação complexos, lidam com incertezas tecnológicas (ROGERS, 1962) e criam novas oportunidades de aprendizado. No âmbito do desenvolvido de um sistema social de inovação, a formação de redes não apenas promove um *feedback* da inovação (CASTELLS, 1996, p. 31) por meio da comunicação, como também amplia a capacidade de pesquisa e viabiliza novos investimentos.

2.3 FORMAÇÃO DE REDES DE COLABORAÇÃO

2.3.1 Incentivos à cooperação

Os motivos que levam ao comportamento cooperativo estão em amplo debate na literatura. Dachs *et al.* (2008) argumentam que a tradicional visão baseada na firma sobre o processo inovador falha ao tentar explicar o porquê (e como) as firmas e indivíduos se engajam em atividades cooperativas. A visão relacional, segundo a leitura dos autores, volta-se às relações externas à firma como uma forma de adquirir vantagens cooperativas. De forma análoga, De Faria *et al.* (2010) identificam na literatura evidências que indicam que as firmas buscam cooperações externas à sua estrutura como forma de acessar habilidades, recursos e tecnologias complementares e cuja produção interna teria um elevado custo.

Sob a ótica individual, a figura do “inventor de garagem” é uma raridade. O atual inventor pode ser identificado como um trabalhador formal, qualificado e que domina um

²Rogers (1962) destaca ainda que alguns autores restringem o termo “difusão” ao espalhamento espontâneo, enquanto “disseminação” remete ao fluxo dirigido e organizado de ideias. O livro, contudo, não faz distinção semântica clara entre os termos ao longo da exposição.

conjunto de informações altamente especializado, que busca acessar novos conhecimentos pelo contato interpessoal em equipe. Este fato pode ser observado com o aumento do tamanho das equipes de pesquisa e desenvolvimento (WUCHTY *et al.*, 2007) e da coautoria acadêmica (HICKS E KATZ, 1996; ADAMS *et al.*, 2005; SIDONE, HADDAD E MENA-CHALCO, 2016).

A cooperação inventiva pode ser dividida em três etapas: (i) contato, (ii) cooperação *de facto* e (iii) invenção. O primeiro estágio é mais idiossincrático e de difícil mensuração; Giuri e Mariani (2013) apontam que, em geral, o contato ocorre no interior de comunidades que proveem oportunidades de cooperação. O ingresso em uma comunidade de colaboração, contudo, não é claro na literatura. Indivíduos podem ora buscar ativamente um parceiro (*e.g.* ir a eventos científicos), ter relacionamentos casuais (*e.g.* colegas de faculdade), ou relacionamentos profissionais (*e.g.* colegas de trabalho). Uma vez estabelecido o contato, a cooperação está apta a ocorrer caso haja mútuo interesse.

Na etapa da cooperação, indivíduos com contato prévio são considerados parceiros potenciais. Caso seja concretizada, durante a cooperação ocorre a troca de conhecimentos e informações, por meio da qual há a difusão. Logo, o processo pelo qual os inventores passam de parceiros potenciais para parceiros concretos é um objeto de estudo. Os determinantes desta mudança de papel é o foco desta dissertação

Por fim, a cooperação concretizada gera uma invenção, que pode ser mensurada. Este produto da cooperação pode revelar informações importantes sobre as etapas anteriores, sobretudo sobre a forma como o inventor opta pela cooperação. Da mesma forma, retroativamente, a identificação das comunidades onde os contatos ocorrem pode ser realizada. Contudo, o conceito de comunidade cooperativa não é consenso uma vez que esta pode ser fruto de diferentes espaços. A subseção seguinte discutirá como tais espaços podem determinar a formação de redes à luz das perspectivas da economia regional e da análise de redes sociais.

2.3.2 Espaço de lugares *versus* espaço de fluxos

A tradição de análise de inovação iniciada por Schumpeter (1943) e estudada posteriormente por Griliches (1979) e Jaffe (1989) busca identificar os determinantes, a capacidade de transbordamento e suas consequências para o desenvolvimento econômico.

Em particular, a formação de sistemas de inovação atua tanto como meio para a criação da inovação quanto como canal para a reprodução deste processo (CASTELLS, 1996). Por muito tempo, estes estudos estiveram restritos a explicar o processo de transbordamento de conhecimentos pelos limites geográficos criados com o aumento do custo de propagar a informação com a distância. No campo da geografia econômica, os trabalhos de Jaffe (1989), Krugman (1991), Jaffe, Trajtenberg e Henderson (1993) e Audretsch e Feldman (1996) fundamentaram o arcabouço teórico no estudo deste fenômeno.

Retomando a discussão da seção 2.1, Marshall (1920) destacou que o aumento da distância física restringe a velocidade de propagação de bens, serviços e informações e, ao mesmo tempo, concentra no território a atividade econômica, facilitando, por isso, a comunicação entre compradores e vendedores. Analogamente, Krugman (1991) ressalta que a aglomeração territorial cria retornos crescentes, dotando a aglomeração produtiva de novas vantagens locais. Há extensa pesquisa que ressalta a facilidade de transbordamentos de conhecimento à medida que a distância entre os agentes econômicos envolvidos diminui (JAFÉ, 1989; AUDRETSCH e FELDMAN, 1996). Henderson (2007) aponta que a difusão da informação e do conhecimento possuem decaimento espacial, o que a torna mais eficiente entre atores próximos. Uma das formas mais eficazes de difusão de conhecimento é por intermédio de redes de colaboração (FRENKEN *et al.*, 2009), que também podem ter como características o fato de serem limitadas espacialmente (COWAN e JONARD, 2004).

A revolução no paradigma das comunicações foi fundamental para a construção de um sistema internacional de cooperação que superasse os limites espaciais antes impostos pelos custos crescentes de comunicação (CASTELLS, 1996). O maior destaque dado aos ganhos da cooperação entre inventores em grande escala tem sido explorado por estudos que buscam introduzir as redes na economia regional e na geografia econômica (GABHER e IBERT, 2006; TER WAL e BOSCHMA, 2008). Embora estudos clássicos anteriores como Coleman, Katz e Menzel (1966) e Rogers (1962) tenham introduzido metodologias de análise de redes de inovação, os estudos recentes neste campo se diferenciam por dar destaque não apenas à formação da rede, mas à posição relativa dos agentes inseridos em redes onde a comunicação tácita pode ser transmitida a um custo menor. Grabher e Ibert (2006) sugerem que a introdução do conceito de redes iniciou uma nova visão sobre o contexto social da ação econômica.

Ter Wal e Boschma (2008) distinguem “espaço de lugares” (*space of places*) de “espaço de fluxos” (*space of flows*). No primeiro, existe a ideia de que a geografia importa para o aprendizado e inovação; já o espaço de fluxos carrega a ideia de que as redes são importantes veículos de transferência de conhecimento e difusão. Embora haja esta distinção, os autores destacam que a formação das redes parece estar geograficamente confinada. Da mesma forma, Henderson (2007) e Cowan e Jonard (2004) apontam que a difusão da informação e do conhecimento nas redes possui decaimento espacial, o que a torna mais eficiente entre atores próximos.

Tem-se, portanto, duas perspectivas dominantes sobre os determinantes da formação das redes. Na primeira, as relações ocorrem no espaço geográfico, ou espaço de lugares, onde o decaimento espacial é o principal fator que levará à formação de redes espacialmente confinadas. Na segunda, as redes se formam no espaço relacional, ou espaço de fluxo, dentro do qual a inserção dos agentes na rede é mais importante para a absorção, compartilhamento e transbordamento de informações e conhecimentos que a distância física para os outros membros da rede (HOEKMAN, FRENKEN e VAN OORT, 2008).

Recentemente, os avanços na agenda institucional³ introduziram nestas duas perspectivas características antes não consideradas na análise dos padrões de invenção e inovação. Mohr e Guerra-Pearson (1998) descrevem um terceiro espaço importante onde são definidos incentivos e barreiras ao processo inovador, o “espaço institucional” (*institutional space*). Neste, características sócio-organizacionais definem espaços comunitários de necessidades e problemas que devem ser tratados de duas maneiras, uma sob a ótica coletiva outra sob a ótica dos atributos individuais e sua relação no meio. No âmbito da economia da inovação, estas características são relevantes na medida em que diferenças culturais, organizacionais, sociais e geracionais podem influenciar a forma como os agentes formam conexões e as áreas de conhecimento que privilegiam.

2.3.3 Cooperação entre indivíduos e entre regiões: dois níveis de análise

Uma vez compreendidos quais espaços propiciam a formação de comunidades pelas quais os nós – indivíduos, firmas ou regiões – selecionam potenciais parceiros,

³A agenda institucional ganhou certo destaque na ciência econômica recente nos últimos 30 anos. Neste período, quatro autores ligados a estes estudos foram laureados com o Prêmio Nobel de Economia: Coase (1991), North (1993) e Williamson e Ostrom (2009).

pode-se realizar uma análise dos incentivos à cooperação. Dois níveis de agregação serão analisados, o nível individual e o nível regional. Esta subdivisão será utilizada na formulação dos modelos empíricos.

Sob a ótica do inventor, este possui uma quantidade limitada de tempo e recursos para dividir em projetos individuais e projetos cooperativos. Por um lado, dedicar-se a projetos individuais demanda custos de obtenção de conhecimentos e, por consequência, resulta em uma produção mais lenta de inovações. Por outro lado, engajar-se em projetos de pesquisa em cooperação com outros inventores possui um custo para se manter em constante contato com potenciais parceiros. Além disso, em invenções cooperativas, o retorno de reputação do inventor é inferior à invenção individual, embora ocorra com maior frequência (WUCHTY *et al.*, 2007). Portanto, na decisão entre cooperar ou não cooperar, o inventor deve conciliar custos e retornos.

Contudo, o retorno da cooperação é heterogêneo, sendo maior para “boas conexões” e menor para “conexões ruins”. Este julgamento, entre conexões boas e ruins, é igualmente heterogêneo entre os inventores e está no campo da percepção, sendo dificilmente mensurável. Kahneman (2002) define a percepção como um processo rápido, automático, associativo e de aprendizado lento. Em outras palavras, a percepção pode ser compreendida como a forma como os indivíduos interpretam um limitado conjunto de informações utilizando filtros pessoais sujeitos a experiências e conhecimentos anteriores, e ao senso comum. Nesta definição, a alocação dos recursos de um inventor entre cooperar ou não é resultado da sua interpretação *ex ante* do retorno da cooperação, baseado na sua percepção de ganhos.

Neste processo, o inventor possui um *portfólio* de parceiros potenciais com os quais teve contato prévio. Sendo o tempo a se alocar na cooperação limitado, o inventor deve optar pelo conjunto de parceiros potenciais a se concretizar a relação. Nesta etapa – a escolha dos parceiros –, o inventor utiliza a percepção dos retornos de cada parceiro potencial a fim de elencar aqueles de quem se espera maior retorno. Owan e Nagaoka (2011) identificaram duas classes de motivos que guiam a escolha do inventor:

Intrinsically motivated behaviors are behaviors which a person engages in to feel competent and self-determining (Deci, 1975) and for R&D researchers overcoming obstacles to contribute to the advancement of science fulfills this definition. They are also influenced by 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. (OWAN e NAGAOKA, 2011, p. 3)

Neste processo decisório, a percepção pode dificultar a análise dos retornos que seriam auferidos pelos inventores. Se as instituições que afetam a conexão são discriminatórias, o inventor pode subjugar o retorno de uma relação. Assim, questões importantes como o sexo do inventor (WHITTINGTON E SMITH-DOERR, 2008; JUNG E EJERMO, 2014), sua região de origem, sua etnia, o idioma que fala ou mesmo a área tecnológica que atua podem ser fontes de discriminação que modificam a percepção dos inventores (CRESCENZI, NATHAN E RODRIGUEZ-POSE, 2016). Em particular, quando se limita as fronteiras físicas de análise, limita-se também o grau de heterogeneidade das comunidades formadas. Por exemplo, ao se analisar o Brasil, questões linguísticas são pouco importantes.

Já sob o nível de agregação regional, a análise segue por outros meios. Por esta ótica, os recursos são menos limitados e as conexões potenciais menos restritas. A questão então passa a ser o que atrai inventores de uma região a buscar parcerias em outra. Uma solução metodológica é proposta por Torre e Gilly (1999) e Boschma (2005) que organizaram em diferentes dimensões de proximidade os elementos que aproximam ou afastam as regiões. Torre e Rallet (2004) interpretam como “próximas” as regiões cujas características são parecidas e que facilitam a concretização da relação. Por esta perspectiva, entende-se a inovação como uma combinação de cinco tipos de proximidade (BOSCHMA, 2005):

- (i) Geográfica: como a inovação é afetada pelo espaço físico. Análogo à leitura individual, regiões buscam outras para complementar sua infraestrutura de pesquisa sem incorrer em altos custos. Contudo, o aumento da distância entre duas regiões aumenta o custo de acesso à estas facilidades (*e.g.* custos de transporte), o que dificulta essas conexões. Por outro lado, uma distância geográfica muito próxima pode ampliar a competição entre as firmas da região e dificultar a conectividade.
- (ii) Social: como a posição social dos nós afeta a sua possibilidade de cooperação. Sob esta dimensão, quanto mais socialmente próximas estão as regiões – por exemplo, com parceiros em comum, maior seria sua possibilidade de conexão.

- (iii) Cognitiva: como o mútuo entendimento entre os nós permite a troca informacional. Boschma (2005) aponta que caso a proximidade cognitiva seja pequena, os nós terão dificuldade de compreender a informação a ser trocada, já caso a proximidade entre os nós seja muito elevada, a falta de novidade informacional reduzirá a busca por complementaridade.
- (iv) Organizacional: como os diferentes mecanismos de controle ou gestão organizacional afetam a cooperação. Nesta dimensão, caso os nós possuam organizações com gestões muito diferentes, então se aumenta a burocracia para se estabelecer o contato, o que reduz a cooperação. Embora esta dimensão seja mais evidente na conexão entre firmas, regiões podem ter perfis organizacionais característicos. Por exemplo, regiões com um elevado número de universidades, ou regiões financeiras, ou regiões industriais.
- (v) Institucional: como as instituições locais geram relações de confiança na cooperação. Caso duas regiões tenham características institucionais bastante distintas, o custo de estabelecer mecanismos de incentivo que propiciem um contato confiável é elevado, o que reduz as potenciais conexões.

Além das proximidades – definidas conjuntamente no laço – características particulares dos nós podem também afetar a visibilidade que este possui e a maior chance de este se conectar. No contexto das regiões, o estoque de infraestrutura de pesquisa, de pesquisadores e de recursos financeiros que custeiam as atividades de P&D também determinam a atratividade de uma localidade. Cohen e Levinthal (1989) apontam que a capacidade de absorção de um conhecimento está relacionada ao estoque prévio de conhecimento de uma região, fortemente relacionado à riqueza de uma região e ao seu gasto em P&D. Se, como anteriormente mencionado, indivíduos e empresas buscam outras regiões a fim de acessar recursos não disponíveis no seu entorno, então regiões que dispõem de mais recursos tendem a atrair mais que as regiões menos desenvolvidas.

Os trabalhos empíricos revisados nas seções seguintes utilizam, de forma intencional ou não, uma ou mais dimensões de proximidade das acima descritas. A divisão como proposta auxilia na identificação das contribuições metodológicas e dos resultados para corroborar, ou não, suas hipóteses de teste.

2.4 EVIDÊNCIAS EMPÍRICAS DO PAPEL DAS REDES NA DIFUSÃO E SEUS DETERMINANTES

Assim como a inovação, a formação de redes está espacialmente concentrada (JAFFE, TRAJTENBERG e HENDERSON, 1993; FELDMAN e KOGLER, 2010). Owen-Smith e Powell (2004) revisam na literatura empírica três características notáveis na formação de redes: (1) redes empresariais são mais propensas a se formar entre firmas localizadas no mesmo espaço (JAFFE *et al.*, 1993); (2) o tamanho e mobilidade das aglomerações de trabalhadores em ciência e engenharias amplia os transbordamentos localizados (ALMEIDA e KOGUT, 1999); e (3) alianças estratégicas entre firmas aumentam a probabilidade de transbordamentos.

Rogers (1983)⁴ elaborou robusta revisão empírica dos trabalhos sobre redes de difusão de inovação produzidos até o início da década de 1980. Os trabalhos analisados estavam inseridos em um contexto social e tecnológico diferente das atuais redes de inventores, influenciadas diretamente pelas novas tecnologias de difusão de conhecimento. Esta revisão, contudo, permite visualizar dois importantes aspectos das pesquisas empíricas em redes de inovação: (1) a mudança na estrutura organizacional das redes de inovação e seu papel na difusão, e (2) a importância da metodologia da Nova Geografia Econômica (KRUGMAN, 1991) e do estudo institucional como complementares teóricos das pesquisas anteriores. A grande contribuição da revisão de Rogers (1983) é identificar na literatura empírica o papel de laços fortes e laços fracos entre inventores (GRANOVETTER, 1973), que dá peso à posição relativa de cada nó (inventores, firmas ou regiões) no espaço de fluxos⁵.

A partir dos trabalhos desenvolvidos pela literatura da geografia econômica, em especial ao modelo de transbordamentos geográficos de Jaffe, Trajtenberg e Henderson (1993), o papel do espaço físico foi inserido no contexto do estudo da inovação. Contudo,

⁴O livro de 1983 é a terceira edição do livro publicado em 1962 e reeditado também em 1971. Em 2003 o autor publicou a última edição do livro *Diffusion of Innovations*. Optou-se por utilizar para esta seção de revisão empírica a terceira edição, de 1983. Em outros momentos neste trabalho foi utilizada a primeira edição, de 1962. O objetivo desta escolha é criar um padrão cronológico das contribuições empíricas e teóricas de Rogers para o estudo da difusão e do papel das redes.

⁵Grosso modo, laços e nós são unidades conjuntas básicas para análise das redes. “Nós” são definidos como pontos distribuídos no espaço relacional. Estes pontos possuem características individuais e que os diferenciam entre si. “Laços” são segmentos de retas formados entre dois nós no espaço relacional. Estes segmentos possuem características conjuntas das conexões. Por exemplo, a idade é uma característica individual do nó, uma vez que não é afetada pela idade de qualquer outro nó, enquanto a distância física é uma característica do laço, pois é afetada pela posição geográfica de ambos os nós. Para uma definição formal de redes, ver seção 2.6.

embora tenha definido o arcabouço tradicional para o estudo deste fenômeno, não houve grandes esforços ao longo da década de 1990 em estabelecer convergências teóricas e empíricas com a literatura de redes anteriormente desenvolvidas. Portanto, embora estes estudos reconhecessem nos laços entre inventores o canal de transbordamento de informação e conhecimento, os determinantes da formação destas conexões foram negligenciados nestes estudos.

Na revisão de Cowan (2005), o estudo do papel das redes ressurgiu na década de 1980, sobretudo devido ao crescente interesse dos economistas pelas tecnologias que facilitam o contato. O retorno deste paradigma metodológico pode ser creditado tanto à maior capacidade computacional que permitiu o tratamento de dados de grandes redes, quanto à maior atenção acadêmica dada à aplicação da teoria dos grafos nas redes sociais, especialmente pela ciência da computação. O autor divide em duas vertentes os trabalhos que voltados à análise da formação da rede: (i) os estudos baseados em jogos, e (ii) os estudos evolucionários. Os trabalhos abaixo revisados se encaixam na primeira definição, que estuda os incentivos à formação de laços em diferentes níveis e estruturas de rede.

Os trabalhos pela abordagem relacional não negaram o papel da geografia na difusão do conhecimento, mas indicaram que, segundo a lógica das proximidades, o espaço físico pode ser complementado ou mesmo substituído pelas características sociais da inovação. Para este fim, instrumentos tradicionalmente utilizados pela geografia econômica e métricas de análise de redes sociais são combinados.

No âmbito das contribuições metodológicas, os trabalhos variam quanto ao nível de análise. Autores como Breschi e Lissoni (2004, 2005) e Cassi e Plunket (2015) adotaram como variável dependente a existência de conexões entre patentes, para o primeiro, e entre regiões, para o segundo. Já Montobbio e Sterzi (2013) realizaram uma análise global para compreender o diferencial de internacionalização entre países desenvolvidos e em desenvolvimento.

Quanto ao modelo de estimação, Ejermino e Karlsson (2006) optaram por um modelo de contagem adequado à categoria da variável analisada. Cassi e Plunket (2015) utilizaram o *logit* condicional e o *probit* multinomial para lidar com a estrutura dos dados, já Crescenzi, Nathan e Rodriguez-Pose (2016) obtiveram resultados utilizando regressão linear robusta. Maggioni, Nosvelli e Uberti (2007) optaram pela utilização de um modelo de defasagem espacial. Neste trabalho, os autores constroem uma matriz de contiguidade

relacional além do modelo espacial convencional. Esta matriz, como descrita por Leenders (2002) considera os vizinhos em rede de cada região, ou seja, as conexões diretas de primeiro grau. Como será melhor discutido na seção metodológica, estes modelos não são eficientes para se trabalhar com casos em que o evento observado (a conexão) é raro frente às possibilidades.

