EMERSON GOMES DOS SANTOS

**A proximidade nas interações: revisão e proposta de novas
abordagens para medição e representação**

São Paulo

2017

EMERSON GOMES DOS SANTOS

**A proximidade nas interações: revisão e proposta de novas
abordagens para medição e representação**

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção
do Título de Doutor em Ciências.

Área de Concentração: Engenharia de
Produção

Orientador: Prof. Dr. Renato Garcia

São Paulo

2017

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, _____ de _____ de _____

Assinatura do autor: _____

Assinatura do orientador: _____

Catálogo-na-publicação

Santos, Emerson Gomes dos

A proximidade nas interações: revisão e proposta de novas abordagens para medição e representação / E. G. Santos -- versão corr. -- São Paulo, 2017. 135 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Produção.

1.Geografia da inovação 2.Interação universidade-empresa 3.Proximidade cognitiva 4.Dimensões da proximidade 5.Técnicas multivariadas
I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Produção II.t.

DEDICATÓRIA

Para Josefa e Manoel (*in memorian*).

AGRADECIMENTOS

Elaborar uma página com os agradecimentos não é uma tarefa simples, pois é necessário resumir todo o percurso percorrido e dificilmente conseguirei expor toda a minha gratidão. De qualquer forma, seguem os principais colaboradores deste trabalho que traduz uma importante meta alcançada na minha vida.

Ao meu orientador Renato Garcia, não só pelo tempo dedicado a este projeto, mas por toda a amizade e convivência dispensada neste período. Sendo um exemplo de profissional para minhas pretensões acadêmicas.

Aos amigos de pesquisa Veneziano, Suelene e Ariana. Com um ambiente rico em discussões e aprendizagem qualquer tema a ser investigado se tornaria facilitado.

Aos professores da banca por contribuírem para que as ideias contidas neste estudo pudessem ser aprimoradas.

À Universidade de São Paulo e a Escola Politécnica da USP, um local de oportunidades de ensino e de pesquisa no qual fiz e continuo fazendo amigos.

À Unifesp, instituição com algumas pessoas queridas que me dão ânimo para continuar contribuindo com a academia.

Um agradecimento especial aos irmãos Jacobianos (Davi, Denis, Fernando, Hommenig, Luis e Paulo) pela amizade exemplar, pelas discussões e inspiração a cada encontro.

À minha família por me fornecer sempre a motivação necessária para continuar nesta jornada contínua de aprendizagem.

Por fim, um agradecimento especial àquela pessoa que me faz refletir que a proximidade possui mesmo diversas dimensões, muito além da dimensão espacial. Obrigado Amália por ser parte importante da minha vida.

RESUMO

SANTOS, E. G. A proximidade nas interações: revisão e proposta de novas abordagens para medição e representação. 2017. 134 f. Tese (Doutorado) – Escola Politécnica, Universidade de São Paulo, São Paulo, 2017.

O objetivo desta tese é a proposição de uma classificação para organizar a perspectiva multidimensional da proximidade. Essa classificação organiza formas para medição, representação e análise das múltiplas dimensões da proximidade. Isto dado que ainda não foram utilizados indicadores suficientemente satisfatórios para medir as várias dimensões da proximidade e que a forma como as medidas são definidas podem influenciar os resultados inferenciais obtidos nos modelos estatísticos. Este estudo levantou as definições conceituais para a proximidade e as suas principais formas empíricas de medição e propõe uma organização que possibilita a escolha entre as diferentes possibilidades de indicadores e de outras técnicas para interpretação e análise dessas medidas. Com isso foi realizada uma aplicação com dados da interação Universidade-Empresa no Brasil. A análise avaliou tanto os limites do ganho com a proximidade quanto os mecanismos de complementariedade e de substituição entre as dimensões geográfica e cognitiva. O foco na proximidade cognitiva favoreceu a comparação entre diferentes formas para a sua medição. Comparação realizada na perspectiva geométrica com uso de Escalonamento Multidimensional (EMD) e Análise de Correspondência (AC). Os resultados mostram a importância que precisa ser dada na forma como as medidas são definidas e representadas já que as interpretações dos resultados inferenciais nos modelos estatísticos estão diretamente relacionadas com esse passo inicial. As contribuições com esse estudo são de natureza metodológica, porém de caráter prático para guiar estudos futuros a partir das técnicas organizadas. Dentro desse escopo, a tese visa que outras técnicas, ainda pouco exploradas se apresentem como alternativas de uso para a ampliação da literatura da área. Por fim, a própria classificação pode ser adaptada para incluir outras técnicas analíticas que incorporem dados não espaciais ou que lidam com dados de não proximidade.

Palavras-Chave: Geografia da Inovação; Interação Universidade-Empresa; Proximidade Cognitiva; Dimensões da Proximidade; Técnicas Multivariadas.

ABSTRACT

SANTOS, E. G. Proximity in interactions: revision and proposal of new approaches for measurement and representation. 2017. 134 f. Tese (Doutorado) – Escola Politécnica, Universidade de São Paulo, São Paulo, 2017.

The aim of this thesis is the proposition of a classification to organize the multidimensional perspective of proximity. This classification organizes forms for measurement, representation and analysis of the multiple dimensions of proximity. This is because satisfactory indicators have not yet been used to measure the proximity and that the way in which the measures are defined can influence the inferential results obtained in the statistical models. This study raised the conceptual definitions for proximity and its main empirical forms of measurement and proposes an organization that allows the choice between the different possibilities of indicators and other techniques for interpretation and analysis of these measures. Hence, an application with data of the University-Industry interaction in Brazil was performed. The analysis evaluated both the limits of the gain with proximity and the mechanisms of complementarity and substitution between the geographic and cognitive dimensions. The focus on cognitive proximity favoured the comparison between different ways of measuring it. Comparison carried out in the geometric perspective using Multidimensional Scaling and Correspondence Analysis. The results show the importance that must be given in the way the measures are defined and represented since the interpretations of the inferential results in the statistical models are directly related to this initial step. The contributions with this study are of a methodological nature, but of a practical nature to guide future studies from the techniques organized. Within this scope, the thesis aims that other techniques, still little explored are presented as alternatives of use for the broaden of the literature of the area. Finally, the classification itself can be adapted to include other analytical techniques that incorporate non-spatial data or that deal with non-proximity data.

Keywords: Geography of innovation; University-industry linkages; Cognitive proximity; Proximity dimensions; Multivariate techniques.

LISTA DE ILUSTRAÇÕES

Figura 1 – Mapa das Mesorregiões segundo Quantidade de Interações de origem nas empresas.....	82
Figura 2 – Mesorregiões segundo Quantidade de Interações de destino nos grupos de pesquisa.....	83
Gráfico 1 – Exemplo 1 de interpretação para modelos Tipo I e Tipo II.....	69
Gráfico 2 – Exemplo 2 de interpretação para modelos Tipo I e Tipo II.....	70
Gráfico 3 – Escalonamento Multidimensional para Distância Geográfica entre as mesorregiões (km)	81
Gráfico 4 – Escalonamento Multidimensional para Distância Cognitiva (Euclidiana) entre as mesorregiões.....	89
Gráfico 5 – Escalonamento Multidimensional para Distância Cognitiva (Cosseno) entre as mesorregiões.....	90
Gráfico 6 – Distribuição dos grupos ocupacionais potencialmente empregados em atividades de ciência e tecnologia e de pesquisa e desenvolvimento (P&D) para mesorregiões selecionadas.....	91
Gráfico 7 – Análise de Correspondência para obtenção da Distância Cognitiva (Qui-Quadrado) entre as mesorregiões – coordenadas “principais” para as linhas (mesorregiões) e as coordenadas “padrão” para as colunas	96
Gráfico 8 – Análise de Correspondência para obtenção da Distância Cognitiva (Qui-Quadrado) entre as mesorregiões — coordenadas “principais” para as linhas (mesorregiões)	97
Gráfico 9 – Gráficos de dispersão com respectivos testes de correlação de Pearson entre Y e as dimensões da proximidade	104
Gráfico 10 – Gráficos de dispersão com respectivos testes de correlação de Pearson entre Y e as variáveis que não representam distâncias	105
Quadro 1 – Índice Cosseno.....	38
Quadro 2 – Variedade Relacionada (RELVAR) e Variedade não Relacionada (UNRELVAR)	40
Quadro 3 – Índice de similaridade.....	41

Quadro 4 – Índice com base no coeficiente locacional	41
Quadro 5 – Índice com base em duas unidades de análises	42
Quadro 6 – Família 1 (com base em lei de potência)	48
Quadro 7 – Família 2 (com base no produto interno)	49
Quadro 8 – Família 3 (com base em lei de distribuição)	50
Quadro 9 – Proposta de classificação para mensuração, representação e análise da proximidade.....	67
Quadro 10 – Descrição das variáveis do estudo	102
Quadro 11 – Resultados da Análise de Correspondência (Stata 13)	133

LISTA DE TABELAS

Tabela 1 – População e quantidade de municípios por mesorregiões	76
Tabela 2 – Quantidade de Interações entre as mesorregiões das empresas (linhas) e mesorregiões dos grupos de pesquisa (coluna)	78
Tabela 3 – Lista dos municípios mais populosos para as mesorregiões e respectivas coordenadas georeferenciadas	79
Tabela 4 – Distância Geográfica entre as mesorregiões (km)	80
Tabela 5 – Distância Cognitiva (Euclidiana) entre as mesorregiões	86
Tabela 6 – Distância Cognitiva (Cosseno) entre as mesorregiões	87
Tabela 7 – Distância Cognitiva (Qui-Quadrado) entre as mesorregiões	94
Tabela 8 – Resumo de Estatísticas Descritivas para as variáveis do estudo	101
Tabela 9 – Estimativas dos parâmetros para Quantidade de Interações considerando Binomial Negativa Inflada de Zeros – Capacidade de Absorção	108
Tabela 10 – Estimativas dos parâmetros para Quantidade de Interações considerando Binomial Negativa Inflada de Zeros – Demais Variáveis	109
Tabela 11 – Estimativas dos parâmetros para Quantidade de Interações considerando Binomial Negativa Inflada de Zeros – Capacidade de Absorção (com termo quadrático)	112
Tabela 12 – Estimativas dos parâmetros para Quantidade de Interações considerando Binomial Negativa Inflada de Zeros – Demais Variáveis (com termo quadrático) ...	113
Tabela 13 – Estimativas dos parâmetros para Quantidade de Interações considerando Binomial Negativa Inflada de Zeros – Capacidade de Absorção (com interações) .	116
Tabela 14 – Estimativas dos parâmetros para Quantidade de Interações considerando Binomial Negativa Inflada de Zeros – Demais variáveis (com interações)	117

LISTA DE SIGLAS

AC	Análise de Correspondência
CBO	Classificação Brasileira de Ocupações
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
DMD	Desdobramento Multidimensional
DGP	Diretório dos Grupos de Pesquisa da base Lattes do CNPq
EMD	Escalonamento Multidimensional
FDP	Função de distribuição de probabilidade
IBGE	Instituto Brasileiro de Geografia e Estatística
IPDA	<i>Ideal Point Discriminant Analysis</i>
IPEA	Instituto de Pesquisa Econômica Aplicada
MLG	Modelos Lineares Generalizados
OMDA	<i>One-Mode Distance-Association</i>
P&D	Pesquisa e desenvolvimento
RAIS	Relação Anual de Informações Sociais do Ministério do Trabalho e Emprego
RELVAR	<i>Related Variety</i>
TMDA	<i>Two-Mode Distance-Association</i>
UNRELVAR	<i>Unrelated Variety</i>

SUMÁRIO

INTRODUÇÃO	14
1 A PROXIMIDADE: TEORIA E CONCEITO	19
1.1 A multidimensionalidade da proximidade	21
1.1.1 Dimensão Geográfica	21
1.1.2 Dimensão Cognitiva	22
1.1.3 Dimensão Organizacional	22
1.1.4 Dimensão Social	23
1.1.5 Dimensão Institucional	24
1.1.6 Outras dimensões e seus desdobramentos	25
1.2 Papel da proximidade em geografia econômica	26
1.2.1 Limites dos ganhos com a proximidade	27
1.2.2 Paradoxo da proximidade e nível ótimo de proximidade	28
1.2.3 Variedade Relacionada	30
1.3 Mecanismos de Complementariedade e de Substituição	31
1.3.1 Relação entre as proximidades geográfica e cognitiva	31
1.3.2 Inter-relacionamentos entre as outras dimensões da proximidade	32
1.4 Síntese conceitual	33
2 ABORDAGENS PARA MEDIÇÃO DA PROXIMIDADE	36
2.1 Medições das dimensões da proximidade	36
2.2 Proximidade cognitiva	37
2.2.1 Índice Cosseno	38
2.2.2 Variedade Relacionada	39
2.2.3 Outros índices para medir proximidade cognitiva	40
3 QUADRO PARA MEDIÇÃO, REPRESENTAÇÃO E ANÁLISE DAS DIMENSÕES DA PROXIMIDADE	43
3.1 Argumento I: O que medir? Sobre a natureza da distância	43
3.2 Argumento II: Como medir? Sobre a natureza da informação nos dados categorizados	51
3.3 Argumento III: Como representar e analisar? Sobre as perspectivas algébricas e geométricas e respectivas abordagens analíticas	53

3.4	Perspectiva Algébrica: Análise Clássica de Dados Categorizados e desenvolvimentos posteriores	54
3.4.1	Modelos Lineares Generalizados	54
3.4.2	Modelos Loglineares	55
3.4.3	Modelos de Associação.....	56
3.4.4	Modelos “ <i>Distance-Association</i> ”	57
3.5	Perspectiva Geométrica	59
3.5.1	Escalonamento Multidimensional	60
3.5.2	Análise de Correspondência.....	62
3.6	Quadro geral e discussão.....	65
4	APLICAÇÃO EMPÍRICA.....	72
4.1	Interação Universidade-Empresa	72
4.2	Definição da Unidade de Análise e Mensuração e Representação da Dimensão Espacial.....	74
4.3	Mensuração e Representação da Dimensão Cognitiva.....	84
4.4	Proposta de medição e representação da distância cognitiva com uso da Análise de Correspondência	92
4.5	Influência da medição da proximidade não espacial na análise inferencial...98	
4.5.1	Limites nos ganhos	110
4.5.2	Complementariedade e Substituição.....	114
5	CONCLUSÃO.....	118
	Referências Bibliográficas.....	122
	Apêndice A – Resultados da Análise de Correspondências	133
	Apêndice B – Resumo das inferências estimadas pelos modelos	135

INTRODUÇÃO

A geografia econômica ocupa-se com a distribuição da atividade econômica no espaço e sua abordagem evolucionária focaliza-se nos processos de aprendizagem que produzem padrões capazes de explicar a evolução espacial de organizações, setores, cidades ou regiões. Uma de suas áreas de estudo trata da natureza das externalidades de aglomeração e investiga se uma composição específica de setores econômicos em uma região proporciona mais externalidades (GLAESER *et al.*, 1992).

Esse aumento relaciona-se com a capacidade de inovar e com o crescimento regional ou econômico (BOSCHMA; FRENKEN, 2011; BOSCHMA; IAMMARINO, 2009). Quanto à inovação, ela ultrapassa as estruturas de conhecimentos existentes e se relaciona às novas combinações de conhecimento heterogêneo. Assim, a transferência de conhecimento depende profundamente da proximidade, já que a assimilação do conhecimento trocado não é possível sem ela (MATTES, 2012).

A proximidade foi analisada como um fenômeno geográfico em virtude do papel da proximidade geográfica (ou espacial) para a transmissão do conhecimento (FRITSCH; FRANKE, 2004; MATTES, 2012). No entanto, esse conceito ganhou caráter multidimensional (TORRE; GILLY, 2000; BOSCHMA, 2005), visto que foi identificada a necessidade de explicar, por exemplo, a ocorrência de interações entre parceiros distantes fisicamente (GERTLER, 2008; MATTES, 2012).

Alguns estudos podem ser considerados precursores na avaliação do conceito proximidade e na popularização do tema no debate (TORRE; GILLY, 2000; KIRAT; LUNG, 1999). Primeiramente, quanto à noção de proximidade baseada nas inter-relações entre a proximidade geográfica e as proximidades não espaciais (TORRE *et al.*, 2005); em um segundo momento, em relação a diferentes dimensões da proximidade (BOSCHMA, 2005; KNOBEN; OERLEMANS, 2006).

Logo, mesmo que a proximidade geográfica tenha um papel importante, ela, por si só, não pode ser condição necessária tampouco suficiente para aprendizagem, criação do conhecimento e inovação. Neste sentido, identifica-se a necessidade de novos

estudos sobre a importância da proximidade geográfica e o seu relacionamento com outras dimensões de proximidade (BOSCHMA, 2005; GIULIANI, BELL, 2005).

Estudos prévios contribuíram com a redução da ambiguidade conceitual relacionada às possibilidades de compreensão de formas não espaciais da proximidade. Uma das propostas mais importantes definiu quatro dimensões: cognitiva, organizacional, institucional e social (BOSCHMA, 2005).