Quanto aos principais resultados encontrados na literatura internacional (cujos equivalentes para o Brasil serão apresentados na seção seguinte, pode-se organizá-los em:

- *Papel da geografia*: ao nível regional, Scherngell e Barber (2009) e Morescalchi *et al.* (2015) encontraram que a distância física é um fator limitante para a formação de conexões dentro das fronteiras da União Europeia. Já sob a ótica da internacionalização das atividades inventivas, Picci (2010) identificou que regiões fronteiriças possuem maior chance de cooperação bilateral. Resultado similar foi encontrado por Ejermo e Karlsson (2006), cujos resultados indicaram que o nível de P&D de regiões próximas aumentam a afinidade inovativa. Já os resultados de Paier e Scherngell (2011) para as conexões interfirmas também indicam o importante papel da geografia para a formação de laços. Além disso, Cassi e Plunket (2015) e Maggioni, Nosvelli e Uberti (2007) destacam a não-linearidade do papel da distância, destacando que esta é mais relevante para a formação de novos laços. Estes resultados corroboram para a formação de redes o decaimento espacial descrito por Jaffe (1989) e Audretsch e Feldman (1996) para a difusão de conhecimento.
- *Proximidades substituíveis*: ao nível do inventor, Cassi e Plunket (2015) identificaram que diferentes formas de proximidade podem ser substituíveis, sobretudo se existe conexão prévia. Dimensões fundamentais para a primeira etapa da formação da rede – o contato – como a geografia e a tecnologia, são gradualmente substituídas pela posição relativa do nó na rede, capturada pela dimensão social. Logo, a substituição entre dimensões de proximidade parece relacionada ao nível de amadurecimento da relação.
- *Nível de desenvolvimento regional*: Montobbio e Sterzi (2013) compararam a internacionalização das atividades de copatenteamento com

o nível de desenvolvimento econômico (mensurado pelo PIB *per capita*) e institucional (pela maturidade do sistema de propriedade intelectual). Os autores encontraram que há uma elevada correlação entre as maturidades econômicas e institucionais. Assim, regiões menos desenvolvidas possuem uma propensão menor a se engajar em cooperações. Por outro lado, a cooperação também permite a essas regiões acesso a recursos necessários à pesquisa.

- *Características individuais do inventor*: no âmbito individual, Crescenzi Nathan e Rodriguez-Pose (2016) buscaram incluir características individuais do inventor, como grupo étnico-linguístico, que obteve um coeficiente positivo na propensão à formação de laços. Já relacionado ao gênero do indivíduo, Whittington e Smith-Doerr (2008) identificaram que existe uma propensão à formação de redes intragrupos mais fortemente que intergrupos. Contudo, como informações pessoais tem acesso restrito, características relevantes apontadas pela literatura supracitada, como nível de renda e nível educacional, não foram exploradas pela literatura.

Portanto, a recente literatura empírica indica que a formação da rede é determinada por um conjunto de dimensões e não apenas pela posição relativa do nó na rede ou no espaço geográfico, embora ambos sejam importantes. Estes estudos indicam que características organizacionais (a cooperação entre inventores com diferentes naturezas jurídicas), cognitivas (a cooperação entre inventores de diferentes áreas de conhecimento), institucionais (a cooperação entre inventores ou regiões cujas normas ou incentivos são diferentes) e étnico-demográficas (a cooperação entre indivíduos cuja cultura ou características sociais sejam diferentes) possuem também influência na determinação da formação, continuidade ou dissolução de laços.

2.5 INOVAÇÃO E REDES DE COLABORAÇÃO NO BRASIL

Se comparado às demais experiências internacionais, o sistema brasileiro de inovação é imaturo, embora em processo de formação (ALBUQUERQUE, 1996; GONÇALVES E ALMEIDA, 2009). Albuquerque (2000) investiga as patentes depositadas entre 1980 e 1995 e reforça esta característica destacando que o sistema brasileiro de inovações tem características particulares se comparados aos sistemas maduros: (1) uma alta participação de inventores individuais e de pequenas firmas e baixa

participação de empresas, sobretudo de porte médio; (3) elevada presença de firmas multinacionais, e; (4) baixo envolvimento contínuo das empresas em atividades inventivas.

Sob a ótica regional, Albuquerque *et al.* (2002) e Lemos *et al.* (2006) identificam que a atividade inventiva brasileira é altamente concentrada na região centro-sul do país, embora se note uma tendência à descentralização de São Paulo para as demais metrópoles industriais emergentes. Destacam ainda possuem forte potencial atrativo à atividade inventiva devido à atividade de P&D devido às condições urbano-espaciais, sobretudo relacionadas à infraestrutura de pesquisa, ao transporte e ao acesso a serviços e profissionais. Gonçalves (2007) encontra ainda uma elevada diferença no padrão de aglomeração no depósito de patentes per capita entre as regiões mais ao norte e a região centro-sul, indicando uma distribuição desigual da atividade no espaço. Em complemento, Gonçalves e Fajardo (2011) ratificaram o papel do espaço na atividade, bem como encontraram que a similaridade tecnológica pode ampliar os transbordamentos entre regiões.

Se, por um lado, o estudo da distribuição espacial da inovação e dos transbordamentos no Brasil é bem consolidado, por outro lado, o estudo de redes de colaboração no Brasil ainda é limitado. Os esforços para o desenvolvimento de uma literatura empírica neste campo são vistos nos trabalhos de Sidone, Haddad, e Mena-Chalco (2016a, 2016b) e Oliveira (2016), embora apenas o segundo seja baseado em redes de copatenteamento.

Sidone, Haddad, e Mena-Chalco (2016a, 2016b) utilizaram os currículos da plataforma *Lattes* para a construção de uma rede nacional de coautoria acadêmica. Os resultados obtidos indicam que a distância ainda importa para a formação desta rede. Por outro lado, observaram uma tendência à descentralização das publicações, com o aumento da participação de regiões menos tradicionais em ciência, como o Sul e o Nordeste.

Já Oliveira (2016) analisa a distribuição geográfica dos copatenteamentos no Brasil e identifica que assim como a atividade inventiva, a atividade colaborativa também está bastante relacionada à estrutura urbano-espacial das metrópoles inventivas. Além disso, metrópoles inventivas regionais fora do “polígono de desenvolvimento” identificado por Lemos *et al.* (2006), tem forte capacidade de polarizar seu entorno e criar novas periferias nos subcomponentes de suas redes locais.

Esta dissertação se insere neste grupo para contribuir no entendimento das redes de inovação, sobretudo para a compreensão de como esta é formada e quem são os inventores que se conectam a ela.

2.6 COPATENTES COMO INDICADOR DE DIFUSÃO POR COLABORAÇÃO

Nos estudos sobre inovação, algumas *proxies* têm sido utilizadas na tentativa de capturá-la quantitativamente (NAGAOKA, MOTOHASHI e GOTO, 2010). Os gastos em pesquisa e desenvolvimento (P&D) e a produtividade total dos fatores (PTF), por exemplo, são variáveis comuns na análise do processo inovador. Contudo, a primeira é antes um indicador de insumo do processo do que o produto da inovação em si e seu uso negligenciaria o valor tecnológico agregado; segundo, a utilização da PTF é contaminada por outros fatores externos à mudança tecnológica que também afetam a produtividade.

Por outro lado, a patente, como um produto do processo inventivo, apresenta algumas vantagens no seu uso como indicador de inovação. Nagaoka, Motohashi e Goto (2010) elencam estas vantagens: (1) existência de base de dados confiáveis e de longo prazo; (2) possui informações detalhadas sobre a tecnologia e seu depositante; (3) contém referências bibliográficas bem documentadas, o que permite a rastreabilidade da evolução tecnológica; (4) tecnologias patenteadas satisfazem rigorosos critérios que garantem a não-obviedade e a aplicação industrial da tecnologia. Por outro lado, os autores destacam os desafios do uso deste tipo de informação, como a dificuldade em lidar estatisticamente com o volume de informações gerado e as tecnologias protegidas por outros meios que não o patenteamento, como o segredo industrial. Ainda assim, a utilização dos dados de patente para analisar a inovação (como produto) e o processo inovador (a relação do insumo ao produto) se justifica pela disponibilidade e confiabilidade dos dados gerados.

A fim de se estudar a colaboração, o uso de patentes permite relacionar inventores por meio de copatenteamento (TER WAL e BOSCHMA, 2008). Define-se copatenteamento, seguindo Breschi e Lissoni (2004), como a cooperação inventiva que produziu uma patente. É cabível, porém, notar uma limitação desta abordagem. O relacionamento existe previamente à concretização do produto. Por outro lado, relacionamentos profissionais que produzem inovação podem ser caracterizadas como tendo maior valor econômico.

Além disso, Ter Wal e Boschma (2008) destacam outras limitações do uso de copatenteamento como indicador de cooperação. Primeiro, as informações contidas nos

registros de patentes sobre os inventores consideram apenas aqueles que tiveram o nome incluído no depósito. Desta forma, o papel dos técnicos e outros profissionais que atuaram na pesquisa é ignorado. Em muitas pesquisas, estes outros atores contribuem ativamente na introdução de novo conhecimento na pesquisa e, portanto, na difusão de novo conhecimento dentro da rede. Segundo, o patenteamento possui variabilidade setorial. Em alguns setores, o uso de segredo industrial e o registro de marcas são preferidos como forma de proteção de propriedade intelectual. Terceiro, o patenteamento está relacionado ao tamanho da firma, uma vez que grandes empresas são mais propensas a incorrer em custos de patenteamento. Por fim, as inovações realizadas em universidades, centros de pesquisa e outros órgãos com menor propensão a patentear podem estar sub-representadas no uso de dados de patentes. Estas limitações podem restringir o acesso dos indicadores às relações sociais existentes no sistema.

3 METODOLOGIA GERAL, BASE DE DADOS E DESCRIÇÃO DA REDE

Este capítulo está dividido em três partes. A primeira, acerca da metodologia geral, engloba as seções 3.1 e 3.2. Nestas seções são formalizados os conceitos de grafos (3.1) e o princípio utilizado para a inclusão de características nodais nos modelos baseados em laços (3.2). Descrições detalhadas sobre os métodos de estimação e construção de contrafactual serão apresentados junto às metodologias específicas de cada modelo, nos capítulos 4 e 5. Em sequência, a seção 3.3 trata da base de dados. Nesta é apresentado o conteúdo desta base, o método de pareamento, e a metodologia utilizada para a construção de redes em diferentes níveis de agregação. Por fim, utilizando as metodologias apresentadas, a rede construída é descrita na seção 3.4, que se subdivide nos níveis de agregação analisados.

3.1 DEFINIÇÕES FORMAIS DE GRAFOS E REDES SOCIAIS

A análise de redes sociais é um campo derivado da Teoria dos Grafos. Nesta, uma rede é caracterizada pela sua estrutura, a posição relativa e papéis dos atores e pelos relacionamentos formados entre esses (WASSERMAN e FAUST, 1995). Este tipo de análise é utilizado por diversos campos de conhecimento que buscam explicar a relação entre atores cujas relações criam e afetam um sistema. A construção de métricas e definição de modelos funcionais que lidam com redes requerem a definição formal de alguns conceitos (WALLIS, 1995):

Definição 1.1. Um *grafo* finito G é formado por um par $(V(G), A(G))$, tal que $V(G) = \{v_1, \dots, v_n\}$ é um conjunto finito não-vazio de vértices (*vertices*) e $A(G)$ é uma família de pares não ordenados chamados arestas (*edges*) formados por elementos contidos em $V(G)$ não necessariamente distintos.

Definição 1.2. Chama-se de “laço” (*tie*) uma aresta se para o nó $v \in V(G)$, $vv \in A(G)$.

Definição 1.3. Se $\{x, y\} \in A(G)$ é u laço, então diz-se que x é adjacente de y , denotado por $x \sim y$. Uma matriz M_G é chamada de matriz de adjacência de G se para cada par $m_{ij} = \{x_i, x_j\}$, tem-se que

$$m_{ij} = \begin{cases} 1 & \text{se } x_i \sim x_j \\ 0 & \text{caso contrário} \end{cases}$$

Uma matriz de adjacência, quando representada em um espaço relacional, permite visualizar nós e laços, o que auxilia a identificar de antemão os nós que possuem mais conexões e induzir noções de centralidade e aglomerações em torno de vértices. O exemplo a seguir mostra um grafo simples, formado por 5 vértices. A Figura 1.1 mostra a matriz de adjacência, enquanto a Figura 1.2 mostra a distribuição do grafo no espaço relacional.

Figura 1. Exemplo de grafo com 5 vértices e 5 arestas

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Figura 1.1

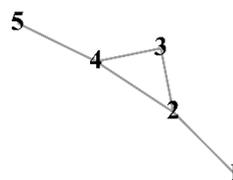


Figura 1.2

A Figura 1.2 foi elaborado com o software Gephi

Definição 1.4. O grau (*degree*) $g(v)$ de um vértice $v \in V(G)$ é definido como o número de arestas que contém v . Em qualquer grafo, a soma dos graus de todos os vértices é igual ao dobro do número de arestas⁶. Define-se grau máximo de G como sendo $\max\{g(v) | v \in V(G)\}$ e o grau mínimo como $\min\{g(v) | v \in V(G)\}$. Chama-se de *pontos isolados* os vértices em que $g(v) = 0$.

⁶A segunda parte desta definição corresponde ao Teorema 1.1 de Wallis (1995).

No exemplo da Figura 1, tem-se 5 vértices conectando 5 arestas. A soma das colunas para cada linha na matriz de adjacência é o grau de cada vértice, de 1 a 5. Neste grafo, os vértices 2 e 4 possuem ambos graus 3, o vértice 3 possui grau 2 e o vértice 1 e 5 possuem ambos graus 1. O grau máximo e o grau mínimo deste grafo é, portanto, 3 e 1, respectivamente. A importância de se saber o grau de cada vértice e o grau máximo e mínimo de um grafo é identificar os nós importantes na rede.

Definição 1.5. Um passeio (*walk*) em G é uma sequência finita de vértices v_0, v_1, \dots, v_n e arestas a_1, \dots, a_n , em que o ponto inicial de a_i é o vértice v_{i-1} para cada i . Um caminho (*path*) é um passeio em que nenhum vértice v se repete na sequência. Chama-se de geodésico $\ell(v_1, v_2)$ a quantidade de arestas no menor caminho entre dois nós.

Na Figura 1, por exemplo, dois caminhos possíveis entre os nós 5 e 3 são formados pelas sequências $v_5, v_4, v_3, a_{45}, a_{34}$ e $v_5, v_4, v_3, v_2, a_{45}, a_{24}, a_{23}$, sendo o primeiro mais curto que o segundo. Este conceito é importante para identificar o quão próximo ou o quão distantes no espaço relacional estão dois nós. Quanto mais distantes são os nós, menor é a força de atração entre eles, bem como quanto mais próximos estes nós estão, maior é a atratividade.

Métricas de Análise de Redes

Já conhecidas as definições básicas de grafos e seus elementos fundamentais, a construção de métricas permite extrair informações sobre a estrutura do grafo, a posição relativa dos nós e a relevância de cada laço para o sistema. Nesta seção será feita uma revisão apenas das métricas que serão essenciais para a análise das redes de inovação nesta dissertação (JACKSON, 2008; WASSERMAN e FAUST, 1995).

- 1. Grau médio $g(v)$** – É o valor médio de $g(v)$ no grafo. Indica a conectividade média de cada nó e pode indicar para diferentes configurações de rede qual grafo possui vértices mais conectados. Em grafos não direcionais, onde as arestas formadas entre os vértices caminham em ambas direções, o grau médio é dado pela equação:

$$\bar{g}(G) = \frac{\sum_{i=1}^n g(v_i)}{n} \quad (1)$$

No exemplo da Figura 1, o grau médio é 2.

- 2. Densidade $\Delta(G)$** – É a proporção entre o número de arestas possíveis em um grafo e o número de arestas existentes (L). Em um grafo não-direcional, a quantidade de arestas possíveis é dado por $n(n-1)/2$. A densidade é descrita algebricamente por:

$$\Delta(G) = \frac{L}{n(n-1)/2} = \frac{2L}{n(n-1)} \quad (2)$$

No exemplo da Figura 1, a densidade é 0,5.

- 3. Centralidade** – Indica a importância relativa da posição do nó na rede. Na realidade, a centralidade é uma família de métricas desenhadas para este fim. Jackson (2008) destaca quatro principais medidas de centralidade comumente adotadas na análise de redes sociais. São elas: centralidade de grau, de proximidade, de intermediação e de vizinhança.

3.1. Centralidade de Grau: indica o quão conectado é um nó;

$$C_G(v) = \frac{g(v)}{n-1} \quad (3)$$

3.2. Centralidade de Proximidade: o quão fácil é para um nó alcançar outros;

$$C_P(v) = \frac{n-1}{\sum_{j \neq i} \ell(i,j)} \quad (4)$$

onde $P_i(kj)$ é o número de geodésicos entre k e j que passa por i, e $P(kj)$ é o número total de geodésicos.

3.3. Centralidade de Intermediação: qual a importância de um nó em termos de sua conectividade com os demais;

$$C_i(v) = \sum_{k \neq j; i \notin \{k, j\}} \frac{P_i(kj)/P(kj)}{(n-1)(n-2)/2} \quad (5)$$

3.4. Centralidade de Vizinhança: o quão importante, central ou influente a vizinhança de um nó é. Esta métrica pode ser aferida como a média das centralidades de grau, de proximidade e de intermediação (que formam uma matriz $C_{n \times 3}$) nos vizinhos de v . Para cada vértice i , tem-se $C_V(v_i)_{1 \times 3}$ na linha i da matriz definida por:

$$C_V(v)_{n \times 3} = \tilde{M}_{G, n \times n} C_{n \times 3} \quad (6)$$

onde \tilde{M}_G é a matriz de adjacência normalizada na linha.

A Tabela 1 calcula para cada vértice do exemplo da figura 1 as centralidades de grau, proximidade e intermediação. A Tabela 2 calcula as centralidades de vizinhança para as mesmas medidas.

Tabela 1. Centralidades calculadas para o exemplo da Figura 1

Vértice	Centralidade de Grau	Centralidade de Proximidade	Centralidade de Intermediação
1	0,25	2,0	0,0
2	0,75	1,25	3,0
3	0,50	1,50	0,0
4	0,75	1,25	3,0
5	0,25	2,0	0,0

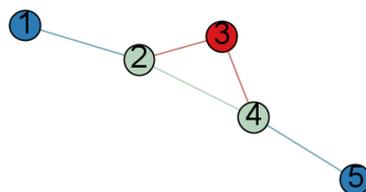
Tabela 2. Centralidades de Vizinhança calculadas para o exemplo da Figura 1

Vértice	Centralidade de Grau	Centralidade de Proximidade	Centralidade de Intermediação
1	0,75	1,25	3,0
2	0,50	1,58	1,0
3	0,75	1,25	3,0
4	0,50	1,58	1,0
5	0,75	1,25	3,0

4. **Agrupamento (*Clustering*)** – Jackson (2008) define *cluster* como um subconjunto de triplas⁷ transitivas, de modo que a partir de um ponto i emanam novos pontos ij e ik , (g_{ij} e g_{ik} , respectivamente), bem como há ligações entre jk (g_{jk}). O coeficiente de agrupamento (*clustering coefficient*) pode ser medido como⁸:

$$Cl_i(g) = \frac{\sum_{j \neq i; k \neq j; k \neq i} g_{ij} g_{ik} g_{jk}}{\sum_{j \neq i; k \neq j; k \neq i} g_{ij} g_{ik}} \quad (7)$$

Para o exemplo o exemplo da figura 1, tem-se os seguintes agrupamentos:

Figura 2. Mapa de agrupamento do exemplo da Figura 1

Elaborado com o software *Gephi*

⁷Triplas são subconjuntos formados por 3 vértices.

⁸Outras formas de se calcular os agrupamentos são estudadas na literatura, embora não tão habituais para grafos não-direcionais (ver Jackson [2008]).

Neste exemplo, tem-se 3 escalas de coeficientes individuais de agrupamento. Os nós 1 e 5, em azul, não formam nenhum conjunto de triplas, uma vez que estão ligados ambos a apenas 1 nó. Neste caso, o coeficiente individual de cluster é 0. Os nós 2 e 4 formam ambos uma tripla com 3, embora não formem com 1 e 5, respectivamente. Portanto, tem-se três triplas transitivas potenciais para 2 ($[1,2,3]$, $[2,3,4]$, $[1,2,4]$) e três triplas positivas potenciais para 4 ($[2,3,4]$, $[2,4,5]$, $[3,4,5]$). Destas, apenas $[2,3,4]$ são formadas. Portanto, para 2 e 4, o coeficiente de agrupamento para estes nós é 0,33. Por fim, o nó 3 possui apenas uma tripla potencial, $[2,3,4]$, que é formada. Portanto, seu coeficiente de agrupamento é 1. O coeficiente médio de agrupamento do grafo é 0,566.