Conceitualmente, cada uma das quatro dimensões é definida considerando características relacionadas à sua denominação: a dimensão cognitiva relaciona-se à base de conhecimento dos agentes envolvidos (NOOTEBOOM, 1999); a dimensão organizacional relaciona-se com as práticas organizacionais, mecanismos, hierarquias, entre outros (BOSCHMA, 2005); a dimensão social estabelece as relações no nível micro, que afetam os resultados econômicos e são baseadas em confiança (GRANOVETTER, 1985; BOSCHMA, 2005). E, por fim, com respeito à dimensão institucional, há as características que consideram as restrições informais (sanções, tabus, costumes, tradições e códigos de conduta) e as regras formais (constituições, leis, direitos de propriedade), que estruturam a interação política, econômica e social (NORTH, 1991; KNOBEN; OERLEMANS, 2006).

Para que ocorresse a supracitada diminuição na ambiguidade conceitual, foram necessários estudos empíricos para propor formas de medir e de representar essas dimensões. Porém, mesmo com o avanço dessa literatura, ainda não foram desenvolvidos indicadores suficientemente satisfatórios para investigar minuciosamente as várias dimensões da proximidade (WETERINGS; BOSCHMA, 2009; BOSCHMA; IAMMARINO, 2009; FRENKEN *et al.*, 2007; MATTES, 2012).

A dimensão espacial possui menor ambiguidade na sua representação, em virtude do uso de medidas diretas, como a distância em quilômetros ou outras. Em contraponto, as dimensões não espaciais possuem a necessidade de uso de indicadores baseados em *proxies*, os quais permitem refletir cada dimensão conceitualmente. Entre as dimensões de proximidade não espaciais, a cognitiva ganhou relevância por estar relacionada com o conhecimento do ponto de vista conceitual e por ter recebido maiores esforços nas propostas para sua medição empírica.

Uma das formas precursoras e mais utilizada nos estudos empíricos propõe medir a proximidade cognitiva pelo chamado índice cosseno, o qual é definido para mensurar a correlação não centrada entre vetores que representam a distribuição dos agentes em uma dada classificação (JAFÉ, 1989; BRESCHI; LISSONI; MALERBA, 2003). Outra proposta, também muito utilizada, é a de variedade relacionada, que mede a proximidade cognitiva considerando o grau de diversificação de um conjunto de dados com base em medidas de entropia (FRENKEN *et al.*, 2007; BOSCHMA; IAMMARINO, 2009; BOSCHMA; FRENKEN, 2011).

O índice cosseno mede o grau de similaridade entre dois vetores em termos de orientação. Contudo, não pode ser diretamente interpretável como distância, pois não necessariamente atende axiomas que definem a distância no sentido métrico como os autores desejariam. Da mesma forma ocorre com a variedade relacionada, que apresenta interpretação útil como índice de diversificação, por ser baseada em entropia. Porém, também não pode ser considerada uma distância, visto que frequentemente falha quanto a esses mesmos axiomas.

Nesse contexto, a partir das definições conceituais e estudos empíricos para medir a proximidade, nota-se a necessidade de renovar os esforços para o desenvolvimento de novas medidas para as dimensões não espaciais da proximidade, além de formas mais claras para interpretá-las. A forma como as medidas são definidas para a sua utilização nos modelos estatísticos é de fundamental relevância, já que podem ter influência direta nos resultados inferenciais obtidos.

Assim, o objetivo desta tese é a proposição de uma classificação para organizar a perspectiva multidimensional da proximidade e realizar uma aplicação empírica. Para tanto, utilizou-se como ponto de partida uma reflexão dos estudos teóricos e empíricos, com uma abordagem multidisciplinar, a fim de realizar uma contribuição de caráter metodológico sobre as técnicas específicas para esse propósito de medição, representação e análise.

Técnicas analíticas foram levantadas e classificadas em duas abordagens, denominadas regra de distância e de projeção. Com a contraposição dessas abordagens, fica mais nítido como as medidas utilizadas para representar a

proximidade podem ser interpretadas. A interpretação de distância é mais intuitiva do que a de projeção, cuja representação implica uma abstração com a interpretação da projeção de pontos.

Para a redução de possíveis caminhos falhos ou de interpretações incorretas no desenvolvimento de pesquisas, as duas abordagens são definidas e justificadas em três argumentos que apresentam a lógica para criação do quadro geral, os quais passam pelas questões mais básicas sobre o que e como medir, até alcançarem a indagação de como representar e analisar as dimensões da proximidade.

Além disso, foi proposta uma classificação em três famílias de medidas, que podem ser utilizadas como índices de similaridades ou dissimilaridades. A separação e a especificação de medidas segundo as suas formas de cálculos também revelam novas possibilidades de índices a serem explorados para medir as dimensões da proximidade.

Por fim, os conceitos resumidos na classificação proposta permite realizar reflexões empíricas com dados das interações entre universidades e empresas no Brasil. Especificamente, esse estudo foi realizado para avaliar as dimensões geográfica e cognitiva, dada a quantidade de interações como variável de interesse e as mesorregiões como unidades de análise.

A proximidade não espacial entre as mesorregiões de empresas que interagem com os grupos de pesquisa de universidades foi obtida de três diferentes formas, segundo as três famílias, para possibilitar uma discussão metodológica do uso de diferentes abordagens analíticas, como o Escalonamento Multidimensional (EMD) e a Análise de Correspondência (AC).

Esta aplicação teve foco no problema da influência das dimensões da proximidade na quantidade de interações em dois aspectos: foram avaliados tanto os limites do ganho com essa proximidade quanto os mecanismos de complementariedade e de substituição entre as dimensões dessa proximidade.

O principal resultado mostra a importância da forma como as medidas são definidas, já que há diferenças de significância nos resultados inferenciais da proximidade

cognitiva nos modelos estatísticos; notadamente, observa-se pela comparação dos resultados nos modelos estimados. As demais variáveis utilizadas como preditoras dessas interações relativas às características internas e externas que motivam o interesse das empresas em interagir com grupos de pesquisa nas universidades permanecem significativas e de acordo com a literatura.

Assim, ressalta-se quão sensível é a escolha da forma como a proximidade é definida e medida e o quanto a interpretação dos resultados nos estudos realizados sobre o tema está diretamente relacionada com esse passo inicial. É viável que alguns deles possam ser questionados e, até mesmo, reavaliados, salvo os trabalhos nos quais foram feitas as devidas ressalvas e limitações nas definições dessas medidas.

A estrutura da tese está composta por 5 capítulos. Além da introdução, a tese apresenta no primeiro capítulo uma revisão dos tópicos relacionados à discussão teórica sobre o conceito de proximidade, suas inter-relações e seus usos. No capítulo dois, é apresentado o levantamento das abordagens empíricas para medir e representar a proximidade. O terceiro capítulo detalha a proposta de classificação, que considera duas abordagens e as respectivas técnicas analíticas. No capítulo quatro, uma aplicação é apresentada, para tanto utiliza-se uma base de dados com interações entre universidades e empresa no Brasil. Por fim, o capítulo final discorre sobre as conclusões e as considerações finais, resumindo os principais resultados das análises e as sugestões de linhas para estudos futuros.

1 A PROXIMIDADE: TEORIA E CONCEITO

Este capítulo apresenta o conceito teórico de proximidade, os estudos realizados sobre o tema e, por fim, um quadro resumo do conceito e de suas dimensões. A seção inicial mostra, a fim de motivar, como o conceito ganhou o *status* de multidimensionalidade e as limitações deste fato, já que uma maior complexidade para essa abordagem traz um *trade-off* em relação à capacidade de interpretação. Duas seções são apresentadas na sequência e estão relacionadas com a discussão sobre a importância que o conceito possui para explicar fenômenos em geografia econômica, como a transferência de conhecimento, a inovação e a interação.

O termo proximidade sempre teve um importante significado em áreas como matemática e geografia (TORRE; GILLY, 2000). O interesse dado à noção de proximidade para a economia remete aos estudos de localização (KIRAT; LUNG, 1999). A integração do espaço na análise econômica esteve presente, mesmo que indiretamente, desde 1890 (MARSHALL, 1961; VON THUNEN, 1926). A atenção ao tema pode estar ligada a uma tendência em economia e, em particular, em economia industrial e espacial, de os estudos se dedicarem ao contexto no qual as empresas estão inseridas para produção e inovação, e em suas relações com os outros atores econômicos (TORRE; GILLY, 2000).