Estas métricas possibilitam incluir nos modelos tradicionais de inovação as características sociais dos nós – indivíduos, firmas ou regiões – para compreender como sua posição na rede afeta sua capacidade de cooperar. Por outro lado, ainda é necessário compreender como as características nodais (que não depende diretamente da sua posição na rede) pode afetar sua capacidade de formar laços. Um desafio a isso, contudo, é como modelar características nodais em um contexto no qual não há hierarquia explícita na relação. A subseção seguinte busca desenvolver uma métrica de modelagem de variáveis nodais na estimação de variáveis dependentes relacionais.

3.2 TRATAMENTO DE CARACTERÍSTICAS NODAIS

A construção do modelo de cooperação parte do pressuposto de que existe uma relação de dominância informacional ou gerencial no laço. Ainda que inexista hierarquia formal identificável na relação entre dois nós, é possível avaliar que é possível que haja influência socioeconômica de uma região sobre outra. Desta forma, em um laço (i, j) , a informação pode ser instrumentalizada pelo conhecimento acumulado na região, pelo potencial inovador desta, e pelo poder econômico. Análogo aos modelos gravitacionais, esta interpretação permite que nós com melhores posições econômicas no laço exerçam maior atração ou possuam maior atratividade perante laços menos favorecidos. Da mesma forma, nós em patamares iguais não possuiriam dominância presumida.

Para a construção da variável que avalie o papel de cada nó no laço, a relação de dominância é descrita por uma equação de diferença absoluta e soma. Por uma

comodidade de análise, a soma será tratada na média, sem que se alterem as interpretações do modelo. Assuma o modelo simples (1) em que a função que determina a variável y é descrita por dois elementos do laço derivados de x : (i) diferença absoluta e (ii) valor médio:

$$y_{ij} = \beta_0 + \beta_1 |x_i - x_j| + \beta_2 (x_i + x_j) \cdot \frac{1}{2} \quad (8)$$

Tem-se pela desigualdade triangular que:

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

Se $x_i, x_j \geq 0 \forall i, j \in V$ e $z_i = \max(x_i, x_j)$ e $z_j = \min(x_i, x_j)$:

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

Valendo-se do lado esquerdo:

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

$$\begin{aligned} \max(x) - \min(x) &= \max(x) + \min(x) \text{ se e somente se } \max(x) \\ &= \min(x) \end{aligned} \quad ^9$$

Portanto:

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

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

$$y_{ij} = \beta_0 + \left(\frac{\beta_2}{2} + \beta_1 \right) \max(x_i, x_j) + \left(\frac{\beta_2}{2} - \beta_1 \right) \min(x_i, x_j) \quad (9)$$

Deste modo, a equação estimada (9) para (i) e (ii) permite construir um coeficiente $\gamma = \left(\frac{\beta_2}{2} \pm \beta_1 \right)$ que estima o efeito da variável x no nó em situação de dominância (*max*) e de dominado (*min*). A vantagem em se estimar em separado os coeficientes β_1 e β_2 permite analisar em separado o efeito da diferença (β_1) e o efeito da escala (β_2). Se β_1 .

⁹ $x = (x_i, x_j)$

Defina $\gamma_1 = \frac{\beta_2}{2} + \beta_1$ e $\gamma_2 = \frac{\beta_2}{2} - \beta_1$, a relação entre o sinal dos coeficientes β e o valor dos coeficientes construídos γ .

Se o coeficiente calculado para a diferença (β_1) é negativo, tem-se que quanto mais diferente é o nó menor é sua conectividade. Já se o coeficiente β_2 é positivo, então nós diferentes são mais propensos à conexão. Por outro lado, se o coeficiente de escala (β_1) for positivo, então quanto maior a média da variável x observada no nó, maior é a propensão à conexão. Se o coeficiente β_1 assumir um valor negativo, então quanto maior a média de x , menor é a conectividade do laço.

O coeficiente do nó em situação de dominância (γ_1) e do dominado (γ_2) no laço pode então ser analisado. Se assumir valores positivos, então a situação do nó é beneficiada na relação e, portanto, existem motivos intrínsecos ao par que melhoram sua percepção dos ganhos da relação. Caso contrário, se o coeficiente γ estimado for negativo, a perspectiva do nó é negativa para a relação, o que o torna menos propenso à formação de laços. O Quadro 1 sistematiza a relação entre o sinal dos coeficientes β e o sinal dos coeficientes γ , auxiliando na interpretação deste.

Quadro 1. Relação de sinal dos coeficientes de dominância

Coeficiente	$\beta_1 > 0$	$\beta_1 < 0$
$\beta_2 > 0$	$\gamma_1 \geq 0$ $\gamma_2 \geq 0$	$\gamma_1 \leq 0$ $\gamma_2 \geq 0$
$\beta_2 < 0$	$\gamma_1 \leq 0$ $\gamma_2 \geq 0$	$\gamma_1 \leq 0$ $\gamma_2 \leq 0$

3.3 BASE DE DADOS E CONSTRUÇÃO DA REDE

Este trabalho está dividido em duas estimações que tratam a cooperação sob diferentes perspectivas e níveis de agregação. A primeira perspectiva é focada na agregação regional dos dados, enquanto a segunda observa a cooperação ao nível do indivíduo. Para cada estimacão, a base de dados foi construída utilizando diferentes parâmetros de cooperação. Em ambos os casos, porém, utilizam-se os dados de patentes brasileiras obtidos no Instituto Nacional de Propriedade Intelectual (INPI) que abrange o

período de 2000 a 2011. Esta base identifica as patentes depositadas no Brasil segundo a classificação internacional de patentes (*International Patents Classification - IPC*), e associa a estas patentes os inventores cujos nomes foram incluídos no registro, identificados por nome, Código de Pessoa Física (CPF) e endereço do inventor, o que permite a associar com as demais bases utilizadas.

Para a perspectiva regional, os dados foram agrupados segundo a proposta metodológica do IBGE (2013) que divide o território brasileiro em 161 Regiões Intermediárias de Articulação Urbana, de acordo com o conceito de Regiões de Influências da Cidades (REGICs), que agrupa aglomerações econômicas sob influência de cidades centrais da produção nacional. A escolha deste nível de agregação regional se baseia na constatação de que centros urbanos formam áreas de influência coincidentes ou não com os limites administrativos metropolitanos, microrregionais ou das unidades federativas. Este conceito está baseado na ideia de que existem cidades capazes de polarizar um território socioeconômico, como na teoria de Christaller (1966). Além disso, os dados geolocalizados da REGIC permitiram a construção de variáveis que capturem as características socioeconômicas das regiões inventoras centrais e do seu entorno direto. O Quadro 2 descreve as variáveis utilizadas e a fonte dos dados.

Quadro 2. Descrição das bases de dados auxiliares utilizadas na análise regional, 161 REGICS, 2000-2011

Variável	Fonte
Produto Interno Bruto <i>per capita</i>	Contas Municipais (IBGE)
	Censo Demográfico de 2000 e 2010 (IBGE)
Capacidade de realizar pesquisa e desenvolvimento (P&D)	Relatório Anual de Informações Sociais (MTE)

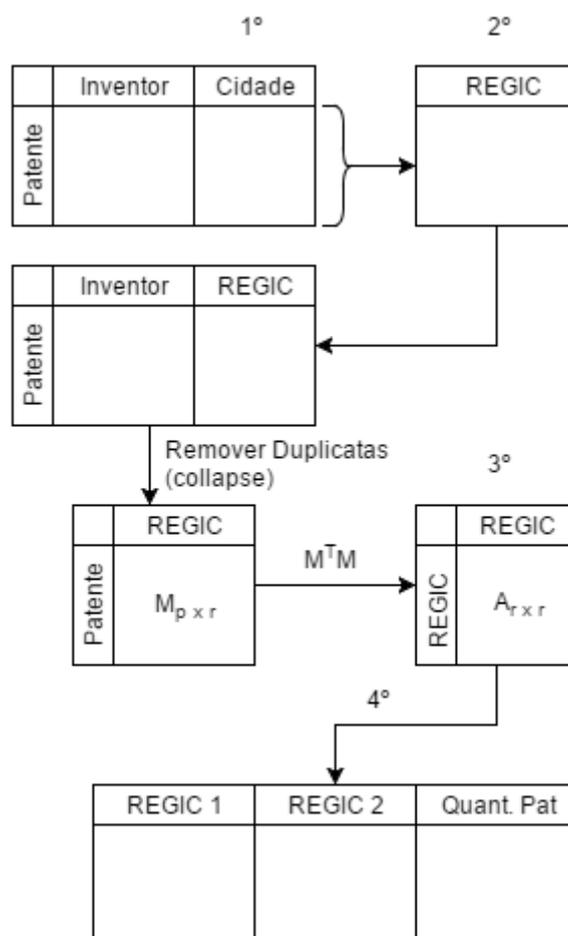
Para a perspectiva individual, foi amplamente utilizada a base de dados do Relatório Anual de Informações Sociais do Ministério do Trabalho e Previdência Social (RAIS/MTE). Esta base contém os dados de controle da atividade trabalhista no Brasil e permite a extração de informações sobre características socioeconômicas dos indivíduos

atuantes no mercado de trabalho. Contudo, apesar da riqueza de informações, esta base também apresenta uma limitação ao incluir apenas dados sobre trabalho formal. Por outro lado, sua utilização pode ser justificada uma vez que a figura do inventor não empregado e que coopera é rara e pontual.

3.3.1 Construção da rede regional

A construção da base regional baseia-se na agregação das patentes identificadas pelo endereço do inventor em regiões de influência das cidades (REGICs), em torno das quais cidades estão polarizadas. O procedimento adotado para este fim segue os seguintes passos: 1º) identificação dos inventores e coinventores por cidade de residência; 2º) agregação das cidades em REGICs; 3º) criação de uma matriz quadrada simétrica que associe a quantidade de cooperações em patentes a cada par de REGICs; 4º) transformação desta matriz em uma lista relacional. Estes passos estão esquematizados na Figura 3.

Figura 3. Esquema de Construção da Base Regional



Fonte: Elaboração própria

A fim de melhor capturar características estruturais e reduzir as variações que podem ocorrer eventualmente de um ano para outro, o período analisado foi agrupado em quatro subperíodos trienais. Dessa forma, o primeiro período compreende o intervalo 2000–2002, o segundo 2003–2005, o terceiro 2006–2008 e o quarto 2009–2011. As variáveis trienais foram criadas pela média dos triênios, de modo que a variação pudesse ser atenuada.

3.3.2. Construção da rede individual

Dois procedimentos podem ser utilizados para criar uma base de copatentes. O objetivo é a partir da base que associa as patentes aos inventores identificados por número de CPF gerar uma nova base que relacione inventores a coinventores. Duas formas são válidas para se fazer isso. Na primeira, para cada patente é formado um par. Neste par, têm-se todas as relações existentes entre inventores para cada código de patente (Figura 4). A vantagem deste método é visualizar as conexões formadas entre cada inventor para cada patente e, assim, extrair características como a classe tecnológica de maior cooperação para cada inventor. A desvantagem deste primeiro método é o tamanho da matriz de adjacência, que terá o tamanho do quadrado do produto entre a quantidade de patentes e da quantidade de inventores.

Figura 4. Primeiro procedimento de conversão de bases¹⁰

Patente	Inventor		Patente	Inventor	Coinventor
1	A	→	1	A	B
1	B		1	A	C
1	C		1	B	C
2	B		2	B	C
2	C				

¹⁰ Chama-se de “inventor” e “coinventor” ambos indivíduos que cooperam em um laço na patente, sem distinção hierárquica. A fim de padronizar a nomenclatura e evitar duplas contagens, utiliza-se a numeração do CPF do indivíduo, com 9 dígitos, e classifica-se o indivíduo com maior numeração de “inventor”, e o com menor numeração de “coinventor”.

Fonte: Elaboração própria

Um segundo procedimento para transformar a base é a partir do apresentado na seção 2.2, em que se extrai a quantidade de copatentes de um par de inventores. Este método consiste em criar uma matriz de adjacência mais simples, porém mais compacta e computacionalmente viável. Sua vantagem é exibir de modo mais claro a força de uma cooperação pela quantidade de laços formados entre dois inventores, sua desvantagem é perder a qualidade de explorar classes tecnológicas no interior da relação. Uma segunda vantagem deste método é exibir na diagonal da matriz criada a quantidade de patentes que tem o nome do inventor associado. A Figura 5 exemplifica a criação desta base.

Figura 5. Segundo procedimento para criação de base de copatentes

Patente	Inventor			Inventor	Coinventor	Laços
1	A	→	$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$	A	B	1
1	B			A	C	1
1	C		↓	B	C	2
2	B		$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 2 \end{bmatrix}$			
2	C					

Fonte: Elaboração própria

Na segunda estimação ambos procedimentos serão utilizados. Para a extração de informações sobre a similaridade tecnológica entre dois inventores será utilizado o primeiro método, que permite visualizar a classe tecnológica das patentes pelo código do IPC. Já para a elaboração de estatísticas de estrutura e posição dos agentes, como exposto na seção 3.1, será utilizado o segundo método, que permite a construção de grafos. Ademais, para o segundo método as matrizes criadas serão divididas em períodos anuais, trienais e duodecenais (que cobre todo o período). Desta forma é possível capturar características temporais da cooperação.

As subseções seguintes utilizam esta metodologia de construção de base e descrevem a topografia das redes ao nível regional e individual, bem como a distribuição das características nodais.

3.4 DESCRIÇÃO DA REDE BRASILEIRA DE COPATENTEAMENTO

Esta seção se divide em duas subseções dedicadas à análise exploratória da rede ao nível regional e ao nível individual. Esta análise é voltada à distribuição da rede em termos das características dos nós e dos laços, e corrobora para as redes de copatenteamento os resultados encontrados na literatura brasileira acerca da distribuição da inovação e da ciência (ver seção 2.5).

3.4.1 Análise exploratória da rede ao nível regional

O objetivo desta subseção é, sobretudo, compreender a estrutura e distribuição dos dados que serão analisados. Para isto, os dados do BATEPI/INPI serão tratados a fim de se obter padrões de distribuição da rede regional.

No contexto da estrutura brasileira de invenções, a exploração analítica dos dados torna-se imprescindível por alguns fatores intrínsecos a esta. Primeiro, o desenvolvimento histórico das regiões brasileiras foi desigual e isto se reflete na capacidade de desempenho das economias regionais. As regiões que compreendem o Centro-Sul, e em especial o Sudeste, dominam grande parte do cenário econômico nacional, enquanto as regiões Norte-Nordeste formam periferias econômicas nacionais. Além disso, também é possível destacar a formação das metrópoles costeiras e o movimento de interiorização ainda recente, que cria outro eixo de desigualdade econômica no Brasil. Sendo assim, tem-se de antemão a ideia de que tal qual a capacidade para o desenvolvimento econômico, a distribuição da inovação brasileira e de sua rede de copatentes é também desigual (GONÇALVES e ALMEIDA, 2009).

A rede é composta de 80.051 depósitos de patentes, distribuídas entre 161 REGICs (IBGE, 2013) que se conectaram 12.490 vezes entre 2000 e 2011. Considerando apenas os laços únicos, 124 regiões se fizeram presente na rede com 1.083 cooperações. Deste total, em média, cada região se conectou aproximadamente a 13 outras regiões. Entre as 10 regiões que se ligaram ao maior número de regiões (Tabela 3), nota-se que, com exceção de Brasília, todas as cidades polarizadoras estão localizadas no Sul-Sudeste do Brasil, com destaque para o Estado de São Paulo, que possui 4 regiões dentre as 10 mais conectadas. Esta configuração coincide com o “polígono de desenvolvimento” identificado por Lemos (2006).

Tabela 3. Relação das regiões (REGICs) mais conectadas a outras regiões, Brasil, 2000-2011

	REGIC	UF	Regiões Conectadas
1	São Paulo	SP	91
2	Rio de Janeiro	RJ	77
3	Belo Horizonte	MG	71
4	Campinas	SP	69
5	Curitiba	PR	61
6	Brasília	DF	58
7	Ribeirão Preto	SP	57
8	Florianópolis	SC	57
9	São Carlos	SP	54
10	Porto Alegre	RS	53
11	Salvador	BA	44
12	Fortaleza	CE	41
13	São José dos Campos	SP	40
13	Londrina	PR	40
13	Joinville	SC	40
13	Goiânia	GO	40
17	Recife	PE	39
18	Vitória	ES	38
19	João Pessoa	PB	34
20	Caxias do Sul	RS	32
Média Brasil			13,45

Tabela 4. Regiões (REGICs) com mais laços, Brasil, 2000–2011

	REGIC	UF	Laços	Média de Laços/Região
1	São Paulo	SP	3986	43,80
2	Campinas	SP	2660	38,55
3	Rio de Janeiro	RJ	2512	32,62
4	Belo Horizonte	MG	1449	20,41
5	Curitiba	PR	1115	18,28
6	Porto Alegre	RS	1068	20,15
7	Ribeirão Preto	SP	802	14,07
8	Florianópolis	DF	799	14,02
9	Brasília	SC	799	13,78
10	São Carlos	SP	646	11,96
11	São José dos Campos	SP	523	13,08
12	Fortaleza	CE	466	11,37
13	Recife	PE	419	10,74
14	Salvador	BA	418	9,50
15	Manaus	AM	407	13,57
16	Joinville	SC	386	9,65
17	Araraquara	SP	328	10,58
18	Goiânia	GO	324	8,10
19	Vitória	ES	311	8,18
20	Natal	RN	273	9,10
Média Brasil			155,15	11,53

Se compararmos com as regiões que formaram o maior número de laços (Tabela 4), uma configuração semelhante é observada, com poucas variações entre suas posições. Sozinhas, estas 10 regiões com maior penetração na rede concentram 63,4% do total de conexões da rede inter-regional de copatenteamento brasileira. É válido ainda destacar o perfil macrorregional em relação à cidade-sede das REGICs isoladas (Tabela 5). Estas regiões não se conectaram a nenhuma outra em um ou mais triênios. Enquanto as regiões sem conexões se concentram no N-NE-CO, englobando grande parte das regiões destas regiões, as macrorregiões S-SE, concentram as regiões que possuem ao menos uma conexão em todos os triênios analisados. Destaque para o Sudeste, em que 63,6% das regiões se conectam a alguma outra em todos os períodos.

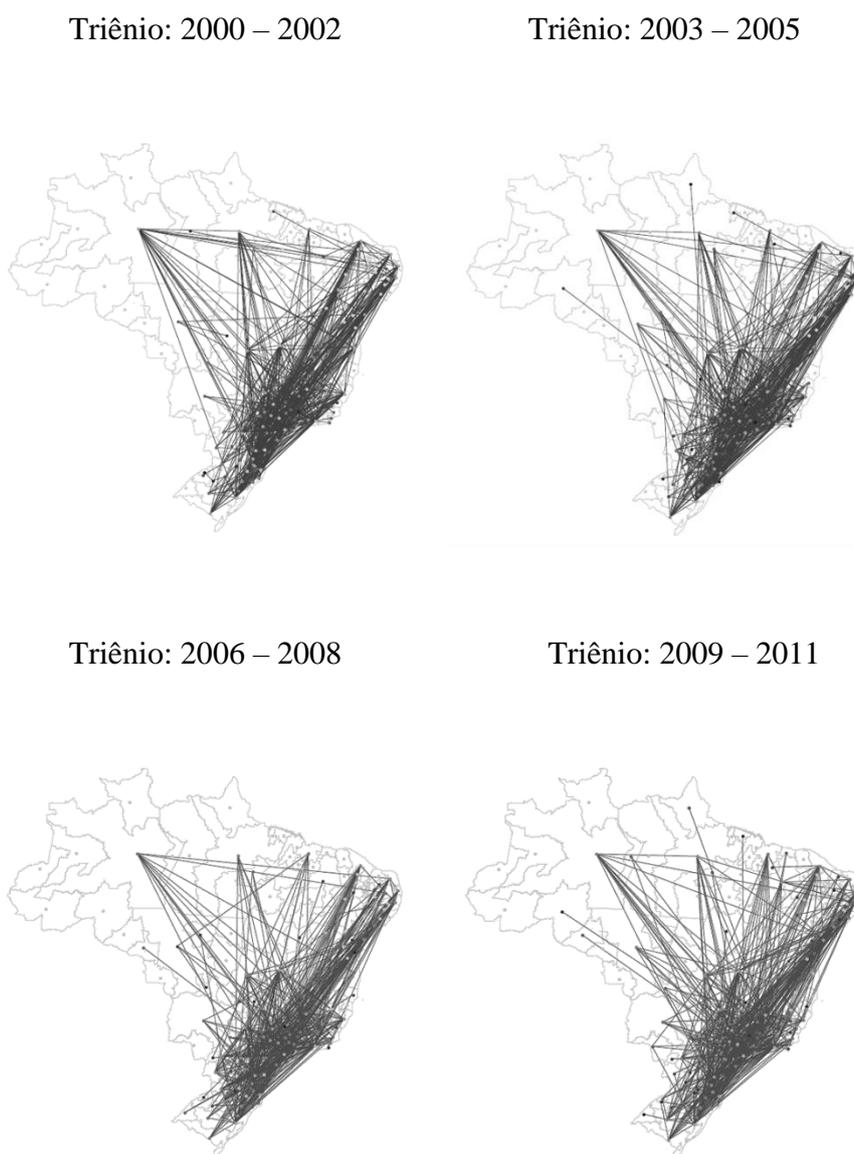
Tabela 5. Distribuição macrorregional dos nós sem conexões ($y = 0$) e com conexões ($y = 1$) em todos os triênios, Brasil, 2000-2011

Macrorregião	#regiões sem conexões em nenhum triênio	% relativo ao total de regiões na Macrorregião	#regiões com conexões nos 4 triênios	% relativo ao total de regiões na Macrorregião
Norte	6	33,3%	3	16,7%
Nordeste	29	51,8%	8	14,3%
Sudeste	1	2,3%	28	63,6%
Sul	6	20,0%	14	46,7%
Centro-Oeste	7	53,8%	2	15,4%
Total	49		55	

Portanto, é cabível notar algumas características em relação ao padrão de conectividade da rede. Primeiro, esta possui forte agrupamento geográfico e relacional. Segundo, existe grande dispersão entre regiões muito conectadas e pouco conectadas no espaço, sendo as primeiras concentradas no Centro-Sul e as segundas nas regiões Norte e Nordeste. Terceiro, um número pequeno de regiões é responsável pela polarização de grande parte das REGICs. Diante dessas características pode-se intuir que as diferenças socioeconômicas regionais possam ter grande influência sobre a forma como se formam os laços. É no interesse pela identificação desses determinantes que as variáveis de interesse e controle foram selecionadas. Estes resultados se aproximam ao encontrado por Gonçalves (2007) e Gonçalves e Almeida (2007) acerca da distribuição espacial da atividade de inovação e do transbordamento de conhecimento.