Trabalhos iniciais buscaram direcionar estudos sobre a noção de proximidade, merecendo destaque um grupo de pesquisadores (conhecidos como a escola francesa de proximidade) que, desde o início da década de 1990, estudava a importância crescente do conceito de proximidade em economia (TORRE; GILLY, 2000). Com a contribuição de que a proximidade teria mais significado do que apenas geográfico, o grupo enfatizou a existência de outras dimensões, distinguindo entre proximidade geográfica e organizacional: a primeira relacionada à distância espacial e a segunda, à forma de interagir e de se organizar. Torre e Gilly (2000) consideraram também a proximidade cognitiva como parte da organizacional, ao incorporarem uma extensão quanto a base de conhecimento, além de discutirem também uma terceira forma, a proximidade institucional, que incorporaria o ambiente institucional (TORRE; GILLY, 2000; KIRAT; LUNG, 1999).

Outras questões negligenciadas ou subestimadas foram trazidas à tona em uma edição especial da revista acadêmica *Regional Studies*, que consolidou estudos sobre o tema, entre eles: as relações entre proximidade e localização; a disjunção entre a necessidade de proximidade geográfica e a colocação dos atores ao introduzir a discussão da mobilidade profissional; o conceito de proximidade geográfica temporária; e os efeitos negativos da proximidade geográfica, que poderiam ser limitados por outras proximidades com potencial na resolução de conflitos e processos de cooperação ou negociação (TORRE *et al.*, 2005).

A proximidade apresentada em cinco dimensões por Boschma (2005) tornou-se o quadro teórico mais amplamente citado e utilizado nos estudos subsequentes sobre proximidade. A preocupação em definir as dimensões geográfica, cognitiva, organizacional, institucional e social fortaleceu de vez a idealização de haver um conceito de proximidade multidimensional e a relevância da proximidade para aprendizagem e inovação (BOSCHMA, 2005).

A proximidade geográfica por si só não determina a capacidade de inovar de uma empresa (GIULIANI; BELL, 2005). Deste modo, pretendia-se saber, especificamente, se a proximidade geográfica ainda tinha papel explicativo importante, visto que outras dimensões também poderiam cumprir esse objetivo. Seguindo essa linha, a seção 1.1 deste trabalho levanta, para além desses estudos precursores, as discussões conceituais sobre as dimensões propostas para o conceito de proximidade e outras classificações possíveis.

A partir da definição de dimensões da proximidade, há outra preocupação na literatura em relação aos inter-relacionamentos entre elas e sua influência conjunta ao explicar fenômenos. Este tópico foi desdobrado em duas seções, ambas relacionadas ao uso do conceito proximidade: uma quanto ao papel da proximidade para transferência do conhecimento e inovação (seção 1.2) e outra a respeito dos mecanismos de compensação ou substituição (seção 1.3). Com base nesse contexto, a última seção (1.4) apresenta um quadro resumo das técnicas levantadas.

1.1 A multidimensionalidade da proximidade

Esta seção apresenta as dimensões da proximidade a partir do quadro teórico apresentado em Boschma (2005). Essas dimensões foram discutidas por outros autores, os quais contribuíram para a apresentação de uma visão conceitual dessas dimensões (KNOBEN; OERLEMANS, 2006). Uma seção final foi incluída para considerar outras dimensões e mostram alguns desdobramentos de outras linhas para a discussão.

1.1.1 Dimensão Geográfica

A proximidade geográfica é definida de uma maneira restrita, referindo-se à distância espacial ou geográfica entre atores econômicos de forma absoluta ou relativa. Em geral, considera uma distância absoluta ou relativa, por exemplo, o tempo de viagem, e apresenta pouca ambiguidade, mesmo que meios indiretos possam levantar diferentes percepções do conceito declaradas pelos próprios agentes (KNOBEN; OERLEMANS, 2006; BOSCHMA, 2005).

Essa dimensão incorpora principalmente os benefícios da colocalização, como as externalidades positivas (JAFFE *et al.*, 1993) ou o compartilhamento de conhecimento tácito, que facilita a interação e favorece a transferência de conhecimento (TORRE; GILLY, 2000; HOWELLS, 2002). Ao facilitarem interações face-a-face entre os atores, essas interações favorecem o intercâmbio de informações de alta qualidade (BOSCHMA, 2005).

Alguns estudos mostram seu lado negativo, por exemplo regiões especializadas podem desenvolver um problema de *lock-in*, que pode ser resolvido com abertura geográfica e diversificação (externalidades de Jacob). Além dos estudos abordarem vantagens e desvantagens relativas à dimensão geográfica, outros a propõem como necessária somente em certas fases da colaboração (KNOBEN; OERLEMANS, 2006) ou por um período de tempo, como a proximidade geográfica temporária (TORRE *et al.*, 2005).

1.1.2 Dimensão Cognitiva

A dimensão cognitiva está relacionada com a base de conhecimento dos agentes envolvidos. Cada agente possui base cognitiva diferente e, portanto, capacidade absorptiva e potencial de aprendizagem diferentes, sendo esta necessária para identificar, interpretar e explorar conhecimento novo (COHEN; LEVINTHAL, 1990). Seu conceito foi desenvolvido por Nooteboom (1999) e comumente é definido como as similaridades na forma de como o mundo é percebido, interpretado, entendido e avaliado por estes agentes (WUYTS *et al.*, 2005; KNOBEN; OERLEMANS, 2006).

Dado o compartilhamento da experiência tecnológica, pode-se considerar um conceito específico de dimensão tecnológica, similar a dimensão cognitiva, porém menos amplo, sendo tecnologia vista como ferramentas, dispositivos e conhecimento intermediários para criar produtos e serviços (TUSHMAN; ANDERSON, 1986), e a base de conhecimento dos agentes sobre ela (KNOBEN; OERLEMANS, 2006). Deste modo, alguns autores utilizam o conceito de proximidade tecnológica de maneira análoga ao de proximidade cognitiva.

Devido ao conhecimento ser frequentemente cumulativo, com resultados localizados dentro da firma e de natureza tácita, as diferenças cognitivas tendem a persistir, o que implica que o conhecimento é disperso entre diferentes empresas (BOSCHMA, 2005). Assim, a criação de conhecimento e aprendizagem depende da combinação da diversidade e de capacidades complementares de agentes heterogêneos dentro e entre empresas (NOOTEBOOM, 1999).

1.1.3 Dimensão Organizacional

Embora a base de competências e o conhecimento sejam pré-requisitos para criação de conhecimento, também é necessária a capacidade de coordenar trocas entre uma variedade de atores, dentro e fora da empresa. Assim, as práticas organizacionais, os mecanismos, as hierarquias, entre outros, são os fatores que definem a proximidade organizacional (BOSCHMA, 2005).

Essa dimensão sofre de uma ambiguidade conceitual relativamente alta, inclusive sendo considerada uma categoria maior, que engloba a dimensão cognitiva (TORRE; GILLY, 2000) ou com outras subdimensões, por exemplo interorganizacional e intraorganizacional (BOSCHMA, 2005). Para Torre *et al.* (2005), a proximidade organizacional é definida com base em agentes cuja interação é facilitada por regras e rotinas de comportamentos, e dividem o mesmo sistema de representações ou de crenças (BROWN; DUGUID, 1991; KNOBEN; OERLEMANS, 2006).

A importância dessa dimensão reside na necessidade de controlar incerteza e oportunismo na criação de conhecimento dentro e entre as empresas (BOSCHMA, 2005); além disso, as interações entre organizações são mais eficientes e conduzem a melhores resultados quando o contexto organizacional de ambos os parceiros que interagem é semelhante. Como tal, a proximidade organizacional gera uma capacidade de combinar informação e conhecimento das partes colaboradoras, e de transferir conhecimento tácito e outros recursos não padronizados entre essas partes (BURMEISTER; COLLETIS-WAHL, 1997; KNOBEN; OERLEMANS, 2006).

1.1.4 Dimensão Social

A noção de proximidade social originou-se na literatura de *embeddedness* (GRANOVETTER, 1985), a qual indica que as relações econômicas são extensões incorporadas a partir do contexto social, principalmente das relações no nível micro, que afetam resultado econômico e são baseadas em confiança (BOSCHMA, 2005).

Alguns autores a consideram proximidade pessoal (SCHAMP *et al.*, 2004) ou relacional (COENEN *et al.*, 2004); outros a incluíram como parte da organizacional (FILIPPI; TORRE, 2003; OERLEMANS; MEEUS, 2005) e (TORRE *et al.*, 2005). Apesar da ambiguidade, o conceito é similar, pois está relacionado a agentes que pertencem ao mesmo espaço de relações. Outra fonte de ambiguidade é relativa aos dois níveis de análise: empresas podem ser analisadas como pertencentes a uma mesma comunidade de prática, devido à sua posição estruturalmente equivalentes na rede, ou avaliando em que medida duas empresas são similares relativamente a uma

terceira, desde que centrando nas características da própria empresa que interage (KNOBEN; OERLEMANS, 2006).

A importância dessa dimensão da proximidade recai sobre a coordenação das transações e permitem a troca de conhecimento em razão da confiança mútua, do parentesco e da experiência, bem como de recursos externos a serem mobilizados (KNOBEN; OERLEMANS, 2006). A proximidade social é reconhecida como um mecanismo chave para compreender os fluxos de conhecimento subjacentes à aprendizagem interativa e à inovação (SORENSEN *et al.*, 2006; AGRAWAL *et al.* 2008; BRESCHI; LISSONI, 2009; CASSI; PLUNKET, 2014).

1.1.5 Dimensão Institucional

A dimensão institucional está associada a um quadro institucional em um nível macro, diferenciando-se do nível micro, considerado tanto na noção de proximidade organizacional quanto na social. A função da instituição seria reduzir incerteza e diminuir custos de transação, visto que tanto as organizações formais quanto as informais influenciam a forma como os atores e organizações coordenam suas ações. Assim, uma estrutura institucional efetiva provém condições balanceadas de estabilidade, abertura e flexibilidade (BOSCHMA, 2005).

A definição é, em grande parte, incontestável, e baseada em North (1991), consistindo em restrições informais (sanções, tabus, costumes, tradições e códigos de conduta) e em regras formais (constituições, leis, direitos de propriedade) que estruturam a interação política, econômica e social (KNOBEN; OERLEMANS, 2006).

Institutions are the humanly devised constraints that structure political, economic and social interaction. They consist of both informal constraints; (sanctions, taboos, customs, traditions, and codes of conduct), and formal rules (constitutions, laws, property rights) (NORTH, 1991, p. 97).

O conceito também apresenta alguma ambiguidade por ser estudado em dois níveis de análise, como a proximidade social. Em um nível geral, ele é baseado na similaridade entre o quadro institucional de países ou de regiões, entretanto há o seu equivalente em um nível menos agregado, relativo aos efeitos dessas instituições

nacionais sobre as normas e rotinas presentes nas organizações (KNOBEN; OERLEMANS, 2006).

A proximidade institucional facilita a aprendizagem coletiva, por permitir a livre transferência de conhecimento entre os agentes, com base em um espaço comum de representações, modelos, normas, procedimentos e regras aplicadas ao pensamento e à ação (KIRAT; LUNG, 1999; KNOBEN; OERLEMANS, 2006).

1.1.6 Outras dimensões e seus desdobramentos

Além das cinco dimensões reconhecidas como relevantes e estudadas com mais atenção ao longo dos anos, outras dimensões e observações foram levantadas e se apresentam como desdobramentos de distintas linhas para a discussão. Entre elas, há as dimensões: cultural, econômica, pessoal e de superproximidade.

Usada com menos frequência e relativamente consistente entre autores, a proximidade cultural, às vezes considerada como social, é tida como padrão de pensamentos, sentimentos, comportamentos e símbolos, vinculando um dado grupo em um tempo específico, o que dá significado para ações e comportamentos. Ela possui dois diferentes níveis de análise (GETLER, 1995): primeiro entre nações e regiões e outro nas diferenças de cultura organizacionais entre atores, medidas no nível relacional, similar a proximidade organizacional (KNOBEN; OERLEMANS, 2006).

Com base na hipótese centro-periferia, cientistas dos países periféricos estão dispostos a colaborar com os países centrais para obterem acesso aos recursos, enquanto as áreas centrais buscam complementaridades (HWANG, 2008; SONNENWALD, 2007). Acosta *et al.* (2011) apresentou uma equação de gravidade que inclui a distância econômica; com base nessa teoria, considerou o conceito de proximidade econômica. Seus resultados mostraram que níveis similares de recursos dedicados a P&D desempenham um papel positivo para facilitar a colaboração científica. Como eles reconhecem, esse achado não está em linha com a hipótese centro-periferia, pela qual se espera que, com os crescentes níveis de colaboração entre as regiões núcleo-periferia, elas se beneficiem com complementaridades. No

entanto, argumentam que esse resultado não é estranho, porque quanto maior a quantidade de recursos, maiores são as oportunidades de mobilidade e de participação em conferências internacionais que estimulam o estabelecimento e o fortalecimento de contatos pessoais para futuras colaborações (FERNÁNDEZ; FERRÁNDIZ; LEÓN, 2016)

Alguns autores pretendiam estudar a proximidade no nível pessoal, pois consideram que, até então, pouco se sabe sobre como as colaborações são afetadas com ela e na sua ausência; além disso, ressaltam que há a hipótese de que elementos pessoais afetam colaborações. Assim, o conceito para a proximidade pessoal considera as características pessoais dos parceiros com que colaboram, o que engloba o grau de similaridade nas características e nos comportamentos pessoais dos agentes. A suposição é que quanto menos os parceiros forem diferentes, mais provável eles terão proximidade pessoal. Especificamente, as colaborações prosperam em um sentimento mútuo de aceitação, apreciação e interesse nas ideias do outro (WERKER; OOMS; CANIËLS, 2016).

A noção de superproximidade foi introduzida para trazer à tona a força da infraestrutura física e virtual de várias escalas geográficas e de tempo, e as implicações econômica-espacial da tendência de redução das distâncias e das implicações da *e-society*. Kourtit (2016) apresenta a proximidade como um conceito central para entender a competitividade entre regiões e resalta a importância da *suprastructure (virtual infrastructure)* para gerar alto valor adicionado da proximidade virtual ou relacional. O autor enfatiza a dinâmica espacial, sendo o desenvolvimento urbano ou regional o resultado de várias forças gravitacionais de natureza material e virtual que, combinadas, formam o espaço econômico. Consequentemente, um determinado nível de desenvolvimento econômico de uma cidade ou região pode ser alcançado com diferentes combinações de medidas de proximidade físicas e virtuais.

1.2 Papel da proximidade em geografia econômica

Esta seção apresenta formas em que a literatura descreve a influência das dimensões da proximidade nos demais conceitos relativos aos fenômenos estudados em

geografia econômica, entre eles a transferência de conhecimento, a inovação e a interação, de modo a captar o potencial papel dos efeitos de transbordamentos.

É consenso que o conhecimento é uma entrada essencial para inovação e se constitui como componente-chave para o crescimento econômico e desenvolvimento (BRESCHI; LISSONI; MALERBA, 2003). A diferenciação entre as várias dimensões da proximidade serve como um ponto de partida interessante para compreender como a inovação pode ser organizada (MATTES, 2012).

1.2.1 Limites dos ganhos com a proximidade

Em geral, os estudos foram direcionados para mostrar que a proximidade, em suas várias dimensões, contribuiu positivamente para inovação pela formação de novas colaborações (CASSI; PLUNKET, 2014). Porém, esses ganhos poderiam ter um limite (BOSCHMA; FRENKEN, 2010).

O direcionamento dos estudos quanto à influência de pouca ou muita proximidade entre os agentes é bem objetivo em Boschma (2005), em razão de discutir os problemas decorrentes dessa influência para cada uma das cinco dimensões e suas possíveis soluções. O autor apresenta, no caso da proximidade cognitiva, que o problema de desentendimento quando há pouca proximidade ou o da falta de novidade quando há muita proximidade poderiam ser resolvidos em um *cluster* geográfico dotado de uma base de conhecimento comum, feita por recursos de conhecimento diversos e complementares; isto é, quando há uma capacidade de absorção aberta para novas ideias (NOOTEBOOM, 2000).

É reconhecido na literatura que agentes espacialmente concentrados se beneficiam de externalidades de conhecimento, ou seja, a proximidade geográfica facilita a troca de conhecimento tácitos ou codificados; quanto menor esta proximidade menor será a intensidade dessas externalidades positivas (HOWELLS, 2002). Empiricamente, estudos já mostraram que essas externalidades são limitadas geograficamente (JAFFE *et al.*, 1993; AUDRETSCH; FELDMAN, 1996).

Assim, a proximidade cognitiva facilita a comunicação e se considera que alguma proximidade cognitiva é necessária para compartilhar conhecimento, a fim de comunicar, entender, absorver e processar nova informação com sucesso. Porém, muita proximidade pode ser prejudicial, visto que alguns dos motivos para manter certa distância cognitiva estão relacionados à necessidade de dissimilaridade e de complementariedade para construir conhecimento; à dificuldade em desaprender rotinas boas que levam ao *lock-in*; por fim, ao risco de transbordamento involuntário, que são barreiras para a difusão (BOSCHMA, 2005).