Quanto ao perfil de confinamento, 29,9% são conexões formadas entre regiões de influência de cidades da mesma UF, enquanto 70,1% são formadas entre regiões de influência de cidades de diferentes unidades da federação. Já entre as macrorregiões, 54,2% são laços entre regiões da mesma macrorregião, enquanto 45% são laços formados fora da macrorregião. Uma forma de visualizar estas características dos laços é pelo grafo espacialmente distribuído da rede de conexões entre as regiões para cada triênio.

Figura 6. Mapas de Redes Trienais de Copatentes, REGICs Brasil, 2000-2011



Os mapas acima (Figura 6) representam a distribuição geográfica de todos os nós da matriz relacional. A cor de cada nó, ou cidade de influência, indica um grau de conectividade maior quanto mais escuro for. Nota-se, a princípio, algumas características

imutáveis da rede ao longo dos períodos: (1) os nós com maior grau de conectividade se aglomeram na Região Sudeste; (2) os nós da Região Norte possuem pouca ou nenhuma conectividade; (3) os laços estabelecidos pelas cidades das regiões Norte e Nordeste tem como destino, principalmente, cidades do Sudeste do país; (4) a região de influência de Brasília parece ser um ponto isolado na região central do país.

As estatísticas de rede são calculadas na Tabela 6. Entre os triênios se observa pouca variação nas estatísticas, embora algumas estatísticas possam revelar alguma tendência. Enquanto o índice de *clustering* médio teve uma trajetória de declínio nos últimos três períodos, a centralidade de proximidade média aumentou nos últimos dois. Embora essa alteração não seja expressiva, pode indicar que a rede tem se tornado menos concentrada em torno de alguns nós centrais à medida em que os nós aumentam a facilidade com que se conectam com os demais vértices.

Tabela 6. Estatísticas da rede agregada por REGIC, Brasil, 2000-2011

Centralidades (Médias)	2000-2011	2000-2002	2003-2005	2006-2008	2009-2011
Grau (Densidade)	0,153028	0,036413	0,0387422	0,0371118	0,0407609
Proximidade	0,3802044	0,2476143	0,2884611	0,2659116	0,3009317
Intermediação	51,254658	33,801242	31,875776	34,354037	46,329193
<i>Clustering</i> Médio	0,470831	0,4647196	0,4741984	0,4518794	0,4196146
Grau Médio	24,484472	5,826087	6,1987578	5,9378882	6,5217391
Média Anual	2,0403727	1,942029	2,0662526	1,9792961	2,173913

Utilizando a abordagem do coeficiente de dominância relacional apresentado em 3.2, pode-se dividir os nós em duas funções: máximo e mínimo. Assim, as características nodais das regiões no laço podem ser decompostas. A Tabela 7 apresenta as informações de correlação entre as variáveis de PIB *per capita* e de número de pesquisadores da região (POTEC). É notável que o PIB *per capita* entre as regiões que cooperam possui uma correlação elevada de 0,5909. O número de pesquisadores possui resultado também considerável, com 0,4392 de correlação. Estes valores indicam, de antemão, que regiões com valores mais elevados tendem a se conectar mais.

Tabela 7. Correlação de características nodais das regiões, 2000-2011

	Max(PIB)	Min(PIB)	Max(POTEC)	Min(POTEC)
Max(PIB)	1			
Min(PIB)	0,5909	1		
Max(POTEC)	0,4665	0,273	1	
Min(POTEC)	0,2185	0,4738	0,4392	1

3.4.2 Análise exploratória da rede ao nível individual

Para esta análise, foi utilizada a base de dados do BADEPI/INPI, tratada para se obter laços entre inventores. Para cada patente na base foi construído o subconjunto de arestas formado entre dois inventores. Em seguida, cada inventor foi pareado com os dados disponíveis na RAIS/MTE pelo CPF disponível. Uma limitação desta abordagem, contudo, é que permite identificar apenas inventores que estavam formalmente empregados no momento da invenção. Por outro lado, o uso da RAIS apresenta a vantagem de capturar informações apenas dos inventores ditos “profissionais” e que cooperam em contraposição aos “inventores de garagem”. Desta forma é garantido que todos os pares formados possuem informação completa para ambos inventores.

As informações oriundas da compatibilização das bases permitiram elaborar perfis socioeconômicos dos inventores. Neste trabalho, as características pessoais extraídas podem ser divididas em três grupos: (1) pessoais, (2) ocupacionais e (3) geográficas, conforme Quadro 3.

Quadro 3. Classificação das características individuais utilizadas, 2000-2011

Tipo de Característica	Característica
Pessoais	Sexo
	Escolaridade
Ocupacionais	Área de atuação
	Rendimento
Geográfica	Coordenadas Municipais
	Densidade geográfica da rede

A construção da rede se deu pela metodologia apresentada na seção 3.3.2. Nesta, para cada par $(i, j) \in E$ tem-se que $g_{ij} = \sum_{h=1}^p g_{ij,h}$, onde $h = 1, \dots, p$ é o conjunto de p patentes na rede. Se $(i, j) \in h$ então $g_{ih} = 1$, caso contrário, 0. Em seguida, foi adicionado um componente temporal, desagregando g_{ij} pelo triênio em que a relação foi observada.

Assim, tem-se que a rede é composta por 7.733 nós que criam 14.679 laços, no período de 2000 a 2011. O grau médio alcançado neste período é de 1,89 laços únicos por nó. A tabela 8 indica a distribuição dos laços, nós e grau médio por triênio. Nota-se que nos triênios o grau médio dos nós varia de 1 a 2 laços, aproximadamente. Se, por outro lado, for considerada a quantidade de conexões formadas (isto é, g_{ij}), tem-se uma medida de conexões médias, ou quantos laços são formados em média na rede, com possibilidade de reconexão no mesmo triênio. Nesta estatística, cada nó formou ao longo de todo o período uma média de 2,452 conexões de copatentes.

Tabela 8. Distribuição trienal dos nós, laços e grau médio, inventores, 2000-2011

Triênio	Nós	Laços Únicos	Grau Médio	Conexões	Conexões Médias
2000-2002	1.374	2.764	2,011	3.022	2,199
2003-2005	1.983	1.999	1,008	2.750	1,387
2006-2008	2.838	4.766	1,679	5.603	1,974
2009-2011	3.509	5.150	1,467	7.610	2,169
2000-2011	7.733	14.679	1,898	18.985	2,452

Utilizando a abordagem de Owan e Nagaoka (2011), segundo a qual as características dos nós possuem funções intrínsecas e extrínsecas na rede (ver 2.3.3), as classificações das características nodais do Quadro 3 podem ser reinterpretadas em quatro funções: (i) Geográficas, (ii) Relacionais, (iii) Intrínsecas e (iv) Extrínsecas. As variáveis geográficas são incluídas para capturar o conhecido papel do espaço na formação dos laços. A centralidade de proximidade é inserida como uma variável relacional para capturar a facilidade de conexão na rede dos indivíduos. As variáveis de motivação intrínseca e extrínseca incluem controles que capturam diferenças de conexão que capturam a percepção das características inventivas do parceiro para o primeiro tipo, e o papel de fatores externos para o segundo tipo. O Quadro 4 esquematiza estas variáveis.

Quadro 4. Classificação de variáveis do laço e do nó sob a ótica individual

Classificação	Nome	Descrição
GEOGRÁFICA	Distância Geográfica	Distância em quilômetros entre os inventores do par, calculado a partir do município registrado na RAIS.
	Mesmo Município	<i>Dummy</i> que indica se o par é formado por inventores residentes no mesmo município (= 1) ou em municípios diferentes (= 0).
RELACIONAL	Centralidade de Proximidade	Calcula o quão fácil é para um nó se conectar aos demais nós da rede.
INTRÍNSECO	Grau de Instrução	Grau de instrução educacional de cada indivíduo do par
	Rendimento	Mensura, em salários mínimos, a remuneração mensal de cada indivíduo do par.
EXTRÍNSECO	Mesma indústria	<i>Dummy</i> que indica se o par é formado por indivíduos da mesma área de atuação industrial (2 dígitos da CNAE 1.0). Busca capturar se o par busca reputação na sua área.
	Academia x Empresa	<i>Dummy</i> que verifica se a ligação é feita entre inventores acadêmicos e empresariais.
	Sexo	<i>Dummy</i> que indica se o par é formado por indivíduos do mesmo sexo (= 1) ou de sexos diferentes (= 0).

Fonte: elaboração própria.

Perfil de Conexões

Uma característica a se analisar na rede é avaliar a distribuição dos tipos de ligação existentes. Esta interpretação possibilita investigar *a priori* quais elementos os inventores buscam ao formar suas expectativas acerca do retorno da conexão. Jung e Ejermo (2014) constroem um panorama demográfico acerca dos inventores suecos. Os autores analisam três características dos inventores a fim de identificar o perfil geral destes: gênero, idade e educação. Estes resultados oferecem um norte acerca das características gerais dos inventores. O objetivo desta análise é identificar os perfis das conexões na rede brasileira

de copatenteamento, e se estes são similares ou apresentam alguma particularidade em relação ao observado nas demais experiências empíricas relatadas na literatura.

A partir dos dados disponíveis, 5 tipos de ligações podem ser extraídos: (i) entre gêneros, (ii) entre níveis educacionais, (iii) entre níveis de renda, (iv) entre academia e empresa, (v) entre municípios. Estas ligações serão descritas nas tabelas abaixo.

Tabela 9. Distribuição de Ligação entre Inventores por Gênero no Brasil, por triênio, 2000-2011

Triênios	2000-2002	2003-2005	2006-2008	2009-2011	2000-2011
Mulher-Mulher	96	208	469	515	1.288
Homem-Homem	1.085	1.742	3.888	2.579	9.294
Homem-Mulher	1.841	800	1.246	4.516	8.403
Participação M nas Conexões (%)	64,10%	36,65%	30,61%	66,11%	51,05%
Participação H nas Conexões (%)	96,82%	92,44%	91,63%	93,23%	93,22%
% Mulheres Inventoras	23,80%	22,34%	25,16%	25,76%	24,738%

Fonte: Elaboração própria.

Ao se dividir a formação do par pelo gênero do inventor e do coinventor é possível intuir a presença de uma discriminação baseada em gênero na rede de inventores (Tabela 9). Nota-se, a princípio, que homens são grande parte dos inventores da rede, correspondendo a 75,26% e representam percentual ainda maior de participação nos pares, 93,22%. Já a presença feminina é bastante inferior, representando 24,74% dos laços e 51,05% das conexões. Ainda vale destacar que enquanto as ligações entre mulheres formam apenas 13,30% do total de participação feminina, as ligações entre homens representam 52,51% das ligações com presença masculina. Esta distribuição pode indicar que as ligações femininas são periféricas na rede e a inserção de mulheres ocorre, principalmente, em colaboração com homens, o que indica que os inventores discriminam a formação de laços pelo gênero do parceiro. Por outro lado, a presença de mulheres entre os inventores mostra uma trajetória levemente ascendente ao longo dos triênios, representando cerca de um quarto do total de inventores que cooperam.

Na literatura internacional, Whittington e Smith-Doerr (2008) encontram evidências para a formação de redes baseadas em gêneros utilizando dados do *U.S. Patent and Trademark Office* (USPTO). Já Gaughan e Bozeman (2016) realizaram entrevistas anônimas com cientistas acadêmicos de universidades americanas. Os resultados encontrados sugerem que a interpretação do benefício da colaboração científica está associada ao gênero do par. Jung e Ejerme (2014) identificaram que a diferença na participação de homens e mulheres no sistema de invenções diminuiu ao longo do período, embora em escala inferior se comparado a outras áreas da sociedade.

A Tabela 10 indica as conexões formadas segundo grau de instrução do inventor e do coinventor. Os indivíduos do laço são classificados em 1 e 2 conforme seu nível educacional observado no laço: se maior, 1; se menor, 2. Esta configuração se adequa à utilizada pelo coeficiente de dominância relacional apresentada na seção 3.2. No geral, é possível observar uma predileção à indivíduos com nível educacional mais elevado de se conectar entre si. De fato, a correlação calculada entre os anos de estudo dos inventores 1 e 2 é de 0,499 (Tabela 18). Além disso, indivíduos com ao menos ensino superior (16 anos, em média) representam grande parte das conexões, 87,26%. Nota-se ainda que aqueles indivíduos que possuem menor grau de instrução buscam conexões com aqueles com maior nível educacional. Vale ressaltar, contudo, que até 2002 a RAIS não identificou indivíduos mestres e doutores, sendo esta informação disponibilizada apenas nos anos seguintes. Desta forma, não foi possível uma avaliação precisa sobre esses indivíduos, agregando-os na classe “16+”.

Em relação à renda, a Tabela 11 exhibe o número de conexões formadas entre quantis de renda dos inventores 1, com maior renda no laço, e 2, com menor renda no laço. Nota-se, para facilitar a análise da tabela, que os valores médios dos quantis não são coincidentes. Enquanto o valor médio para os inventores com maior renda no laço varia de 10,77 a 74,03, o valor médio para os inventores de menor renda no laço varia de 1,81 a 33,47. Esta variação já revela, de antemão, uma desproporção na distribuição dos laços. A correlação entre as rendas dos inventores é 0,6330 (Tabela 12) e 76% dos laços são formados entre inventores com 30 salários mínimos ou menos.

Tabela 10. Distribuição do número de pares formados entre indivíduos, por ano de estudo, 2000-2011

		Anos de Estudo do Inventor 1								Total
		3	5	7	9	11	12	14	16+	
Anos de Estudo do Inventor 2	5	0	3	0	0	0	0	0	0	3
	7	3	0	1	0	0	0	0	0	4
	9	1	0	3	10	0	0	0	0	14
	11	0	3	1	3	5	0	0	0	12
	12	1	2	11	33	25	195	0	0	267
	14	0	1	1	6	5	96	33	0	142
	16+	4	5	15	47	38	937	381	12.810	14.237
Total		9	14	32	99	73	1.228	414	12.810	14.679

Nota: Para fins de exibição dos dados sem incorrer em dupla-contagem de conexões, os inventores foram classificados entre 1 e 2, sendo, respectivamente, o com maior escolaridade e o com menor escolaridade observada no laço.

Fonte: Elaboração própria.

Tabela 11. Distribuição do número de pares formados entre indivíduos por quantil de renda (em salários mínimos), 2000-2011

		Quantil de Renda do Inventor 1										Total	
Quantil		1	2	3	4	5	6	7	8	9	10		
		Valor Médio	10,77	16,00	18,35	21,06	23,21	25,17	29,95	35,64	46,74	74,03	24,85
Quantil de Renda do Inventor	1	1,81	409	340	261	101	1.679	11	-	-	-	-	2.801
	2	4,58	107	136	1.684	105	170	208	82	-	-	-	2.492
	3	7,81	82	74	114	72	117	141	211	114	-	-	925
	4	9,03	66	67	79	39	83	139	160	170	59	-	862
	5	11,49	68	65	79	50	64	130	197	182	211	6	1.052
	6	13,90	50	53	49	25	56	103	120	133	1.765	102	2.456
	7	16,04	54	58	49	33	72	110	150	120	164	257	1.067
	8	18,82	50	39	60	34	62	84	113	114	172	394	1.122
	9	23,20	41	47	48	31	44	60	119	162	167	720	1.439
	10	33,47	11	9	13	13	18	18	21	25	32	303	463
Total	15,51	938	888	2.436	503	2.365	1.004	1.173	1.020	2.570	1.782	14.679	

Nota [1]: Para fins de exibição dos dados sem incorrer em dupla-contagem de conexões, os inventores foram classificados entre 1 e 2, sendo, respectivamente, o com maior renda em salários mínimos e o com menor renda em salários mínimos observada no laço.

Nota [2]: A linha e a coluna “Valor Médio” indica o valor médio do quantil de renda do inventor 1 e 2 em salários mínimos.

Fonte: Elaboração própria.

Tabela 12. Correlação entre máximos e mínimos de educação e renda no laço, Brasil, 2000 - 2011

	Max(Renda)	Min(Renda)	Max(Educação)	Min (Educação)
Max(Renda)	1,0000			
Min (Renda)	0,6330	1,0000		
Max(Educação)	0,0941	0,0258	1,0000	
Min(Educação)	0,0835	0,1721	0,4909	1,0000

A Tabela 13 exibe a distribuição trienal das ligações em três categorias: (i) ligações entre inventores ligados às empresas (excluindo universidades), (ii) ligações entre inventores ligados às universidades, e (iii) ligações entre inventores de empresas e universidades. Duas situações podem ser percebidas ao longo do período analisado. Primeiro, as ligações são, em grande parte, formadas entre inventores empresariais. Segundo, inventores associados a universidades aumentaram sua participação nas ligações, passando de 10,4% de participação nas conexões no primeiro triênio para 33,6% de participação nas conexões do último triênio. No período de 2000 a 2011, inventores universitários participaram de 27,3% das conexões formadas. Portanto, embora haja uma continuidade na predominância das ligações formadas entre inventores empresariais, a maior participação da universidade na rede de inventores pode revelar algumas características da evolução institucional da rede.

Tabela 13. Distribuição trienal de frequência por conexões entre empresas e universidades, Brasil, 2000-2011

Tipo de Ligação	2000- 2002	2003- 2005	2006- 2008	2009- 2011	2000- 2011
Empresa-Empresa	2.707	2.109	3.937	5.051	13.804
Empresa-Universidade	192	359	876	841	2.268
Universidade- Universidade	123	282	790	1.718	2.913
Total	3.022	2.750	5.603	7.610	18.985

Por fim, o perfil do laço pode ser descrito segundo a força gravitacional exercida pela proximidade geográfica ou fronteiras políticas. A Tabela 14 expressa a quantidade de conexões estabelecidas, por triênio, para dois casos: se ambos inventores são residentes do mesmo município ou não. Com exceção do terceiro triênio, os demais revelam que

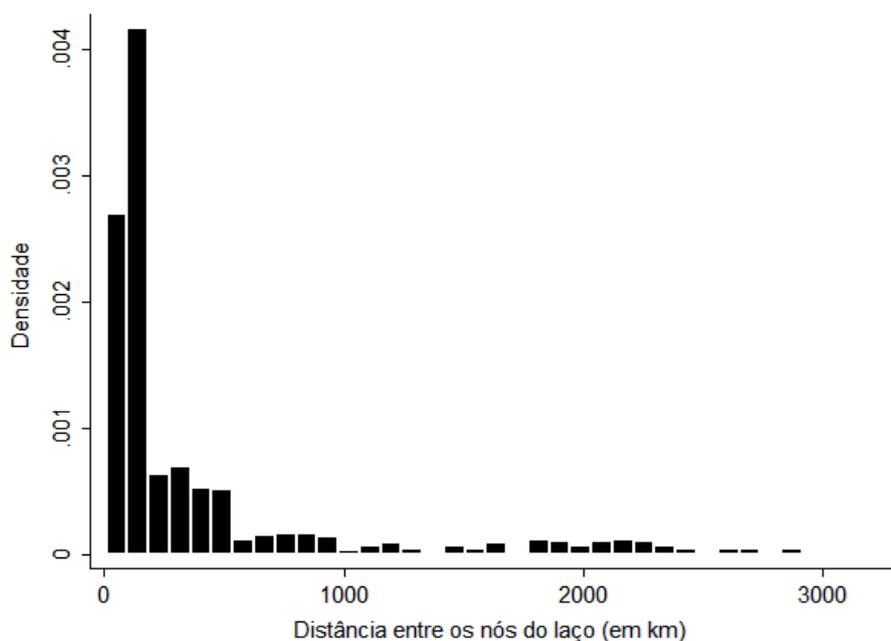
grande parte dos laços são formados entre indivíduos no mesmo município. Com o auxílio do Gráfico 1, que dispõe o histograma da distância dos laços intermunicipais, vê-se que mesmo estes laços estão, em grande maioria, restritos a menos de 1000 km de distância, mesmo com a grande dimensão territorial brasileira.