Pode-se considerar raciocínio similar para as demais dimensões em relação aos limites nos ganhos com a proximidade. A proximidade organizacional é considerada necessária para controlar a incerteza e o oportunismo na criação de conhecimento dentro e entre organizações. Enquanto muita proximidade organizacional é acompanhada por falta de flexibilidade, pouca proximidade seria falta de controle, o que aumenta o perigo de oportunismo. Relações sociais baseadas em confiança facilitam a troca de conhecimento tácito, assim a proximidade social pode ser apontada como outro pré-requisito para aprendizado. Porém, muita proximidade pode ter impactos adversos, sendo desfavorável para novas ideias, dado que relações de longa data ou muito comprometimento podem resultar em redes fechadas para interações. Por fim, e da mesma forma, muita proximidade institucional é desfavorável para novas ideias e inovação em razão de obstruções para novas possibilidades, e pouca proximidade é danosa para ações coletivas devido a instituições formais fracas (BOSCHMA, 2005).

1.2.2 Paradoxo da proximidade e nível ótimo de proximidade

Em linha com os limites dos ganhos com a proximidade, um argumento conhecido como o paradoxo de proximidade implicaria que um efeito positivo desta no desempenho da inovação depende de um nível ótimo de proximidade e deve ser identificado dimensão por dimensão (CASSI; PLUNKET, 2014).

Embora um alto grau de proximidade, seja qual for a definição, e o envolvimento da rede sejam os principais motores da formação de redes e da difusão do conhecimento,

o impacto no desempenho inovador é bastante ambíguo, já que a proximidade entre os atores não se traduz necessariamente em maior desempenho inovador (BOSCHMA; FRENKEN, 2010). O chamado paradoxo de proximidade traz o argumento de que os condutores da formação de redes devem ser distinguidos dos determinantes do desempenho inovador. Se a proximidade e a chamada *embeddedness* explicam claramente a formação de relações de rede, o aprendizado e os fluxos de conhecimento não necessariamente podem beneficiar o desempenho inovador e até ser prejudiciais para o aprendizado interativo (BOSCHMA; FRENKEN, 2010; BROEKEL; BOSCHMA, 2012; CASSI; PLUNKET, 2014).

Para além de um nível ótimo de proximidade por dimensão, há ainda que se considerar a relação entre as proximidades. Por exemplo, a proximidade tecnológica estaria sujeita ao paradoxo, no sentido de que mais proximidade reduz o desempenho da inovação; quando a proximidade tecnológica aumenta, atinge um nível ótimo, após o qual o desempenho da inovação diminui. No entanto, as proximidades sociais e organizacionais parecem moderar esse impacto, deslocando o nível ótimo de proximidade tecnológica. Observando de forma diferente, quando os atores são sociais ou organizacionalmente próximos, eles gerenciam mais efetivamente sua similaridade, ganham efeitos de especialização e, como consequência, combinam níveis efetivamente maiores de proximidade tecnológica. De qualquer forma, em relação ao desempenho inovador, tem sido teoricamente e empiricamente demonstrado que a proximidade tecnológica tem uma relação curvilínea com o desempenho da inovação (MOWERY *et al.*, 1998; FAFCHAMPS *et al.*, 2010; CASSI; PLUNKET, 2014) e seus resultados mostram que o nível ótimo de proximidade tecnológica crítico para as invenções tecnológicas varia em função do contexto organizacional e social.

Deste modo, alguns estudos investigaram se existe um nível ótimo de proximidade, no sentido de que uma proximidade excessiva pode prejudicar o desempenho inventivo dos agentes e, se for o caso, verificaram se a complementaridade entre as diferentes dimensões de proximidade permite aos agentes econômicos lidarem com isso (BOSCHMA; FRENKEN, 2010; CASSI; PLUNKET, 2014).

1.2.3 Variedade Relacionada

Oportunidades de aprendizagem estão presentes quando a proximidade cognitiva entre os agentes de uma região e de fora desta não é nem muito pequena nem muito grande. Assim, o conhecimento externo contribui para o crescimento regional. Boschma e Iammarino (2009) propuseram o conceito relacionado ao papel da variedade: a variedade relacionada (*Related Variety*) como uma característica de regiões com setores complementares em termos de competências e a variedade não relacionada (*Unrelated Variety*) para setores que não compartilham competências complementares.

O conhecimento flui entre dois setores se a distância cognitiva não for muito grande. Logo, algum grau de proximidade cognitiva seria necessário, mas não suficiente para que esse fluxo ocorra (NOOTEBOOM, 2000). Além disso, seria crucial a variedade relacionada para que o fluxo de conhecimento extrarregional seja entendido e transformado em crescimento regional. Em seus estudos, o autor encontrou forte evidência que a variedade relacionada contribui para crescimento econômico regional. Regiões dotadas de setores complementares em termos de competências, ou seja, que mostram variedade relacionada, têm melhor desempenho.

O crescimento regional não é afetado simplesmente por ser bem conectado com o mundo de fora ou por ter alta variedade de conhecimento circulando dentro da região; em vez disso, os indícios apontam que o conhecimento extrarregional relacionado provoca aprendizagem intersetorial entre as regiões, mostrando a necessidade de especificar a natureza dessa ligação extrarregional. Boschma e Iammarino (2009) também ressaltaram a necessidade de estudos de caso ou outros estudos qualitativos para expandir o quadro conceitual, por exemplo indicando o tipo de mecanismos de transferência.

Assim, alguma proximidade cognitiva, mas não em excesso, é um pré-requisito; as outras quatro dimensões são consideradas mecanismos que podem aproximar os atores dentro e entre organizações. Embora pela sua definição cada dimensão possa fornecer indícios teóricos isoladamente, em teoria, se combinadas, elas podem enriquecer ainda mais as análises e fornecer outros subsídios para uma visualização

completa do quadro teórico da influência da proximidade. Assim, a próxima seção mostra as inter-relações entre as dimensões por mecanismos de compensação e de substituição.

1.3 Mecanismos de Complementariedade e de Substituição

Tanto teórica quanto empiricamente, há problemas para se estudarem as dimensões isoladamente (KNOBEN; OERLEMANS, 2006); apesar da proximidade geográfica poder ser importante, ela é simultaneamente influenciada pelas outras dimensões, complementando-as ou as substituindo (MATTES, 2012). Assim, as diferentes dimensões da proximidade desempenham diferentes papéis e se conectam (MATTES, 2012; HANSEN, 2014).

Apesar da complexidade metodológica Paci, Marrocu e Usai (2014) avaliaram empiricamente a articulação e a eficácia complementar de quatro diferentes dimensões: geográfica, tecnológica, social e organizacional, para transbordamentos de conhecimento em nível regional, na Europa. Os autores analisaram o efeito das várias dimensões na capacidade inovadora das regiões com a função de produção do conhecimento, na qual P&D e capital humano são as principais entradas; também avaliaram se as dimensões eram substitutas ou complementares para o transbordamento de conhecimento. Concluíram que as quatro dimensões têm um papel complementar significativo na geração de um importante fluxo de conhecimento entre as regiões, com a proximidade tecnológica, mostrando o efeito mais importante.

1.3.1 Relação entre as proximidades geográfica e cognitiva

Na literatura, atenção especial é dada para as dimensões geográfica e cognitiva (WUYTS *et al.*, 2005), devido à importância teórica e às limitações empíricas. Em teoria, a proximidade geográfica combinada com algum nível de cognitiva é suficiente para aprendizagem interativa ocorrer, entretanto as demais dimensões podem agir como substitutas da proximidade geográfica (BOSCHMA, 2005). Para se considerarem as relações entre as múltiplas dimensões simultaneamente nas estratégias de estimação, elas devem especificar um modelo abrangente, porém as

informações disponíveis, ainda que detalhadas, consideram apenas contextos geográficos limitados ou setores produtivos específicos (PACI; MARROCU; USAI, 2014).

1.3.2 Inter-relacionamentos entre as outras dimensões da proximidade

É provável que as dimensões se relacionem entre si, podendo se fortalecer ou se enfraquecer no tempo. Mesmo que algumas dimensões possam estar mais fortemente interconectadas teoricamente, como a organizacional, a social e a institucional, seria possível encontrar uma maneira de estabelecer um balanço entre as dimensões (BOSCHMA, 2005).

Diferentes quadros conceituais foram apresentados para conectar bases distintas de conhecimento e proximidades. Por exemplo, as proximidades cognitiva, organizacional e institucional podem ser classificadas como orientadas ao conhecimento, ao controle e à estrutura, enquanto que a geográfica e a social atuam como mecanismo de reforço. Assim, as proximidades geográfica e social seriam auxiliares, enquanto a organizacional, a institucional e a cognitiva seriam críticas para permitir a aprendizagem (MATTES, 2012). Nesta linha, primeiramente um certo grau de proximidade cognitiva seria crucial e necessário para a troca de conhecimento, dada a natureza da base de conhecimento; em seguida, ocorreria um *trade-off* entre as proximidades organizacional e institucional (MATTES, 2012).

Alguns estudos ocuparam-se com as dimensões da proximidade como mecanismo de complementariedade. Por exemplo, Kirat e Lung (1999) e Boschma (2005) ressaltaram que a proximidade organizacional pode ser facilitada pela geográfica, pois esta poderia ser complementar na construção e no fortalecimento das outras proximidades, assim como uma estrutura de governança pode resultar em um nível cognitivo satisfatório, o que implica que as dimensões organizacional e cognitiva da proximidade podem ser complementares (BOSCHMA, 2005).

Parte da literatura também mostra a proximidade como mecanismo de coordenação, com suas dimensões atuando como substitutas, ou seja, sua interação em um modelo seria negativa, significando que a posição importa menos quando os agentes já estão

próximos na rede ou trabalham para a mesma organização (CASSI; PLUNKET, 2014). Entre os estudos, as dimensões organizacional e cognitiva mostraram-se substitutas, enquanto a proximidade social poderia reduzir a cognitiva ao longo do tempo e a proximidade organizacional poderia causar a falta de social, apesar de ambas serem baseadas em fortes laços. Além disso, a proximidade geográfica estimula a social, pois a aglomeração pode compensar falta de proximidade social (BOSCHMA, 2005).

Por fim, sobre a existência e relevância de efeitos sinérgicos entre as dimensões, Capello e Caragliu (2015) indicaram o possível efeito positivo da proximidade social para a cooperação científica, porém com o impacto decrescendo, à medida que a distância espacial aumenta; quanto à proximidade cognitiva e à tecnológica, a sugestão é que alguma forma de complementaridade parece existir em relação à distância espacial.

1.4 Síntese conceitual

Pela revisão apresentada, o conceito de proximidade foi utilizado em diferentes contextos, sendo natural que haja o risco de que as suas dimensões tenham sido definidas de forma contraditórias, sobrepostas e sobre ou subespecificadas. Após realizado o levantamento do que compõe o conceito proximidade e de sua relevância para estudos em geografia econômica, esta seção apresenta uma síntese, a fim de facilitar a apresentação das tentativas de operacionalização empírica do conceito no capítulo subsequente.

Entre os trabalhos iniciais que merecem destaque, estão a escola francesa de proximidade (TORRE; GILLY, 2000; KIRAT; LUNG, 1999) e o quadro conceitual de Boschma (2005), pois buscam a redução das ambiguidades desse conceito, das quais cinco fontes foram apresentadas: 1) Os diferentes rótulos para as dimensões; 2) Os rótulos grosseiros ou rudimentares, por exemplo “não espacial”; 3) A sobreposição entre dimensões, por exemplo cultural e institucional; 4) As diferentes definições da mesma dimensão e 5) Os diferentes níveis de análise, concluindo pela redução da ambiguidade em apenas três dimensões: geográfica, organizacional e tecnológica (KNOBEN; OERLEMANS, 2006).

Apesar dos detalhes para mostrar a coerência nas decisões apresentadas, o processo não está isento de críticas e reforça a dificuldade em se limitar as dimensões discutidas na literatura. Como exemplo, a proximidade cognitiva, indicada como parte da proximidade organizacional, também foi comparada com a proximidade tecnológica (RALLET; TORRE, 1999; TORRE *et al.*, 2005).

De qualquer forma, as contribuições apresentadas na literatura visaram a redução da ambiguidade conceitual e a melhora na comparabilidade entre os estudos, para que os achados fossem mais consistentes e facilitassem o desenvolvimento cumulativo do conhecimento, por exemplo em estudos que explicam fenômenos nas áreas de inovação, de organizações e de regional (KNOBEN; OERLEMANS, 2006).

Já o estudo das relações entre as dimensões da proximidade contribuiu com para apresentar que, embora a proximidade geográfica tivesse papel importante, ela isolada não poderia ser condição necessária e nem suficiente para aprendizagem e inovação, sendo provavelmente fortalecida pelas demais (BOSCHMA, 2005). Com os desenvolvimentos teóricos, também foram realizados estudos com o objetivo de evidenciar empiricamente o papel dessas formas complexas de proximidade, enriquecendo a suposição simplista de que a proximidade espacial aumenta os transbordamentos de conhecimento (MAGGIONI; UBERTI, 2007; MORA; MORENO, 2010; AUTANT-BERNARD; LESAGE, 2011; FRENKEN; PONDS; VAN OORT, 2010).

Mais recentemente, surgiram algumas reflexões adicionais sobre o papel da proximidade espacial. Em primeiro lugar, a proximidade espacial pode realmente esconder outros tipos de semelhanças entre agentes nas esferas social, tecnológica e cognitiva; nesse sentido, é uma *proxy* para outros tipos de proximidades, como as proximidades sociais e cognitivas, maiores em áreas geográficas compactas. Em segundo lugar, o espaço em que as interações econômicas ocorrem é muito mais complexo do que aquele resumido pela geografia pura; assim, as teorias foram estendidas para abranger diferentes formas de conceber o espaço na análise das interações econômicas (CAPELLO, 2014; TORRE; WALLET, 2014; CAPELLO; CARAGLIU, 2015).

Por fim, ainda há espaço para estudar os efeitos da proximidade. Capelo e Caragliu (2015) apresentam os seguintes pontos: 1) Efeitos das dimensões de forma conjunta; 2) Efeitos de sinergias e complementariedades entre proximidades, em especial entre a espacial e as não espaciais; 3) Efeitos de não linearidade do impacto de proximidades não espaciais. De qualquer forma, somente pesquisas empíricas podem direcionar outros estudos para apoiar a teoria e ajudar no esclarecimento das diversas questões ainda em aberto. Deste modo, ao longo do próximo capítulo, foram detalhadas as tentativas empíricas para tornar mensurável o conceito da proximidade.

2 ABORDAGENS PARA MEDIÇÃO DA PROXIMIDADE

Estudos com propostas de criar medidas para as dimensões da proximidade, ou seja, criar formas para medir e representar essas dimensões não apresentam muitas variações e a literatura aponta que ainda não foram desenvolvidos indicadores satisfatórios para as várias dimensões da proximidade (WETERINGS; BOSCHMA, 2009; BOSCHMA; IAMMARINO, 2009; FRENKEN *et al.*, 2007; MATTES, 2012), devido a menor empenho metodológico sobre a forma ou às técnicas específicas para esse propósito de medição e representação.

A próxima seção apresenta algumas das principais formas de representação empírica das proximidades; a proximidade cognitiva será exposta em uma seção subsequente, por ter recebido mais esforços no seu desenvolvimento, visto que as tentativas empíricas para medir a proximidade cognitiva foram mais elaboradas do ponto de vista do cálculo para sua obtenção; desta forma, há três subseções específicas para mostrar as ideias precursoras e as variações de seus indicadores.

2.1 Medições das dimensões da proximidade

De forma geral, a proximidade geográfica é a dimensão que apresenta menos ambiguidade na sua medição, dada a possibilidade de utilização da distância em quilômetros como distância geográfica (MARROCU; PACI; USAI, 2013) ou o uso do inverso da distância como proximidade geográfica (CASSI; PLUNKET, 2014). Além dessas formas de representação da medida, o tempo de viagem entre os agentes também pode ser utilizado, como apontaram Ejermo e Karlsson (2006, *apud* MAREK *et al.*, 2016).

Para as dimensões não espaciais da proximidade, há mais variações para a sua medição empírica, já que esta não pode ser feita diretamente e são utilizadas *proxies* (outras características indiretas) para a representação de cada dimensão, as quais consideram os contextos específicos da aplicação dos estudos, porém podem ser resumidos como apresentado a seguir. Para a proximidade organizacional, é considerada, por exemplo, a informação de estabelecimentos na mesma empresa, de departamentos na mesma universidade ou de profissionais na mesma companhia,

além do uso das patentes como indicador para a mesma filiação dos aplicantes (MARROCU; PACI; USAI, 2013; CASSI; PLUNKET, 2014). Para a dimensão social, em grande parte dos estudos é utilizada a coautoria entre agentes em uma dada região ou entre regiões, seja com uso de patentes (MARROCU; PACI; USAI, 2013) ou com outras medidas, a partir da análise de redes sociais (CASSI; PLUNKET, 2014). Já em relação à representação da proximidade institucional, a forma mais comum e simples utilizada na literatura como *proxy* pode ser obtida por meio de variáveis dicotômicas, por exemplo para países ou para regiões no mesmo país a que pertençam as mesmas unidades de análise (MARROCU; PACI; USAI, 2013).