Tabela 14. Distribuição trienal de frequência por conexões entre indivíduos do mesmo município ou de municípios diferentes, Brasil, 2000-2011

	2000-2002	2003-2005	2006-2008	2009-2011	2000-2011
Municípios Diferentes	438	697	3.021	1.402	5.558
Mesmo Município	2.584	2.053	2.582	6.208	13.427
Total	3.022	2.750	5.603	7.610	18.985
% Mesmo Município	85,51%	74,65%	46,08%	81,58%	70,72%

Fonte: Elaboração Própria.

Gráfico 1. Histograma da distância dos laços intermunicipais, ponderado pelo número de conexões, Brasil, 2000-2011



4 DETERMINANTES DA FORMAÇÃO DE LAÇOS INTER-REGIONAIS

4.1 MÉTODO DE ESTIMAÇÃO

Como um dos objetivos do trabalho é identificar os determinantes da formação de laços entre duas regiões, propõe-se a estimação de um modelo de contagem. Em comparação aos modelos probabilísticos tradicionais de resposta binária, como o *probit* e o *logit*, os modelos de contagem possuem uma clara vantagem na atribuição de pesos à ocorrência do evento. Desta maneira, a ligação entre duas regiões i e j pode ser descrita por $y_{ij} \in [0, z]$, em que se não há ligação observada y assume valor 0, caso contrário, y é igual as z ligações observadas.

Na elaboração da rede são considerados todos os laços possíveis e não-direcionais partindo de uma região. Nesta configuração, não há hierarquia aparente entre os nós, sendo a informação duplamente compartilhada no laço. Desta forma, para cada período analisado é possível a formação de $[n(n - 1)]/2$ laços potenciais na rede, sendo n o número de regiões, de modo que cada uma das 161 regiões pode formar laços com as demais 160. Portanto, seja y o somatório de laços binários (1 caso formado e 0, caso contrário), tem-se $y - [n(n - 1)]/2$ é o número de laços possíveis sem conexão observada no período. Em um universo com poucos nós, em que há um considerável número de subunidades formadoras do nó regional, é razoável supor que existe a possibilidade real de que cada nó possa se ligar aos demais $n - 1$ nós. Este pressuposto é importante para a construção do contrafactual dos laços formados.

A escolha do modelo de contagem adequado deve ter em conta as características estatísticas da variável dependente a estimar. De modo a investigar as características das variáveis utilizadas, será feita uma análise da topografia de rede dos laços inter-regionais. Na Tabela 15 estão descritos os percentuais de laços formados ($y > 0$) em relação à quantidade de laços potenciais em cada triênio. Nota-se, de antemão, que as regiões são pouco conectadas em relação às conexões potenciais totais. Esta informação, complementa o obtido na tabela 6. Em média, as regiões se conectam a outras 6,12 regiões em cada triênio, dentre as 160 ligações possíveis, um valor relativamente baixo se comparado às regiões mais conectadas (Tabela 3).

Tabela 15. Percentual de Laços Formados entre REGICs por Triênio, Brasil, 2000-2011

Y	2000-2002	2003-2005	2006-2008	2009-2011
Laço não formado ($y = 0$)	96,36%	96,13%	96,29%	95,92%
Laço Formado ($y > 1$)	3,64%	3,87%	3,71%	4,08%

Fonte: Elaboração Própria.

Considerando apenas as regiões que se conectaram a alguma outra no triênio, de modo que pontos isolados não sejam considerados, existem valores significativamente maiores (Tabela 16). É possível observar também o aumento do número de regiões que possuem alguma conexão ao longo do período, atingindo dois terços do total de REGICs.

Tabela 16. Percentual de Laços Formados entre REGICs por Triênio, regiões não isoladas, Brasil, 2000-2011

Y	2000-2002	2003-2005	2006-2008	2009-2011
Laço não formado ($y = 0$)	88,29%	89,06%	89,06%	90,91%
Laço Formado ($y > 0$)	11,71%	10,94%	10,94%	9,09%
#Regiões com 1 laço ou mais	89	95	93	107
% de Regiões conectadas	55,3%	59,0%	57,8%	66,5%

Fonte: Elaboração Própria.

Uma importante consequência da baixa conectividade dos nós se reflete na baixa média do número de ligações e no alto desvio padrão, conforme descrito na Tabela 17. Hilbe (2011, p. 141) descreve que se o desvio padrão é bastante superior à média, então há superdispersão na variável dependente que leva a requerer a elaboração de modelos de contagem. Sugere-se utilizar a estatística qui-quadrada de Pearson dividido pelo número de graus de liberdade do modelo de Poisson estimado por GLM para avaliar o grau de dispersão. Se a estatística for superior a 1,25 então é necessário um modelo corretivo nos dados. Esta estatística será calculada na definição do modelo.

Lawless (1987) ressalta que os efeitos da superdispersão dos dados nos modelos de contagem baseados na distribuição de Poisson levaram ao estudo de métodos que melhor acomodem essas características, como o modelo binomial negativo. Hilbe (2011) escreve que a distribuição Poisson pode não ser adequada para descrever a dispersão, o que leva à adoção do modelo binomial negativo que utiliza uma mistura Gama-Poisson para correção deste problema.

Tabela 17. Média e Desvio Padrão das ligações por Triênio, Brasil, 2000-2011

Triênio	Média	Desvio Padrão
2000-2002	0,234317	3,837151
2003-2005	0,259162	3,024301
2006-2008	0,217003	3,08451
2009-2011	0,259239	3,244838
2000-2011	0,2424301	3,244838

Fonte: Elaboração Própria.

A estimação binomial negativa aqui descrita segue a formulação de Lawless (1987). Uma vez que o objetivo da dissertação não é contribuir metodologicamente para o método de estimação, o modelo será apenas apresentado sem aprofundamentos. Seja Y a variável de contagem, e x um vetor $p \times 1$ de variáveis explicativas. O modelo Binomial Negativo é escrito como:

$$P(y = Y|x) = \frac{\Gamma(y + a^{-1})}{y! \Gamma(a^{-1})} \left(\frac{a\mu(x)}{1 + a\mu(x)} \right)^y \left(\frac{1}{1 + a\mu(x)} \right)^{a^{-1}}$$

onde Γ é a função de distribuição binomial negativa, a e μ são parâmetros. O modelo é estimado por máxima verossimilhança.

Além da estimação para todo o período 2000-2011, o modelo binomial negativo será aplicado também ao painel balanceado dos 4 triênios entre 2000 e 2011. Estas estimações serão comparadas com demais estimações em painel (Poisson e GLS), bem como serão realizados testes para corroborar o ganho da estimação. Serão utilizados modelos com efeitos fixos e efeitos aleatórios a fim de capturar possíveis diferenças entre os modelos.

4.2 DESCRIÇÃO DAS VARIÁVEIS

Variáveis de interesse

A estimação a ser realizada visa capturar os efeitos de três componentes a seguir:

- (1) Distância Geográfica (Geo_{ij}): mede se há um efeito de decaimento espacial na possibilidade de conexão. Quanto maior a distância geográfica entre duas regiões, menor deve ser a propensão de conexão, o que deve se refletir num

sinal negativo para este coeficiente. Esta variável se baseia na distância geodésica entre o centroide das cidades polarizadoras da região.

- (2) Proximidade Relacional (Rel_{ij}): é medida pelo Índice de Jaccard que tenta captar o efeito das conexões indiretas ou a ligação indireta de primeira ordem (vizinhos de vizinhos). Neste sentido, espera-se que duas regiões que possuam muitas conexões indiretas tenham maior propensão a se conectar, ou seja, um sinal positivo para este coeficiente. O Índice de Jaccard (*Jaccard index*) é uma métrica topológica baseada em vizinhanças utilizadas para predição de ligações entre dois nós na rede, segundo Wasserman e Faust (1995) e Jackson (2008). Esta estimativa usa este índice para mensurar a proximidade relativa de dois pontos. Considera a razão entre o número de vizinhos comuns entre dois pontos i e j ($|\Gamma_i \cap \Gamma_j|$) e o total de vizinhos dos dois pontos ($|\Gamma_i \cup \Gamma_j|$):

$$J(i, j) = \frac{|\Gamma_i \cap \Gamma_j|}{|\Gamma_i \cup \Gamma_j|} \quad (10)$$

- (3) Proximidade Tecnológica¹¹ (Tec_{ij}): tenta medir se a similaridade tecnológica entre duas regiões aumenta as suas chances de formarem uma nova conexão. É utilizado o índice de Jaffe (1986), que é construído com as classes de patentes para mensurar a proximidade tecnológica entre duas regiões. Quanto maior for o grau de patenteamento nas mesmas classes tecnológicas, maior é o índice, que indica similaridade tecnológica. Esta proximidade pode ter dois tipos de efeitos: (i) regiões mais próximas tecnologicamente possuem maior afinidade cognitiva, o que facilita a formação de laços; (ii) regiões podem buscar outras com menor semelhança tecnológica a fim de alcançar complementaridade tecnológica. Logo, o sinal do coeficiente pode ser positivo ou negativo.

¹¹A proximidade tecnológica é utilizada aqui como uma *proxy* institucional para a diferença cognitiva existente entre duas regiões. A ideia para a escolha deste nível institucional é de que nós cognitivamente diferentes possuem maneiras distintas de tratar problemas, criar soluções e gerar novas ideias. Disso, advém o duplo efeito apresentado para explicar como a proximidade cognitiva-tecnológica pode afetar a propensão a cooperar. Por um lado, o maior entendimento ou capacidade de absorção pode aumentar e tornar mais fácil a difusão de conhecimento em uma aresta. Por outro lado, inventores podem buscar conhecimentos complementares aos seus, o que cria um efeito contrário do primeiro: nós cognitivamente distantes aumentam a probabilidade de conexão.

Variáveis de controle

Para controlar o efeito que outros fatores podem ter sobre a propensão de formação de laços, foram incluídos (as): (i) nível de riqueza da região; (ii) posição relativa do nó na rede e (iii) esforço regional em atividades de pesquisa.

(i) PIB *per capita* (PIB_{ij}): pretende capturar se as conexões tendem a se formar entre regiões nível de desenvolvimento similar (MAGGIONI, NOSVELLI e UBERTI, 2007; MAGGIONI e UBERTI, 2007). O PIB *per capita* das regiões é uma medida construída com base em dados do Produto Interno Bruto calculado pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e de dados de população da Pesquisa Nacional por Amostragem de Domicílios (PNAD) e Censos Demográficos de 2000 e 2010. A medida aqui usada refere-se à diferença absoluta entre o logaritmo natural do PIB per capita das regiões i e j que formam laços potenciais.

(ii) Grau de centralidade ($CENT_{ij}$): tenta captar a posição relativa do nó na rede, ou seja, se o nó é uma região central e tem um *score* com amplitude de 0 a 100. A variável é construída pela diferença absoluta entre o logaritmo natural da centralidade relativa das regiões i e j . É esperado, como identificado pela hipótese 1, que regiões com centralidade próxima se conectem mais facilmente entre si (CRESCENZI, NATHAN e RODRIGUEZ-POSE, 2016). Logo, espera-se coeficiente positivo para esta variável. Com base em Wasserman e Faust (1995) e Jackson (2008), esta estimação usa a métrica de Centralidade de Proximidade (*closeness centrality*), que mensura o quão fácil é para um nó se conectar a outros dentro da rede. Para isto, utiliza como medida de proximidade o número de ligações com menor caminho ℓ entre um nó i e os demais j nós da rede:

$$CC_i = \frac{(n - 1)}{\sum_{i \neq j} \ell_{ij}} \quad (11)$$

(iii) Capacidade de realizar pesquisa & desenvolvimento ($P\&D_{ij}$): a variável é construída pela diferença absoluta entre o logaritmo natural de pessoal em ocupações técnicas (POTEC), compostas por profissionais de biotecnologia e biomédicos, engenheiros, pesquisadores, profissionais da matemática e de

estatística, analistas de sistemas computacionais, físicos, químicos, profissionais do espaço e da atmosfera, arquitetos das regiões i e j das regiões.¹² A variável foi construída com dados extraídos da Relação Anual de Informações Sociais, do Ministério do Trabalho e Emprego (MTE), na ausência de dados regionalizados de gastos de P&D. Ejermo e Karlsson (2006) apontam dois possíveis efeitos para variável de P&D. De um lado, regiões com elevada capacidade de pesquisa realizam mais conexões com outras regiões. Por outro lado, estas mesmas regiões podem ser autossuficientes em pesquisa, tendo menos incentivos a se conectar com outras. Num caso espera-se sinal positivo para este coeficiente; noutro, sinal negativo.

No modelo a ser desenvolvido neste capítulo, metodologia para tratamento das características individuais do nó seguirá a decomposição apresentada na seção 3.2. Desta forma, não apenas a diferença relativa dos nós será considerada, como também o peso do laço em relação aos demais. Laços pouco influentes, formados entre regiões pouco relevantes na rede, possuem um peso menor que laços formados entre regiões relevantes ou laços cruzados, entre regiões muito relevantes e pouco relevantes.

4.3 MODELO EMPÍRICO

O modelo proposto pretende estimar a relação das três dimensões (geográfica, relacional e institucional) com a probabilidade de formação de novos laços, e a inclusão de outros controles socioeconômicos. As variáveis de interesse estão em logaritmo natural e as variáveis de controle incluem a diferença e média do logaritmo das variáveis dos nós i e j , conforme descrito anteriormente.

$$\begin{aligned}
 y_{ij} = & \beta_0 + \beta_1 \ln(Geo_{ij}) + \beta_2 \ln(Rel_{ij}) + \beta_3 \ln(Tec_{ij}) & (12) \\
 & + \beta_4 [\ln |\Delta_{ij} PIB|] + \beta_5 [\ln |\Delta_{ij} Cent|] + \beta_6 [\ln |\Delta_{ij} Pesq|] \\
 & + \beta_7 [\sum \ln(PIB)_{ij} / 2] + \beta_7 [\sum \ln(Cent)_{ij} / 2] \\
 & + \beta_7 [\sum \ln(Pesq)_{ij} / 2] + \beta_8 UF + \varepsilon
 \end{aligned}$$

¹²Segundo Araújo, Cavalcante e Alves (2009) e Gusso (2006) há 90% de correlação entre os gastos de P&D externos e internos de firmas e a POTEC, o que indica que esta variável é uma proxy adequada para capacidade de P&D privado.

em que Geo_{ij} indica a distância geográfica entre i e j ; Rel_{ij} , o índice de Jaccard entre i e j ; Tec_{ij} , o índice de Jaffe calculado entre i e j . Para os controles, PIB, indica o produto interno bruto per capita no subscrito indicado; CENT, indica a posição relativa da região (nó) na rede; Pesq, a proporção de trabalhadores empregados em atividades técnicas; e UF indica se i e j são regiões polarizadas por cidades dentro da mesma Unidade de Federação quando igual a 1, e de Unidades diferentes se igual a 0. Para esta especificação empírica, o grau de dispersão de Pearson é 9,057, o que sugere o uso de métodos com correção de superdispersão, como o Binomial Negativo.

4.4 RESULTADOS DA ESTIMAÇÃO

Os resultados do painel trienal do período 2000-2011 são descritos na Tabela 18, divididas em Binomial Negativo, Poisson e GLS, e subdivididas em Efeitos Fixos (EF) e Efeitos Aleatórios (EA). O modelo Binomial Negativo possui melhor ajustamento para variáveis com superdispersão, enquanto os modelos de Poisson e GLS são estimados para controle das variáveis, embora não sejam estimações ideais. As linhas são divididas em 5 blocos: (1) variáveis do laço, (2) variáveis de diferença, (3) variáveis de média, (4) *dummy* de UF e constante. A última linha indica o teste qui-quadrado que avaliar se o modelo de Efeitos Aleatórios é estatisticamente superior à estimação por Painel Agrupado.

Quanto as variáveis do primeiro bloco para a estimação Binomial Negativa, tanto na estimação de efeitos fixos quanto de efeitos aleatórios o resultado foi bastante próximo em magnitude e com sinal igual. O coeficiente estimado é compatível com o esperado, e com o encontrado pela literatura.

A variável Distância Geográfica revela que, quanto maior a distância entre as regiões, menor é a conexão entre duas regiões. Este resultado indica para o Brasil o mesmo efeito observado pela literatura empírica analisada (seção 2), de que o espaço ainda possui grande influência na formação de redes, descrito como um decaimento espacial da força das conexões.

Em relação à segunda variável de interesse (Conectividade) que representa a dimensão relacional, o coeficiente possui sinal positivo, revelando que regiões conectadas com diversas outras possui maior propensão a formar laços. Cassi e Plunket (2015) encontram resultado positivo para a variável "número de parceiros comuns" entre inventores franceses e para o resto da Europa. Breschi e Lissoni (2004) utilizam graus de

distância geodésica¹³, e encontram que nós mais próximos na rede são mais conectados. O modelo estimado por Ejermo e Karlsson (2006) inclui a dimensão relacional como variável dependente e encontram que o sinal para a variável independente "número de patentes" é positiva na Suécia. O trabalho de Crescenzi, Nathan e Rodríguez-Pose (2016) possui paralelo mais próximo à variável Rel_{ij} , e encontra valores positivos para a proximidade social no Reino Unido.

¹³ Menor caminho que une dois nós na rede.

Tabela 18. Resultado das Estimações do Painel Inter-regional, 2000-2011

Var, Dependente: Ligações	Binomial Negativo				Poisson				GLS			
	EF		EA		EF		EA		EF		EA	
Distância Geográfica	-0,1450 ***	(0,0313733)	-0,1443 ***	(0,0313563)	-0,1049 ***	(0,0160586)	-0,1056 ***	(0,0160508)	-0,1464 ***	(0,0219788)	-0,1486 ***	(0,0220026)
Conectividade	105,4943 ***	(5,356739)	105,1829 ***	(5,320424)	92,7087 ***	(2,804755)	92,9206 ***	(2,800962)	479,4715 ***	(7,117761)	489,2251 ***	(7,062084)
Proximidade Tecnológica	3,1950 ***	(0,2197315)	3,1967 ***	(0,2193103)	3,6475 ***	(0,1238855)	3,6467 ***	(0,1238483)	-0,4469 ***	(0,0862225)	-0,5153 ***	(0,0852995)
Diferença no PIB	-0,0661 ***	(0,019878)	-0,0663 ***	(0,0198578)	-0,0612 ***	(0,0101809)	-0,0616 ***	(0,0101803)	0,0797 ***	(0,0141535)	0,0583 ***	(0,0139889)
Diferença na Centralidade	-1,5485 ***	(0,3132983)	-1,5880 ***	(0,3131813)	-1,4649 ***	(0,1810845)	-1,4677 ***	(0,1807008)	2,6086 ***	(0,0944412)	2,6255 ***	(0,0944032)
Diferença no # de Pesquisadores	31,3742 ***	(6,425746)	31,4321 ***	(6,387841)	15,9128 ***	(3,201307)	16,2740 ***	(3,204013)	-138,0748 ***	(7,375614)	-117,9410 ***	(7,141655)
Média do PIB	0,1449 **	(0,0655643)	0,1195 **	(0,0580971)	0,3147 ***	(0,0482217)	0,2927 ***	(0,0486802)	-0,3448 ***	(0,0448345)	-0,0779 **	(0,0311064)
Média da Centralidade	19,1618 ***	(0,5439765)	19,0200 ***	(0,5428458)	20,1330 ***	(0,3411138)	20,0801 ***	(0,3411246)	-4,1747 ***	(0,1603532)	-4,1667 ***	(0,1584256)
Média do # de Pesquisadores	133,3466 ***	(10,37981)	137,9942 ***	(10,45459)	135,3745 ***	(5,408058)	136,9331 ***	(5,456409)	349,7835 ***	(12,96525)	282,1887 ***	(11,39876)
Mesma UF	0,7279 ***	(0,0579097)	0,7291 ***	(0,0578975)	0,8567 ***	(0,0288164)	0,8570 ***	(0,0288092)	0,6654 ***	(0,0678593)	0,6643 ***	(0,0679383)
Constante	-13,2323 ***	(0,7037904)	-12,9626 ***	(0,6356902)			-13,8671 ***	(0,5127596)	2,6861 ***	(0,3707915)	0,6693 **	(0,2743133)
Teste de Razão de MV			0,0920 *				0,0000 ***					

Erro padrão entre parênteses. Nível de significância: * 10%, ** 5%, *** 1%

Tabela 19. Resultado das Estimações Trienais Inter-regionais, Brasil, 2000-2011

Var. Dependente: Ligações	2000-2002		2003-2005		2006-2008		2009-2011	
Distância Geográfica	-0,227	***	-0,385	***	-0,199	***	-0,478	***
	(0,076)		(0,077)		(0,075)		(0,075)	
Conectividade	90,341	***	233,909	***	98,372	***	-19,327	
	(20,709)		(21,340)		(22,791)		(29,409)	
Distância Tecnológica	0,0385		1,697		2,529		1,381	***
	(0,478)		(0,503)		(0,504)		(0,4983)	
Diferença no PIB	-0,030		-0,066		0,049		0,069	
	(0,047)		(0,055)		(0,052)		(0,054)	
Diferença na Centralidade	1,085		-1,043		-0,908		-1,768	
	(1,190)		(0,825)		(1,051)		(1,338)	
Diferença no # de Pesquisadores	-35,937	*	7,589		-14,750		-104,601	***
	(18,785)		(23,356)		(22,387)		(24,895)	
Média do PIB	-0,357	*	0,0998		-0,447	**	-1,145	***
	(0,199)		(0,227)		(0,212)		(0,200)	
Média da Centralidade	40,104	***	19,903	***	31,819	***	51,709	***
	(2,790)		(1,751)		(2,744)		(3,610)	
Média do # de Pesquisadores	122,749	***	268,845	***	174,309	***	240,642	***
	(30,737)		(39,694)		(39,261)		(42,827)	
Mesma UF	1,507	***	1,263	***	1,043	***	1,536	***
	(0,173)		(0,184)		(0,180)		(0,182)	
Constante	-13,838	***	-11,147	***	-11,737	***	-9,778	***
	(1,997)		(2,261)		(2,295)		(2,240)	

Erro padrão entre parênteses. Nível de significância: * 10%, ** 5%, *** 1%

O coeficiente obtido para a dimensão de proximidade tecnológica mostra que regiões similares tecnologicamente tendem a ser mais propensas a se conectar. Neste caso, a afinidade cognitiva tende a prevalecer porque pode facilitar a comunicação e o aprendizado dos inventores. Tal efeito também é observado por Cassi e Plunket (2015) na França e Crescenzi, Nathan e Rodríguez-Pose (2016) para o Reino Unido.

Os resultados encontrados para as variáveis de interesse no Brasil estão condizentes com os encontrados pela literatura: (1) existe um decaimento do peso espacial conforme aumenta a distância entre os nós (HOEKMAN, FRENKEN e VAN OORT, 2009; HENDERSON, 2007); (2) regiões com mais conexões indiretas tendem a se conectar mais facilmente e formar *clusters* de rede (BATHELT, MALMBERG e MASKELL, 2004; EJERMO e KARLSSON, 2006); (3) a proximidade tecnológica possui uma influência significativa na conexão de duas regiões (CASSI e PLUNKET, 2015; CRESCENZI, NATHAN e RODRIGUEZ-POSE, 2016).

Os blocos (2) e (3), que analisam a diferença e a média das variáveis de cada nó devem ser analisadas conjuntamente, para que os resultados capturem a particularidade das relações. Para auxiliar a análise, a Tabela 20 constrói as estatísticas Gama, em que Gama 1 é o coeficiente estimado para o nó em dominância e Gama 2 para o nó em posição de dominado.

Tabela 20. Estimação dos coeficientes de dominância, ótica regional

Gama	Binomial Negativo		Poisson		GLS	
	EF	EA	EF	EA	EF	EA
Gama 1						
<i>PIB</i>	0,0064	-0,0065	0,0962	0,0847	-0,0927	0,0193
<i>Centralidade</i>	8,0324	7,9220	8,6016	8,5724	0,5212	0,5421
<i># de Pesquisadores</i>	98,0475	100,4292	83,6000	84,7405	36,8170	23,1534
Gama 2						
<i>PIB</i>	0,1386	0,1260	0,2186	0,2080	-0,2521	-0,0973
<i>Centralidade</i>	11,1293	11,0981	11,5315	11,5077	-4,6959	-4,7088
<i># de Pesquisadores</i>	35,2991	37,5650	51,7745	52,1926	312,9666	259,0354

Os resultados obtidos para o PIB indicam um sinal negativo na relação de diferença e um sinal positivo na relação de média. Isto indica que, quanto maior for a distância na riqueza de duas regiões, há menor propensão à conexão. Assim como quanto

mais elevada for a média de riqueza desses nós, mais propensos eles são. Por outro lado, analisando os Gamas, pode-se observar que enquanto Gama 2 possui sinal positivo, o que mostra que os nós menos favorecidos possuem um ganho em formar conexões com nós mais ricos, o sinal de Gama 1 não possui sinal constante, sendo positivo para a estimação de EF e negativo para EA.

Para a centralidade, o coeficiente de diferença estimado é negativo, enquanto o de média é positivo. Por estes valores, tem-se que os nós menos centrais costumam se relacionar com nós de centralidade mais próxima a sua. Além disso, quanto maior a média dos nós, mais conexão é esperada entre estes (Tabela 18). A análise dos Gamas revela que tanto o nó em hierarquia maior quanto aquele em hierarquia menor se beneficiam do ganho de centralidade para a formação de laços, embora os nós com valores mais baixos tenham um ganho marginal mais elevado. Este resultado pode ser interpretado como o canal pelo qual os nós menos centrais utilizam para se conectar ao resto da rede.

Os coeficientes estimados para o número de trabalhadores alocados em atividades de P&D revela um outro efeito. Para o coeficiente da diferença, é estimado que quanto maior a diferença no número de trabalhadores entre as regiões do laço, maiores são as conexões formadas pelo laço. Quanto mais trabalhadores em P&D existem no laço, maiores são as possibilidades de conexão. Com este resultado, o benefício do nó com mais pesquisadores no laço é maior, uma vez que se amplia suas possibilidades de conexão. Este resultado é refletido para Gama 1 > Gama 2, embora ambos sejam positivos. Este resultado é similar ao encontrado por Ejerme e Kalrsson (2006), que destacaram o fato de regiões com poucos trabalhadores em pesquisa tenderem a se conectar com regiões com maior capacidade de P&D, para fins de suprir deficiências regionais em P&D.

Os resultados obtidos para as estimações dos coeficientes do controle podem, então, ser resumidos em: (1a) quanto maior a riqueza do laço, mais conexões são esperadas formar; (1b) regiões com menor riqueza buscam as mais ricas para como meio de acessar fluxos mais ricos; (1c) não existem ganhos aparentes das regiões mais ricas buscarem conexões com as menos ricas; (2a) regiões menos centrais encontram nas mais centrais um caminho para a conexão na rede, embora haja uma penalidade no *spread* da centralidade; (2b) quanto maior a média da centralidade, mais conexões são esperadas; (3a) existe grande benefício no relacionamento de regiões com grande diferença na

quantidade de trabalhadores dedicados ao P&D; (3b) regiões com elevada capacidade de P&D possuem um elevado benefício marginal na relação com as regiões com menor capacidade.

Em uma análise mais ampla, pode-se extrair que a formação das redes inter-regionais de invenções no Brasil é determinada por dois tipos de efeitos. Primeiro, as relações se beneficiam do relacionamento com pares similares, em que o ganho de conexões se faz, sobretudo, pela elevada capacidade produtiva de invenções. Segundo, as relações também se beneficiam pela diferença, que permite às regiões mais periféricas, menos ricas e com menor capacidade de P&D de acessar os transbordamentos das regiões mais favorecidas e suprir uma carência interna.

Por fim, é feito o teste da Razão de Máxima-Verossimilhança, que compara os resultados obtidos na estimação de Efeitos Aleatórios com a estimação de um painel agrupado. O resultado obtido é significativo a 10%, o que indica que o modelo binominal negativo possui melhor ajustamento às propriedades estatísticas estimadas. Entre os modelos de Efeitos Fixos e Efeitos Aleatórios não há diferenças relevantes na magnitude ou significância dos coeficientes, o que não altera as conclusões.

Em complemento, a Tabela 19 exhibe as estimações trienais utilizando o modelo binomial negativo. Em comparação com a estimação do painel, os resultados obtidos nas estimações trienais parecem menos conclusivos. Embora algumas variáveis tiveram seu desempenho e sinal confirmados em comparação com o painel anterior, variáveis de adaptação mais lenta, como distância tecnológica ou variáveis de diferenças, em grande parte associadas à percepção de ganho esperado, não tiveram resultados significativos. Isto pode indicar um fator já esperado para a rede brasileira e observado na literatura internacional analisada: redes são construções sociais do tempo e não conjecturais. Ou seja, a relação intra-trienal possui um tempo de vida curto demais para que a intensidade dos laços seja detectada. Estas passam a ser mais um subproduto das relações individuais momentâneas do que das características regionais menos flexíveis no tempo. De fato, observando a Tabela 21, pode-se observar que existe alta correlação entre o número de ligações realizadas pelos laços em cada período. Assim, tem-se indícios de que as relações formadas são construídas e determinadas no tempo.

Tabela 21. Autocorrelação das ligações entre os períodos, Brasil, 2000-2011

	2000-2002	2003-2005	2006-2008	2009-2011
2000-2002	1,0000			
2003-2005	0,8828	1,0000		
2006-2008	0,8426	0,8716	1,0000	
2009-2011	0,7987	0,8253	0,8683	1,0000

Fonte: Elaboração Própria.

A estimação seguinte irá ajudar a elucidar os determinantes dos laços entre inventores. Embora a região tenha um efeito importante na construção de espaços inventivos e propícios à cooperação, o processo de tomada de decisão de cooperar ou não cabe ao inventor. Este seleciona parceiros com base em um conjunto de situações e variáveis observáveis ou não na tentativa de maximizar o retorno da cooperação. A análise conjunta das estimações regional e individual permitirá construir um arcabouço analítico de como as regiões importam na tomada de decisão individual.

5 DETERMINANTES DA REDE BRASILEIRA DE INVENTORES

Os resultados obtidos na primeira estimação, sobre redes inter-regionais, revelaram que a rede brasileira de inventores possui uma clara divisão entre as regiões do país. Tem-se uma influente região inventora no Centro-Sul do país formado por cidades centrais e sua polarização próxima. Por outro lado, regiões periféricas e interiores possuem menor autonomia inventiva devido as deficiências socioeconômicas destas localidades, que a tornam dependentes dos grandes centros regionais, como Manaus na região Norte e as capitais litorâneas no Nordeste. Portanto, entendida como se distribui a rede entre as regiões no país e a influência desta na formação de novas ligações, cabe compreender a motivação no nível do inventor para o copatenteamento.

5.1 MÉTODO DE ESTIMAÇÃO

A fim de se identificar quais os padrões de determinação da conexão na rede brasileira de inventores, é proposta a aplicação de um modelo de estimação adequado ao tratamento de eventos raros. Este tipo de modelo apresenta vantagens para lidar com conjuntos de dados que possuem dados truncados e inflados de zero (em que $y_{ij} = 0$ indica o evento não-observado). Ao se trabalhar com fenômenos sociais organizados em rede, a raridade de ocorrência de um evento dentre as possibilidades disponíveis é comum.

Portanto, deve-se compreender *ex ante* como identificar este fenômeno e construir um contrafactual comparativo. Este entendimento irá fornecer mecanismos úteis na proposição de um modelo adequado.

A probabilidade de conexão de um laço entre dois inventores está diretamente relacionada à possibilidade destes inventores possuírem algum contato prévio. Enquanto na rede formada entre regiões é razoável supor que, a nível agregado, é plausível que ao menos um inventor localizado em cada nó terminal do laço possa se conhecer e estabelecer uma conexão potencial, para o caso de inventores esta suposição é irreal, ou ao menos altamente improvável.

5.1.1 Construção do contrafactual

Para o período analisado de 2000-2011, o grafo exibe 7.733 nós distintos. Caso suposto que estes nós poderiam potencialmente se conectar a todos os outros demais nós, ter-se-ia 29.895.778 laços potenciais. Contudo, por mais amplo que seja a reputação de um inventor na rede, seria pouco provável supor que este disporia de tempo hábil para estabelecer relações formais, e estas gerarem um produto, com outros 7.732 indivíduos.

Diante deste fato, a construção de uma segunda rede que simule as conexões potenciais não-realizadas pelo inventor se faz necessária. Esta, contudo, possui desafios, limitações e algumas soluções possíveis. De maneira ideal, o verdadeiro contrafactual para cada inventor deveria ser formado por aqueles indivíduos a quem este já estabeleceu algum contato formal e profissional. Embora em redes com poucos nós esta informação pudesse ser facilmente obtida por meio de questionários padronizados, para redes com muitos nós, onde a privacidade do inventor deve ser preservada, este tipo de informação não pode ser obtido. Um método alternativo é a construção de um contrafactual que mimetize as relações potenciais.

Diferentes métodos de construção de contrafactual – ou laços de controle – são propostos na literatura de redes (KING e ZENG, 2001; SORENSON, RIVKIN e FLEMING, 2006; CASSI e PLUNKET, 2014; CHANDRASEKHAR, 2015). Em comum, a estratégia adotada por estes modelos é a seleção aleatória de casos-controle em que $g_{ij} = 0$.

Enquanto Cassi e Plunket (2014) optam por selecionar um número fixo de casos-controle não-conectados para a construção do pseudo-contrafactual, a abordagem adotada nesta dissertação adotará o Modelo de Ęrdos-Renyi (ERM) (ERDÖS e RENYI, 1959), conforme interpretado por Chandasekhar (2015). Este modelo possui três implicações que serão úteis na sua aplicação:

Dada que p é a probabilidade de $P(A_{ij} = 1)$ e que $i, j, k \in E$:

- (1) O número esperado de laços de um dado nó é $(n - 1)p$;
- (2) A probabilidade que dois vizinhos de i, j e k sejam ligados é p ;
- (3) A probabilidade de que i, j e k sejam mutuamente ligados é p^3 .

Estas implicações, destaca Chandasekhar (2015), são úteis para a realização de trabalhos empíricos. Primeiro, é esperado na rede observada que o número de laços possíveis de conexão é relativamente pequeno em comparação a n . Segundo, as redes empíricas são usualmente agrupadas em torno de poucos nós.

Para este trabalho foi construída um grafo aleatório para cada triênio, de modo a garantir que cada triênio receba uma quantidade proporcional aos nós nele presentes, com probabilidade de formação p igual a 0,05.¹⁴ A Tabela 22 dispõe a quantidade de laços de controle criados para cada triênio, bem como para a rede no geral. Ao agregar à rede de conexões entre inventores, estes laços receberam $g_{ij} = 0$, enquanto os demais laços formados foram ponderados segundo o número de conexões. Após a associação de ambas as redes (ERM) e *de facto*, foi obtida uma média de 2,20% de laços formados na rede.

A base de dados obtida pela combinação de laços formados *de facto* e formados pelo pseudo-contrafactual obtido pelo ERM possui indicações de ser um caso de evento raro. Segundo King e Zeng (2001), eventos raros podem ser vagamente definidos como os casos em que a quantidade de eventos ocorridos é muito inferior aos possíveis (ou não-eventos). Nas ciências sociais, um evento torna-se mais raro quanto maior é o número de agentes envolvidos no sistema. É razoável, para este caso, supor que a conectividade de um nó é dada antes pelo seu conhecimento prévio do par que pela inserção de um novo nó na rede. Assim, a raridade como a razão entre a quantidade de nós formados e a

¹⁴ O modelo foi construído utilizando o pacote *igraph* do software R.

quantidade de nós possíveis de serem formados. O Teorema 1 demonstra a existência de uma relação inversa entre esta razão e a quantidade de nós.

Tabela 22. Distribuição trienal dos laços de controle criados no ERM, Brasil, 2000-2011

	2000-2002	2003-2005	2006-2008	2009-2011	2000-2011
Laços Formados	2.764	1.999	4.766	5.150	14.679
Laços ERM	68.198	109.644	209.144	266.150	653.136
Laços Totais	70.962	111.643	213.910	271.300	667.815
% Laços Formados	3,90%	1,79%	2,23%	1,90%	2,20%

Fonte: Elaboração Própria.

Teorema 1. Se as conexões formadas por i independem de n , então a possibilidade de conexão dada pela razão entre laços formados e possíveis tende a zero conforme se amplia a quantidade de nós na rede.

Prova: Seja L o número de laços possíveis em um grafo não-direcional:

$$L = \frac{n(n-1)}{2} \quad (10)$$

Seja F_i o grau de i , dado pela probabilidade p de conexão com c nós conhecidos de i . Assuma que c é estritamente menor que e independente de n :

$$F_i = p \cdot c \quad (11)$$

Assuma ainda que p e c são iguais para todo nó na rede. São formados F laços:

$$F = \frac{1}{2} \sum_{i=1}^n F_i = \frac{npc}{2}$$

A razão F/L define a proporção de laços formados em relação ao total possível:

$$\frac{F}{L} = \frac{npc/2}{[n(n-1)]/2} = \frac{pc}{(n-1)} \quad (12)$$

Assim, se $n > 1$ (existe ao menos um par potencial):

$$\lim_{n \rightarrow \infty} \frac{F}{L} = 0$$

5.1.2 Estimador PPML para eventos raros

Para a rede trabalhada, a conectividade média, ou o percentual de laços formados em comparação ao potencial – já controlando para os níveis de razoabilidade pelo ERM – indica que no período de 2000-2011, apenas 2,20% dos laços se formaram, o que indica um caso de evento raro. King e Zeng (2001) apontam que a presença deste fenômeno estatístico implica na necessidade de modelos que corrijam a elevada proporção de valores zero. Santos Silva e Tenreyro (2011a) indicam o estimador Poisson de Pseudo-Máxima Verossimilhança (ou *Poisson Pseudo-Maximum Likelihood (PPML) estimator*) para tratar deste tipo de problema.

O estimador PPML é construído sob o modelo de regressão de Poisson, que especifica que y_i dado x_i possui uma distribuição Poisson com densidade (CAMERON e TRIVEDI, 2013)

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

e média $E[y_i|x_i] = \mu_i = \exp(x_i'\beta)$.

Este modelo também é conhecido como *log-linear*, uma vez que o logaritmo da média condicional é linear nos parâmetros. A função *log-probabilidade* do modelo é dada por

$$\ln L(\beta) = \sum_{i=1}^n \{y_i x_i' \beta - \exp(x_i' \beta) - \ln y_i!\}.$$

O coeficiente estimado $\hat{\beta}_p$ é a solução das condições de primeira ordem:

$$\sum_{i=1}^n (y_i - \exp(x_i' \beta)) x_i = 0.$$

Discute-se na literatura como este modelo responde a uma estrutura de dados de contagem dispersos, em que a variância condicional é maior que a esperança condicional. Embora Santos Silva e Tenreyro (2011a) e Fally (2015) tenham oferecido evidências de que os estimadores PPML são consistentes mesmo na presença deste problema, Cameron e Trivedi (2013) indicam que o uso de estimadores baseados em distribuição binomial negativa talvez seja desejável. Assim, frente a este debate, o modelo Binomial Negativo,

como descrito por Cameron e Trivedi (2013) será utilizado para fins comparativos de consistência.

Vale ressaltar ainda que a estimação por Pseudo-Máxima Verossimilhança do modelo de *Poisson*, conforme sugerido por Santos Silva e Tenreyro (2006, 2011a), possui uma limitação prática. A estimação dos coeficientes se dá por um mecanismo de iterações baseadas na estimação GML (*Generalized Linear Models*), o que o sujeita às limitações de convergência deste tipo de estimação. Santos Silva e Tenreyro (2010, 2011b) analisam os problemas de convergência dos modelos de *Poisson* e suas causas, e propõem uma solução baseada na realização de subamostra das variáveis com dificuldade de convergência. Contudo, no caso trabalhado, a realização de subamostras sem o controle de exclusão de observações *de facto* ou ERM pode implicar na eliminação de dados importantes no modelo, sendo preferível a exclusão das variáveis problemáticas.

5.2 VARIÁVEIS E MODELO EMPÍRICO

O modelo empírico proposto busca identificar variáveis que conseguem explicar a conexão formada entre dois inventores na rede brasileira de patentes. Para este fim, é proposto um modelo que compatibilize as características da rede, bem como indiquem a existência de hierarquia nos laços. Desta forma, será possível compreender melhor o processo de formação da percepção de ganhos do inventor na decisão de formar o laço.

Utilizando as variáveis descritas no Quadro 4, estas serão inseridas no modelo de quatro formas: (i) diferença absoluta, (ii) valor médio do laço, (iii) *dummy* de controle e (iv) em nível. As formas (i) e (ii) são definidos conforme seção 5.2.1. e serão aplicadas às variáveis de grau de instrução formal e renda do indivíduo. A forma (i) é também utilizada para a centralidade. A forma (iii) indica similaridade e dissimilaridade entre os nós que formam (ou podem formar) o laço, e é utilizada na inserção das variáveis que indicam laços intrassetoriais, laços intramunicipais, inventores de mesmo gênero, laços formados entre acadêmicos e empresas, e para indicar o triênio do laço. Por fim, a variável que indica a distância geográfica do laço é posta em nível.

O modelo empírico estimado fornecerá, portanto, não apenas os determinantes da formação de laços, mas a também indicação de posição no laço pelo coeficiente de dominância calculado posteriormente. O Quadro 5 indica o nome das variáveis utilizados no modelo, suas descrições e os sinais esperados.

Quadro 5. Descrição das variáveis do modelo sob a ótica do inventor

Variáveis	Sinal Esperado	Descrição
CONT		Número de conexões no laço
DIST	(-)	Distância geográfica em km
DCCENT	(-)	Diferença na Centralidade de Proximidade
DEDUC	(-)	Diferença na Educação em anos
DREND	(-)	Diferença na Renda em salários mínimos
RENDMED	(+)	Renda Média do Laço em salários mínimos
EDUCMED	(+)	Educação Média do Laço em anos
SSEXO	(+)	Mesmo Sexo (= 1)
SMUN	(+)	Mesmo Município (= 1)
SCNAE	(+)	Mesma divisão industrial (= 1)
LUNIVEMP	(?)	Conexão Universidade-Empresa
T1	(?)	2000-2002 (= 1)
T2	(?)	2003-2005 (= 1)
T3	(?)	2006-2009 (= 1)

5.3 RESULTADOS E INTERPRETAÇÕES

Os resultados encontrados para a estimação por máxima verossimilhança dos estimadores PPML e Binomial Negativo (NBRM) estão na Tabela 23. Em seguida serão discutidos os resultados gerais e específicos dos modelos estimados.

Estimaram-se inicialmente quatro modelos: duas versões de BNMR e duas versões de PPML. Na Tabela 23, os dois primeiros modelos indicam os modelos sem a inclusão de *dummies* trienais, enquanto os dois últimos incluem este controle, excluindo o último triênio para referência. Os resultados obtidos na estimação dos modelos de determinação espacial serão discutidos segundo sua classificação (Quadro 3). No quadro geral, tanto o modelo BNMR quanto o PPML produziram resultados similares, sem variações de sinal e com magnitude próxima.

Tabela 23. Resultado das estimações PPML e BNRM (y: CONT), Brasil, 2000-2011

Variáveis	BNRM	PPML	BNMR ¹	PPML ¹
DIST	-0,0009*** (2,97e-05)	-0,0009*** (4,80e-05)	-0,000919*** (2,97e-05)	-0,0009*** (4,77e-05)
DCCENT	-2,194*** (0,0415)	-2,195*** (0,0579)	-2,225*** (0,0424)	-2,215*** (0,0603)
DEDUC	-0,143*** (0,00506)	-0,140*** (0,0066)	-0,138*** (0,00512)	-0,137*** (0,0064)
DREND	-0,0118*** (0,000795)	-0,0119*** (0,0012)	-0,0130*** (0,000812)	-0,0133*** (0,0012)
RENDMED	-0,0087*** (0,000809)	-0,0079** (0,00399)	-0,00655*** (0,000861)	-0,00531 (0,0041)
EDUCMED	0,0232*** (0,0058)	0,0195** (0,00774)	0,0267*** (0,00610)	0,0192** (0,0086)
SSEXO	0,199*** (0,0162)	0,168*** (0,0172)	0,181*** (0,0164)	0,149*** (0,0172)
SMUN	2,263*** (0,0230)	2,255*** (0,0289)	2,287*** (0,0230)	2,275*** (0,0287)
SCNAE	1,601*** (0,0192)	1,607*** (0,0258)	1,603*** (0,0192)	1,606*** (0,0252)
LUNIVEMP	0,479*** (0,0185)	0,484*** (0,0209)	0,513*** (0,0188)	0,521*** (0,0218)
Constante	-4,394*** (0,0995)	-4,334*** (0,114)	-4,527*** (0,106)	-4,386*** (0,117)
Observações	667,815	667,815	667,815	667,815
R-2		0,262		0,264

¹ Controlado por *dummies* trienais, Exclui-se último triênio.

Os valores entre parênteses indicam o erro padrão do coeficiente.

* = significativo a 1%, ** = significativo a 5%, *** = significativo a 10%.

A variável geográfica DIST, que indica a distância em quilômetros entre as cidades do indivíduo *i* e do indivíduo *j* possui, conforme esperado, sinal negativo e significativo. Análogo à análise regional realizada no primeiro ensaio, este resultado reflete o decaimento espacial esperado para a formação dos laços. Contudo, enquanto a literatura empírica aponta o efeito do decaimento espacial na difusão de informação (JAFFE, 1989; AUDRETSCH e FELDMAN, 1996; HENDERSON, 2007), este impõe também um limite às capacidades de conexão formal entre indivíduos (TER WAL, 2014). A concentração das conexões da rede nas cidades é marcada pela *dummy* SMUN, que possui valor positivo e significativo.

Sob a ótica do desenvolvimento regional, a confirmação da existência (ou continuidade) do decaimento espacial entre inventores é um indício de que a existência de regiões propícias à criação, à difusão e à absorção de conhecimento – os distritos

marshallianos – é fomentada por redes proximais, ou, na terminologia de Saxenian (1994) e Granovetter (1985), pelas estruturas sociais. Por outro lado, a formação de distritos socialmente confinados pode ter um efeito de contenção do fluxo informacional que propiciaria a formação de novos distritos ou integração distrital.

A variável de mensuração relacional (DCENT) mensura a diferença absoluta na centralidade de proximidade entre os nós do laço. Se os indivíduos possuem uma facilidade de conexão na rede que é similar, então pouco incentivo estes terão à conexão. O valor negativo e significativo indica este fenômeno. Interpreta-se que o indivíduo nota a capacidade de outro em se conectar na rede com facilidade para julgar a capacidade de auferir ganhos de futuras conexões. Por outro lado, indivíduos com alta reputação teriam baixo incentivos a se conectar com outros com centralidade muito inferior à sua. Por outro lado, indivíduos com baixa centralidade não encontram (com facilidade) meios de acessar indivíduos mais importantes, e recorrem àqueles que possuem centralidade próxima a sua.

Os motivos intrínsecos são indicados pelas variáveis indicativas de renda e educação. Conforme demonstrado na seção 5.2.1, a inclusão desta variável em duas dimensões permite separar os efeitos que estas exercem na formação de laços. Com exceção do coeficiente da renda média calculado na estimação PPML controlada por triênios, todas as demais foram significativas. Observando-se as duas primeiras estimações, a Tabela 24 indica a distribuição dos efeitos. Chamam-se γ_1 e γ_2 os coeficientes de dominância da educação, e ω_1 e ω_2 os coeficientes de dominância da renda, sendo β são os coeficientes da diferença e α os coeficientes da média

Baseando-se nos valores calculados, foi calculado tanto para a educação quanto para a renda um coeficiente de máximo (ou dominância) negativo e um coeficiente de mínimo (ou de dominado) positivo. Este resultado indica que na relação entre dois inventores em que educação e renda são diferentes, o indivíduo desfavorável possui maior propensão a formar conexões, enquanto a propensão do indivíduo com maior *status* é menor. Sob a ótica da percepção do indivíduo acerca dos retornos da relação, os valores calculados para estes coeficientes indicam que os indivíduos tendem a buscar aqueles que possuem nível educacional ou nível de renda maior que o seu, pois esperam que, desta forma, o retorno auferido seja maior.

Já os valores obtidos para os coeficientes de diferença e de média são diferentes para renda e educação. Enquanto no caso da educação tem-se o esperado, $\beta < 0$ e $\alpha > 0$, para a renda o coeficiente α teve sinal não previsto e significativo. Este resultado mostra que, quanto maior a média de rendimento do laço, menor é a propensão do laço ser formado. Embora contraintuitivo, a análise *ad hoc* possibilita uma explicação. Indivíduos com renda elevada usualmente atuam em altas posições hierárquicas nas empresas inventoras. Estes indivíduos se ocupam primordialmente de tarefas gerenciais e estratégicas, sendo pouco influentes na posição de invenção. Portanto, quanto maior o rendimento do laço, mais provável é que nesta relação hajam indivíduos com pouca propensão inventiva, o que justifica o sinal do coeficiente.

Tabela 24. Cálculo de Coeficientes de Dominação

	$\beta > 0$	$\beta < 0$
$\alpha > 0$	-	$\gamma_{1BNMR} = -0,1314$ $\gamma_{1PPML} = -0,1303$ $\gamma_{2BNMR} = 0,1546$ $\gamma_{2PPML} = 0,14975$
$\alpha < 0$	-	$\omega_{1BNMR} = -0,0162$ $\omega_{1PPML} = -0,0159$ $\omega_{2BNMR} = 0,0118$ $\omega_{2PPML} = -0,0119$

Em relação às variáveis de motivação extrínseca, tem-se um valor positivo e significativo para o coeficiente da *dummy* que indica pares formados entre indivíduos do mesmo sexo. Este coeficiente é positivo e significativo, o que indica que a maior propensão à conexão ocorre entre indivíduos do mesmo sexo, o que é consonante com as características da rede descritas em 5.1. Sob a ótica setorial, existe forte propensão à conexão entre indivíduos na área industrial. O mesmo ocorre, com menor intensidade, com as ligações entre universidade e empresas, que torna mais propensa a conexão no laço.

As *dummies* indicativas de triênio na terceira e quarta estimação revelam que a propensão a formar conexões nos primeiros dois triênios (2000-2005) foi menor em comparação com último triênio (2009-2011). No terceiro triênio (2006-2008), em

comparação com o último, foi sutilmente mais propenso às conexões. Por este fato, entende-se que houve uma melhora na percepção dos ganhos da rede ao longo do período, sobretudo após 2006.

A Tabela 25 exibe os resultados obtidos para as regressões trienais. Uma modificação foi necessária na estimação deste modelo. Seguindo Silva Santos e Tenreiro (2010, 2011b), a convergência dos modelos pôde ser obtida apenas com a exclusão de observações que criavam dificuldades de convergência à estimação por pseudo-máxima verossimilhança. Portanto, a fim de se preservar a estrutura dos dados e após a identificação da variável com restrições, optou-se pela exclusão da variável DCENT, que indica a diferença na centralidade do nó. Se, por um lado, a exclusão desta variável elimina do modelo a capacidade de identificar a posição relativa do nó na rede, a preservação desta variável mediante exclusão de observações poderia acarretar em viés de seleção.

A análise dos valores obtidos para a estimação do PPML trienal mostra que, com exceção da distância (sinal esperado), renda média no laço (sinal contrário), da *dummy* de mesmo sexo no primeiro triênio (não significativo e negativo) e no último (negativo), e da ligação entre universidade e empresa no terceiro triênio (negativa), todas as demais variáveis foram significativas.

Se comparada à estimação para todo o período (2000-2011), os resultados obtidos possuem interpretação próxima: (i) existe decaimento espacial e confinamento municipal, em acordo com as estimações anteriores; (ii) os valores obtidos para a educação são compatíveis com o resultado anteriormente encontrado e possuem a mesma interpretação; (iii) o resultado encontrado para a renda difere nos dois primeiros períodos – sendo não-significativo e positivo no primeiro triênio e positivo e significativo no segundo, mantendo-se a interpretação para os dois últimos triênios. Este resultado pode indicar que o papel hierárquico na firma recebe maior peso após 2006; (iv) a educação possui sinais e valores compatíveis com a estimação para todo o triênio; (v) a variável mesmo sexo, por sua vez, possui um comportamento mais errático que o esperado: sendo não significativo e negativo no primeiro período e negativo e significativo no último. Esta característica reflete as características descritas na seção 5.2 e os encontrados na literatura internacional de que o papel feminino ganha espaço na rede. É cabível ressaltar, porém, que o valor negativo na variável não indica participação feminina superior ou igualitária

à masculina, mas que a cooperação entre estes sexos possui maior benefício à formação de novas conexões; (vi) mantém-se a maior propensão em colaborar dentro da mesma área industrial; e, por fim, (vii) com exceção do terceiro triênio, a conexão entre universidade e empresa aumenta a propensão a conectar.

Tabela 25. Resultado das estimações trienais por PPML, ótica individual, 2000-2011

Variáveis	2000-2002	2003-2005	2006-2008	2009-2011
DIST	-0,000133 (0,000124)	-0,000465*** (0,000114)	-0,00178*** (0,000161)	-0,000948*** (7,51e-05)
DCCENT	<i>Variável omitida para convergência do modelo</i>			
DEDUC	-0,0978*** (0,0223)	-0,127*** (0,0207)	-0,170*** (0,0167)	-0,0893*** (0,00824)
DREND	-0,0115** (0,00461)	-0,00741*** (0,00176)	-0,0160*** (0,00209)	-0,0148*** (0,00191)
RENDMED	0,0266 (0,0189)	0,0159** (0,00706)	-0,0215*** (0,00296)	-0,0197*** (0,00283)
EDUCMED	-0,150*** (0,0502)	-0,123*** (0,0304)	0,190*** (0,0264)	-0,0533*** (0,00838)
SSEXO	-0,0735 (0,0578)	0,361*** (0,0601)	0,748*** (0,0388)	-0,384*** (0,0251)
SMUN	2,675*** (0,123)	2,624*** (0,0941)	1,813*** (0,0528)	2,973*** (0,0567)
SCNAE	2,648*** (0,105)	2,199*** (0,0931)	0,800*** (0,0351)	2,039*** (0,0624)
LUNIVEMP	1,152*** (0,125)	0,523*** (0,108)	-0,786*** (0,0483)	0,368*** (0,0723)
Constante	-3,638*** (0,444)	-3,810*** (0,460)	-6,168*** (0,409)	-3,545*** (0,136)
Observações	70.962	111.643	213.910	271.300
R-2	0,248	0,124	0,119	0,484
γ_1 (EDUC)	-0,1728	-0,1885	-0,075	-0,11595
γ_2 (EDUC)	0,0228	0,0655	0,265	0,06265
ω_1 (REND)	0,0018	0,00054	-0,02675	-0,02465
ω_2 (REND)	0,0248	0,01536	0,00525	0,00495

Os valores entre parênteses indicam o erro padrão do coeficiente.

* = significativo a 1%, ** = significativo a 5%, *** = significativo a 10%.

5.4 INTERPRETAÇÕES GERAIS, LIMITAÇÕES E CONCLUSÕES

Os resultados empíricos obtidos pelo modelo que identifica os determinantes da rede brasileira de inventores corroboram a nível micro a ótica regional (Cap. 4). Assim como os resultados obtidos estão a par dos encontrados pela literatura internacional e pelos efeitos esperados pela teoria da inovação e pela teoria de redes.

Entende-se, portanto, que os motivos que levam os inventores a formar novas conexões ou preservar antigas conexões, fundamentam-se, principalmente, na sua percepção de retorno da conexão. Embora a medida de retorno à conexão não seja precisa, a interpretação do laço formado em patentes permite estimar aquelas conexões que geraram algum valor econômico. Se, por um lado, os retornos sociais, como a reputação, não são capturados por esta metodologia, por outro lado pode-se supor que este é altamente relacionado ao retorno econômico, ao menos em nível agregado.

O estudo sob a ótica do indivíduo identificou algumas relações importantes que estes levam em consideração na decisão de conexão. Deve-se ter em conta, porém, que a decisão do laço não é unilateral, mas requer que ambos nós possuam conhecimento prévio das intenções e formem suas percepções de retorno. Ademais, questões como assimetria informacional, racionalidade limitada e características psicológicas não podem ser acessadas nas bases de dados comumente adotadas, uma vez que requerem estudos específicos baseados em questionários individuais para a formação de perfis de inventores.

Uma segunda limitação deste modelo é considerar apenas a rede formada por inventores formalmente empregados, imposta pela utilização da RAIS como base de microdados complementar. Se, por um lado, existe esta limitação, por outro lado, a utilização da RAIS permite acessar dados em níveis individual e comumente sigilosos, como rendimento e residência do indivíduo. Qualquer outro método de obtenção destes dados seria dispendioso e ineficiente.

Desta forma, o resultado obtido para na segunda estimação proposto atende seus objetivos propostos de identificar variáveis sensíveis ao processo de copatenteamento entre inventores, e os efeitos que estas variáveis causam na formação das conexões. A

interpretação conjunta dos dois níveis investigados, regional e individual, possibilita a construção de um *framework* para análise de redes de inventores no Brasil e pontos de atenção na formulação de mecanismo de incentivo à integração tecnológica nacional.

CONCLUSÕES

Nesta última seção os resultados obtidos serão analisados à luz da pergunta inicialmente levantada no estudo. Desta forma, será possível indicar caminhos mais concretos para a compreensão do problema identificado pela dissertação bem como identificar futuros estudos a se realizar.

O objetivo geral da pesquisa realizado era responder a seguinte questão: quais os determinantes da formação de laços entre inventores no Brasil? Este foi dividido em dois objetivos específicos. Primeiro, buscou-se compreender os fatores que criam condições de cooperação inventiva entre as regiões brasileiras, utilizando o conceito de Região de Influência do IBGE. Em sequência, o estudo focou na identificação das características dos inventores que propiciam a construção de redes de invenção.

A relevância na realização deste estudo é permitir maior compreensão do processo que leva à formação de redes que permitem a difusão de informações para “fora” dos centros tradicionalmente inventores. O papel dos laços sociais, no sentido de Saxenian (1994), para a consolidação de uma organização em rede (CASTELLS, 1996), pode ter padrões identificados que possibilitam elaborar planos mais amplos de cooperação inventiva entre regiões. Uma vez que a inovação, suas causas e consequências, constitui elemento decisivo para o desenvolvimento de uma região, a integração de áreas mais remotas com centros inventivos pode trazer benefícios à rede como um todo.

Com relação à estimação do capítulo 4, de foco regional, alguns resultados valem ser destacados. Primeiro, o espaço geográfico ainda é bastante importante para a cooperação entre as regiões inventoras, principalmente para regiões periféricas da rede. Isto indica que a prevista “morte do espaço” ainda não é uma realidade. Segundo, as regiões inventoras centrais possuem um papel-chave na rede, atuando como *hub* social e tecnológico para regiões que dependem diretamente destas. Esta influência regional cria *polos inventivos*, que concentram grande parte dos subcomponentes de invenção brasileiros. Em paralelo com desenvolvimento econômico histórico das regiões brasileiras, tem-se um padrão bem conhecido: regiões mais desenvolvidas se localizam

no entorno dos polos inventivos enquanto regiões menos desenvolvidas estão na periferia da rede.

De modo análogo, a estimação do capítulo 5 buscou identificar o que leva os indivíduos a se conectar mais ou menos à rede. Partindo do conceito de percepção de retornos da relação, o estudo focou em encontrar um conjunto de variáveis que refletisse como um inventor enxerga os ganhos sociais, incluindo noções socioeconômicas. Assim, foram identificados certos padrões de atratividade entre indivíduos na rede. Em média, os inventores buscam se conectar com indivíduos do mesmo sexo, da mesma área de atuação, com grau de instrução e rendimentos similares, que estão geograficamente próximos e que possuem relevância similar na rede.

Os resultados obtidos para ambas estimações evidenciam para o Brasil características similares ao encontrado nas experiências internacionais, embora com sutis diferenças. Este resultado pode ser apresentado sob duas óticas: geográfica e relacional. O primeiro trata dos aspectos do espaço físico, a nível regional e individual, e como este influencia na rede, já o segundo nas características internas da rede que afetam a forma como os laços se formam.

Sob a ótica da geografia, o espaço possui um papel fundamental na construção de novos laços, sobretudo para os nós que possuem poucas conexões. Conforme o nó se torna mais relevante na rede, em termos de conexões e centralidade, as afinidades sócio-cognitivas parecem desempenhar um papel mais importante que a distância dos nós. Além disso, os resultados individuais indicam algum efeito de contenção de fronteira, levando os indivíduos a se conectar mais dentro da mesma cidade ou do mesma unidade federativa. Esta característica, quando extrapolada para o agregado, se traduz na formação dos polos inventivos.

Sob a ótica relacional, os resultados ratificaram o esperado pela teoria e o obtido na literatura internacional. Regiões centrais atuam na rede de modo a suprir carências regionais, sobretudo tecnológicas e humanas, bem como são intermediária nos fluxos relacionais. Já indivíduos centrais – comumente denominados “inventores estrelas” – possuem não apenas um papel inventivo como também conectivo, uma vez que acumulam capital social e atuam como referência na rede.

Embora haja paralelos entre os resultados obtidos na primeira e na segunda estimação (capítulos 4 e 5, respectivamente), os motivos que levam o laço a se formar no

primeiro difere do segundo em alguns aspectos. Enquanto sob a ótica regional os laços formados são determinados, principalmente, pela capacidade do espaço em fornecer condições de cooperação inventiva, sob a ótica do inventor o prestígio – e a percepção deste – é um fator mais destacado. Em nível macro, este tipo de formação de laço se reflete no acúmulo de inventores prestigiados e centrais em localidades que oferecem condições de invenção.

Metodologicamente, o trabalho buscou desenvolver uma forma de estimar variáveis quantitativas individuais para o laço. Para isto, introduziu-se um coeficiente de dominância que permite calcular separadamente os coeficientes para o laço e para o nó, quanto existe uma tendência à hierarquização não-observada no laço. Esta ferramenta se mostrou útil quando aplicada a redes bilaterais sem hierarquia definida e pode ser utilizada em aplicações futuras de redes com características e limitações similares.

Algumas limitações na execução do trabalho devem ser levantadas. Primeiro, o uso da RAIS como base de dados de informações individuais é restritiva no sentido em que só captura indivíduos com registro de emprego formal no ano de coleta. Duas situações refletem esta restrição: caso o inventor não tenha registro de emprego formal (inventor liberal) ou caso o tenha participado do processo inventivo, registrado como inventor, mas não esteja na RAIS no ano de depósito da patente. Neste caso, não se terá as informações para o inventor e os dados. Segundo, a RAIS possui dados faltantes, o que não permite o uso de todos os inventores identificados caso a informação seja imprescindível. Terceiro, o período utilizado de 12 anos (ou quatro triênios), 2000-2011, não captura grandes mudanças estruturais. Desta maneira, variáveis sensíveis às mudanças de longo prazo não possuem grande variabilidade em períodos mais curtos, o que impossibilita avaliar efeitos de longo prazo na determinação de relações. Quarto, tem-se como premissa de que o contato social ocorre apenas no momento de registro da invenção, não sendo avaliados efeitos por contatos prévios, como trabalho ou universidade.

O estudo das redes brasileiras de invenção ainda são iniciais e possuem uma ampla gama de questionamentos ainda a explorar. A partir da identificação da formação da rede feita neste trabalho, é necessário compreender quais os efeitos da cooperação no desempenho econômico da região. É possível ainda estudar a evolução e alcance da rede, bem como a sobrevivência das conexões. Já sob uma ótica normativa, pode-se explorar políticas de incentivo à rede. Além disso, a especificidade estrutural e institucional

brasileira possibilita o estudo dos efeitos da topologia de rede brasileira em comparação com sistemas de inovação já bastante institucionalmente desenvolvidos.

REFERÊNCIAS BIBLIOGRÁFICAS

ADAMS, J. D., BLACK, G. C., CLEMMONS, J. R., STEPHAN, P. E. Scientific teams and institutional collaborations: Evidence from US universities, 1981–1999. **Research Policy**, v. 34, n. 3, p. 259-285, 2005.

ALBUQUERQUE, Eduardo da Motta. Sistema nacional de inovação no Brasil: uma análise introdutória a partir de dados disponíveis sobre a ciência e a tecnologia. **Revista de Economia Política**, v. 16, n. 3, p. 56-72, 1996.

ALBUQUERQUE, Eduardo da Motta. Domestic patents and developing countries: arguments for their study and data from Brazil (1980–1995). **Research Policy**, v. 29, n. 9, p. 1047- 1060, 2000.

ALBUQUERQUE, E., SIMÕES, R., BAESSA, A., CAMPOLINA, B., & SILVA, L.. A distribuição espacial da produção científica e tecnológica brasileira: uma descrição de estatísticas de produção local de patentes e artigos científicos. **Revista Brasileira de Inovação**, v. 1, n. 2, p. 225-251, 2002.

ALMEIDA, Paul; KOGUT, Bruce. Localization of knowledge and the mobility of engineers in regional networks. **Management science**, v. 45, n. 7, p. 905-917, 1999.

ANCORI, Bernard; BURETH, Antoine; COHENDET, Patrick. The Economics of Knowledge: The Debate about Codification and Tacit Knowledge. **Industrial and Corporate Change**, vol. 9,n. 2, 2000.

ARAÚJO, Bruno César; CAVALCANTE, Luiz Ricardo; ALVES, Patrick. Variáveis proxy para os gastos empresariais em inovação com base no pessoal ocupado técnico-científico disponível na Relação Anual de Informações Sociais (RAIS). **Radar: Tecnologia, Produção e Comércio Exterior**: n. 05, 2009.

AUDRETSCH, David B.; FELDMAN, Maryann P. R&D spillovers and the geography of innovation and production. **The American Economic Review**, p. 630 - 640, 1996.

BATHELT, Harald; MALMBERG, Anders; MASKELL, Peter. 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.

BECATTINI, Giacomo. **Les districts industriels en Italie. La flexibilité en Italie**, Paris, Syros, p. 261-270, 1989.

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

BRESCHI, Stefano; LISSONI, Francesco. Knowledge networks from patent data: Methodological issues and research targets. **KITeS**, Centre for Knowledge, Internationalization and Technology Studies, Università Bocconi, Milano, Italy, 2004.

BRESCHI, Stefano; LISSONI, Francesco. "Cross-Firm" Inventors and Social Networks: Localized Knowledge Spillovers Revisited. **Annales d'Economie et de Statistique**, p. 189-209, 2005.

BUSCHMANN, Tobias; PYKA, Andreas. The evolution of innovation networks: the case of a publicly funded German automotive network, **Economics of Innovation and New Technology**, 24:1-2, 114-139, 2015.

CAMERON, A. Colin; TRIVEDI, Pravin K. **Regression analysis of count data**. Cambridge university press, 2013.

CASSI, Lorenzo; PLUNKET, Anne. Proximity, network formation and inventive performance: in search of the proximity paradox. **The Annals of Regional Science**, v. 53, n. 2, p. 395-422, 2014.

CASSI, Lorenzo; PLUNKET, Anne. Research Collaboration in Co-inventor Networks: Combining Closure, Bridging and Proximities. **Regional Studies**, vol. 49, n. 6, p. 936-954, 2015.

CASTELLS, Manuel. **The Rise of the Network Society: The Information Age: Economy, Society, and Culture, Volume I**. Oxford: Blackwell Publishers. 1996.

CHANDRASEKHAR, A. **Econometrics of network formation**. Oxford Handbook on the Econometrics of Networks, forthcoming, v. 12, 2015.

CHRISTALLER, Walter. **Die zentralen Orte in Süddeutschland: eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen**. University Microfilms, 1933.

CHRISTALLER, Walter. **Central places in southern Germany**. Prentice-Hall, 1966.

COHEN, Wesley M.; LEVINTHAL, Daniel A. Innovation and learning: the two faces of R & D. **The economic journal**, v. 99, n. 397, p. 569-596, 1989.

COLEMAN, James; KATZ, Elihu; MENZEL, Herbert. **Medical innovation: A diffusion study**. Indianapolis: The Bobbs-Merrill Company, 1966.

COWAN, Robin. Network models of innovation and knowledge diffusion. **Clusters, networks and innovation**, p. 29-53, 2005.

COWAN, Robin; JONARD, Nicolas. Network structure and the diffusion of knowledge. **Journal of Economic Dynamics & Control**, vol. 28, p. 1557 – 1575, 2004.

CRESCENZI, R.; NATHAN, M.; RODRÍGUEZ-POSE, A. Do Inventors Talk to Strangers? On Proximity and Collaborative Knowledge Creation. **Research Policy**, vol. 45, n. 1, 177-194, 2016.

DACHS, Bernhard; EBERSBERGER, Bernd; LÖÖF, Hans. The innovative performance of foreign-owned enterprises in small open economies. **The Journal of Technology Transfer**, v. 33, n. 4, p. 393-406, 2008.

DE FARIA, Pedro; LIMA, Francisco; SANTOS, Rui. Cooperation in innovation activities: The importance of partners. **Research Policy**, v. 39, n. 8, p. 1082-1092, 2010.

DECI, E.L. **Intrinsic motivation**. Nova York: Plenum Publishing Co, 1975

EJERMO, Olof; KARLSSON, Charlie. Interregional inventor networks as studied by patent coinventorships. **Research Policy**, Elsevier, vol. 35, n. 3, p. 412-430, 2006.

ier, vol. 35, n. 3, p. 412-430, 2006.

ERDÖS, Paul; RÉNYI, Alfréd. On random graphs, I. **Publicationes Mathematicae** (Debrecen), v. 6, p. 290-297, 1959.

FALLY, Thibault. Structural gravity and fixed effects. **Journal of International Economics**, v. 97, n. 1, p. 76-85, 2015.

FELDMAN, M. P.; KOGLER, D. F. **Stylized facts in the geography of innovation**. In: Hall; B. H.; Rosenberg, N. (eds.) *Handbook of Economics of Innovation*. Elsevier, 2010, cap. 8.

FITJAR, Rune Dahl; RODRÍGUEZ-POSE, Andrés. Nothing is in the air. **Growth and Change**, 2016.

FRENKEN, K., HOEKMAN, J., KOK, S., PONDS, R., VAN OORT, F., & VAN VLIET, J.. Death of Distance in Science? A Gravity Approach to Research Collaboration. **Innovation Networks**, Understanding Complex Systems, p. 43-57, 2009.

GABHER, Gernot; IBERT, Oliver. Bad company? The ambiguity of personal knowledge networks. **Journal of Economic Geography**, vol. 6, p. 251 – 271, 2006.

GAUGHAN, Monica; BOZEMAN, Barry. Using the prisms of gender and rank to interpret research collaboration power dynamics. **Social Studies of Science**, v. 46, n. 4, p. 536-558, 2016.

GIURI, Paola; MARIANI, Myriam. When distance disappears: Inventors, education, and the locus of knowledge spillovers. **Review of Economics and Statistics**, v. 95, n. 2, p. 449-463, 2013.

GLAESER E. L., KALLAL H. D., SCHEINKMAN J. A., SHLEIFER, A. Growth in cities. **Journal of Political Economy**, vol. 100, p. 1126-1152, 1992.

GONÇALVES, Eduardo. O padrão espacial da atividade inovadora brasileira: uma análise exploratória. **Estudos Econômicos (São Paulo)**, v. 37, n. 2, p. 405-433, 2007.

GONÇALVES, Eduardo; ALMEIDA, Eduardo. Innovation and spatial knowledge spillovers: evidence from Brazilian patent data. **Regional Studies**, v. 43, n. 4, p. 513-528, 2009.

GONÇALVES, Eduardo; FAJARDO, Bernardo de Abreu Guelber. A influência da proximidade tecnológica e geográfica sobre a inovação regional no Brasil. **Revista Econômica Contemporânea**, v. 15, n. 1, p. 112-142, 2011.

GRANOVETTER, Mark S. The strength of weak ties. **American journal of sociology**, p. 1360-1380, 1973.

GRANOVETTER, Mark. Economic action and social structure: The problem of embeddedness. **American journal of sociology**, p. 481-510, 1985.

GRILICHES, Zvi. Issues in assessing the contribution of research and development to productivity growth. **The Bell Journal of Economics**, p. 92-116, 1979.

GRILICHES, Zvi. The search for R&D spillovers. **Scandinavian Journal of Economics**, Wiley Blackwell, vol. 94, n. 0, p. S29-47, 1992.

GUSSO, D. **Agentes da inovação: quem os forma, quem os emprega. Tecnologia, exportação e emprego**. Brasília: Ipea, p. 397-444, 2006.

HENDERSON, J. Vernon. Understanding knowledge spillovers. **Regional Science and Urban Economics**, vol. 37, n. 4, p. 497-508, 2007.

HICKS, Diana M.; KATZ, J. Sylvan. Where is science going?. **Science, Technology, & Human Values**, v. 21, n. 4, p. 379-406, 1996.

HILBE, Joseph M. **Negative binomial regression**. Cambridge University Press, 2011.

HOEKMAN, Jarno; FRENKEN, Koen; VAN OORT, Frank. Collaboration networks as carriers of knowledge spillovers. **KITeS Working Papers** 222, 2008.

HOEKMAN, Jarno; FRENKEN, Koen; VAN OORT, Frank. The geography of collaborative knowledge production in Europe. **The Annals of Regional Science**, vol. 43, n. 3, p. 721- 738, 2009.

IBGE - Instituto Brasileiro de Geografia e Estatística. Divisão **Urbano-Regional**. Rio de Janeiro, 2013.

ISARD, Walter. **Location and space-economy: A General Theory Relating to Industrial Location, Market Areas, Land Use, Trade, and Urban Structure**. The MIT Press, 1956.

JACKSON, Matthew. **Social and Economic Networks**. Princeton University Press, 2008.

JAFFE, Adam B. **Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value**. 1986.

JAFFE, Adam B. Real effects of academic research. **The American Economic Review**, p. 957- 970, 1989.

JAFFE, Adam; TRAJTENBERG, Manuel. HENDERSON, Rebecca M.; Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. **Quarterly Journal of Economics**, v. 434, p. 578-598, 1993.

JUNG, Taehyun; EJERMO, Olof. Demographic patterns and trends in patenting: Gender, age, and education of inventors. **Technological Forecasting and Social Change**, v. 86, p. 110-124, 2014.

KAHNEMAN, Daniel. Maps of bounded rationality: A perspective on intuitive judgment and choice. **Nobel prize lecture**, v. 8, p. 351-401, 2002.

KING, Gary; ZENG, Langche. Logistic regression in rare events data. **Political analysis**, v. 9, n. 2, p. 137-163, 2001.

- KRUGMAN, Paul R. **Geography and trade**. MIT press, 1991.
- LAWLESS, Jerald F. Negative binomial and mixed Poisson regression. **Canadian Journal of Statistics**, v. 15, n. 3, p. 209-225, 1987.
- LEENDERS, Roger Th AJ. Modeling social influence through network autocorrelation: constructing the weight matrix. **Social Networks**, v. 24, n. 1, p. 21-47, 2002.
- LEMOS, Mauro Borges; CAMPOS, Bruno; BIAZI, Elenice; SANTOS, Fabiana. Capacitação Tecnológica e Catching Up: o caso das regiões metropolitanas emergentes brasileiras. **Revista de Economia Política**, v. 26, n. 1, p. 95-118, 2006.
- MAGGIONI, Mario; NOSVELLI, Mario; UBERTI, Erika. Space vs. Networks in the Geography of Innovation: A European Analysis. **Papers in Regional Science**, vol. 86, n. 3, p. 471 - 494. Agosto, 2007.
- MAGGIONI, Mario A.; NOSVELLI, Mario; UBERTI, Teodora Erika. Space versus networks in the geography of innovation: A European analysis. **Papers in Regional Science**, v. 86, n. 3, p. 471-493, 2007.
- MARSHALL, Alfred. **Principles of Economics**. London: Mcmillan and Co., Ltd. 8ª edição disponibilizada pela Library of Economics and Liberty, 1920.
- MOHR, John W.; GUERRA-PEARSON, Francesca. The differentiation of institutional space: Organizational forms in the New York social welfare sector, 1888-1917. **Remaking the Iron Cage: Institutional Dynamics and Processes**. Chicago: University of Chicago Press, forthcoming, 1998.
- MONTOBBIO, Fabio; STERZI, Valerio. The globalization of technology in emerging markets: a gravity model on the determinants of international patent collaborations. **World Development**, v. 44, p. 281-299, 2013.
- MORESCALCHI, A., PAMMOLLI, F., PENNER, O., PETERSEN, A. M., & RICCABONI, M. The evolution of networks of innovators within and across borders: Evidence from patent data. **Research Policy**, v. 44, n. 3, p. 651-668, 2015.
- NAGAOKA, S.; MOTOHASHI, K.; GOTO, A. Patent statistics as an innovation indicator In: HALL, Bronwyn H.; ROSENBERG, Nathan (Ed.). **Handbook of the Economics of Innovation**. Elsevier, 2010.
- NELSON, Richard R; WINTER, Sidney G. **An evolutionary theory of economic change**. Cambridge: Harvard University Press, 1982.
- NORTH, Douglass C. **Institutions, institutional change and economic performance**. Cambridge university press, 1990.
- OLIVEIRA, Priscila. Determinantes Espaciais da Capacidade de Invenção no Brasil: O Papel das Redes de Inventores. 2016, 112f. Dissertação (Mestrado em Economia). Juiz de Fora, Faculdade de Economia – Universidade Federal de Juiz de Fora, 2016
- OWAN, Hideo; NAGAOKA, Sadao. Intrinsic and extrinsic motivation for inventors. **Discussion papers**, v. 11022, 2011.

OWEN-SMITH, Jason; POWELL, Walter. Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community. **Organization Science**, 2004.

PAIER, Manfred; SCHERNGELL, Thomas. Determinants of collaboration in European R&D networks: empirical evidence from a discrete choice model. **Industry and Innovation**, v. 18, n. 1, p. 89-104, 2011.

PICCI, Lucio. The internationalization of inventive activity: A gravity model using patent data. **Research Policy**, v. 39, n. 8, p. 1070-1081, 2010.

POLANYI, M. **The tacit dimension**. Londres: Routledge and Kegan Paul, 1967.

ROGERS, Everett. **Diffusion of innovations**. (1a Ed.) Nova York: Free Press of Glencoe, 1962.

ROGERS, Everett. **Diffusion of innovations**. (3a Ed.) Nova York: Free Press of Glencoe, 1983.

SANTOS SILVA, JMC; TENREYRO, Silvana. The log of gravity. **The Review of Economics and statistics**, v. 88, n. 4, p. 641-658, 2006.

SANTOS SILVA, JMC; TENREYRO, Silvana. On the existence of the maximum likelihood estimates in Poisson regression. **Economics Letters**, v. 107, n. 2, p. 310-312, 2010.

SANTOS SILVA, JMC; TENREYRO, Silvana. Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. **Economics Letters**, v. 112, n. 2, p. 220-222, 2011a.

SILVA, J. M. C. S.; TENREYRO, Silvana. Poisson: some convergence issues. **Stata journal**, v. 11, n. 2, p. 207-212, 2011.

SAXENIAN, AnnaLee. **Regional networks: industrial adaptation in Silicon Valley and route 128**. Cambridge: Harvard University Press, 1994.

SCHERNGELL, Thomas; BARBER, Michael J. Spatial interaction modelling of cross-region R&D collaborations: empirical evidence from the 5th EU framework programme. **Papers in Regional Science**, v. 88, n. 3, p. 531-546, 2009.

SCHUMPETER, Joseph Alois. The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle. **Transaction publishers**, 1934.

SCHUMPETER, Joseph. **Capitalism, Socialism and Democracy**. Routledge, 1943.

SIDONE, Otávio José Guerci; HADDAD, Eduardo Amaral; MENA-CHALCO, Jesús Pascual. Scholarly publication and collaboration in Brazil: The role of geography. **Journal of the Association for Information Science and Technology**, 2016a.

SIDONE, Otávio José Guerci; HADDAD, Eduardo Amaral; MENA-CHALCO, Jesus Pascual. Science in Brazilian regions: Development of scholarly production and research collaboration networks. **Transinformação**, v. 28, n. 1, p. 15-32, 2016b.

- SINGH, Jasjit. Collaborative networks as determinants of knowledge diffusion patterns. **Management science**, vol. 51, n. 5, p. 756-770, 2005.
- SORENSEN, Olav; RIVKIN, Jan W.; FLEMING, Lee. Complexity, networks and knowledge flow. **Research policy**, v. 35, n. 7, p. 994-1017, 2006.
- TER WAL, Anne LJ. The dynamics of the inventor network in German biotechnology: geographic proximity versus triadic closure. **Journal of economic geography**, v. 14, n. 3, p. 589-620, 2014.
- TER WAL, Anne; BOSCHMA, Ron. Applying social network analysis in economic geography: framing some key analytic issues. **The Annals of Regional Sciences**, v. 43, p. 739 – 756, 2008.
- TORRE, Andre; GILLY, Jean-Pierre. On the analytical dimension of proximity dynamics. **Regional studies**, v. 34, n. 2, p. 169-180, 2000.
- RALLET, Alain; TORRE, André. Proximité et localisation. **Économie rurale**, v. 280, n. 1, p. 25-41, 2004.
- VON THÜNEN, JH. **Der isolierte Staat. Beziehung auf Landwirtschaft und Nationalökonomie**, 1826.
- WALLIS, W.D. **A Beginner's Guide to Graph Theory**. Springer Science & Business Media, 1995.
- WASSERMAN, Stanley; FAUST, Katherine. Social Network Analysis: Methods and Applications. **Structure Analysis in Social Sciences**, book 8, Cambridge University Press, 1995.
- WHITTINGTON, Kjersten Bunker; SMITH-DOERR, Laurel. Women inventors in context: Disparities in patenting across academia and industry. **Gender & Society**, 2008.
- WUCHTY, Stefan; JONES, Benjamin F.; UZZI, Brian. The increasing dominance of teams in production of knowledge. **Science**, v. 316, n. 5827, p. 1036-1039, 2007.
- ZARATIEGUI, J. M. Marshallian Industrial Districts Revisited. Part I. **Problems and Perspectives in Management**, v. 2, p. 80-97, 2004.