Atenção também foi dada ao risco na utilização de indicadores que sobrepõem as informações utilizadas nas medições, como indicam Marrocu, Paci e Usai (2013) em relação às dimensões organizacional e social. Por fim, uma forma geral de obter informações para todas as dimensões é por meio de questionários (GELDES *et al.*, 2015; PRESUTTI; BOARI; MAJOCCHI, 2011; HUBER, 2012; HANSEN, 2014; ENKEL; HEIL, 2014; MUSCIO; POZZALI, 2013), o que implica na limitação inerente da declaração dos respondentes. Essas informações podem ser utilizadas de forma direta (MOLINA-MORALES; GARCÍA-VILLAYERDE; PARRA-REQUENA, 2014) ou após cálculos de indicadores, como a distância euclidiana normalizada (CAPELLO; CARAGLIU, 2015).

2.2 Proximidade cognitiva

Os trabalhos empíricos com intuito de criar uma medida para a dimensão cognitiva ou tecnológica propõem indicadores que medem similaridades entre as bases de conhecimento dos agentes; em relação à proximidade tecnológica, considera-se como base de conhecimento a distribuição setorial ou de outra informação relevante. Utiliza-se, dependendo do contexto, a produção medida pela quantidade de patentes em uma dada classe tecnológica se o estudo for com parcerias entre empresas, ou pelos artigos científicos se o caso for entre parceiros acadêmicos (KNOBEN; OERLEMANS, 2006; GONÇALVES; FAJARDO, 2011; BRESCHI; LISSONI; MALERBA, 2003).

2.2.1 Índice Cosseno

Uma das principais abordagens utilizada em muitos estudos considera o índice cosseno para medir a correlação entre vetores que representam a distribuição de uma variável de interesse. Jaffe (1989) utilizou esse índice para caracterizar a posição tecnológica com uma medida de proximidade como separação angular (ou correlação não centrada) dos vetores, considerando a distribuição das patentes de empresas nos vários setores.

O quadro 1 apresenta a equação para o cálculo do índice cosseno, conforme Jaffe (1989). Essa medida foi definida a partir da separação angular entre os vetores, que é igual ao cosseno do ângulo, denominando-a de Proximidade Tecnológica. Sendo f_{ik} a fração de patentes da empresa i na área k , quanto maior o valor da medida (P_{ij}) maior será a proximidade. Deste modo, a autor procurou uma medida que refletiria a similaridade de foco de pesquisa da firma, porém como foi definida implica que essa proximidade tem relação com direção do vetor e não com seu comprimento.

Quadro 1 – Índice Cosseno

$$P_{ij} = \frac{\sum_{k=1}^K f_{ik} f_{jk}}{\left(\sum_{k=1}^K f_{ik}^2 \right)^{1/2} \left(\sum_{k=1}^K f_{jk}^2 \right)^{1/2}} .$$

Fonte: Jaffe (1989)

Com uma abordagem similar, outros autores propuseram o uso do mesmo índice (BRESCHI; LISSONI; MALERBA, 2003; ENGELSMAN; RAAN, 1994), analisando as co-ocorrências dos códigos de classificação de patentes. Aplicando essa contagem de ocorrências conjuntas a todos os pares possíveis de códigos de classificação, obtém-se uma matriz simétrica quadrada com a qual deriva a medida a partir da matriz utilizando o índice cosseno. O índice mede a separação angular entre os vetores que representam as co-ocorrências dos campos tecnológicos i e j com todos os outros campos.

Em outros exemplos de uso do índice cosseno, autores consideram a utilização do quadrado do índice para testar os limites de ganhos com a proximidade ou o efeito de

“U-invertido” e crescimento e decaimento a partir de um certo ponto (WUYTS *et al.*, 2005; CASSI; PLUNKET, 2014). Outros autores apresentam pequenas variações, como uso de coeficiente de correlação dos pares de perfis como medida da redundância de conhecimento entre firmas a partir da análise de redes (ENKEL; HEIL, 2014), ou o uso em anos diversos para verificar se há diferenças do índice ao longo do tempo (FERNÁNDEZ; FERRÁNDIZ; LEÓN, 2016).

2.2.2 Variedade Relacionada

A variedade relacionada foi operacionalizada em nível regional pelo cálculo de indicadores de relação tecnológica entre setores. Esse conceito foi representado por meio de medidas de entropia (FRENKEN *et al.*, 2007) a partir de dados de emprego ou de valor agregado em vários níveis de desagregação setorial (CAPELLO; CARAGLIU, 2015). Assim, as características das regiões quanto aos setores são consideradas complementares ou não em termos de competências e são definidas por variedade relacionada ou variedade não relacionada, respectivamente *Related Variety* (RELVAR) e *Unrelated Variety* (UNRELVAR).

Esses conceitos foram utilizados em outros estudos como grau de diversificação de exportação (BOSCHMA; IAMMARINO, 2009; BOSCHMA; FRENKEN, 2011). O quadro 2, representa o índice como apresentado em (BOSCHMA; FRENKEN, 2011). Considera-se p_i a participação dos setores de uma determinada representação agregada de setores, por exemplo três dígitos. Além disso, seja P_g a soma dessas participações para os respectivos setores considerando dois dígitos (S_g setores) e j a soma da informação para um dígito.

Quadro 2 – Variedade Relacionada (RELVAR) e Variedade não Relacionada (UNRELVAR)

$$RELVAR = \sum_{g=1}^G P_g H_g,$$

$$H_g = \sum_{i \in S_g} \frac{p_i}{P_g} \log_2 \left(\frac{1}{p_i/P_g} \right).$$

$$UNRELVAR = \sum_{j=1}^N P_j \log_2 \left(\frac{1}{P_j} \right),$$

Fonte: Boschma e Frenken (2011)

Estendendo o conceito existente de variedade relacionada a um cenário inter-regional, Capello e Caragliu (2015) propõem a construção de uma nova medida de "proximidade cognitiva inter-regional". Para tanto, identificaram as proximidades tecnológicas e sociais como semelhanças entre pares de regiões em termos dos perfis regionais de especialização social e industrial. Duas regiões são cognitivamente próximas se tiverem um conjunto complementar de habilidades e competências pertencentes a uma base de conhecimento comum, caracterizando um domínio tecnológico.

2.2.3 Outros índices para medir proximidade cognitiva

Para medir a proximidade cognitiva ou tecnológica entre regiões, Marrocu, Paci e Usai (2013) utilizaram um índice de similaridade entre a região i e j , com base na distribuição da atividade de patenteamento ou do emprego para diferentes setores. A fim de atrair novos conhecimentos de fora, as empresas e as regiões podem precisar aumentar a capacidade de absorção em torno da base de conhecimento existente e realizar atividades tecnológicas em domínios semelhantes.

Em outras palavras, a capacidade cognitiva é limitada e as empresas e regiões que compartilham uma base de conhecimento análoga podem trocar informações e conhecimentos e aprender umas com as outras mais facilmente. Seja l_{ik} a participação do setor k na região i , o índice t_{ij} é definido entre 0 e 1: quanto maior o seu valor mais a estrutura tecnológica das regiões será similar, como apontado no quadro 3.

Quadro 3 – Índice de similaridade

$$t_{ij} = 1 - \left(\frac{1}{2} \sum_{k=1}^{K=44} |l_{ik} - l_{jk}| \right)$$

Fonte: Marrocu, Paci e Usai (2013)

Já Capello e Caragliu (2015) utilizaram um coeficiente locacional para medir a proximidade tecnológica entre regiões, diferenciando-a da proximidade cognitiva, para a qual utiliza variedade relacionada. A seguir, o LQ_{qi} representa o quociente locacional para a região i e Q indica o número de setores.

Quadro 4 – Índice com base no coeficiente locacional

$$tech_{ij} = \frac{\sqrt{\sum_{q=1}^Q (LQ_{qi} - LQ_{qj})^2}}{Q}$$

Fonte: Capello e Caragliu (2015)

Por fim, considera-se um exemplo de índice que inclui dados a partir de duas fontes ou unidades de análises diferentes, como apresentado (SCHARTINGER; RAMMER; FRÖHLICH, 2002). Os autores utilizam-se das interações entre universidade e empresa e desenvolvem um índice em um estudo para o qual seja m a participação de interações do tipo t entre o setor econômico s e área científica r , tem-se de forma normalizada pelo valor esperado p_{rs} o índice apresentado no quadro 5.

Quadro 5 – Índice com base em duas unidades de análises

$$p_{rs} = \frac{(P_r \times P_s)^{0.5}}{\sum_{r=1}^R \sum_{s=1}^S (P_r \times P_s)^{0.5}} \quad {}^t n_{rs} = \frac{{}^t m_{rs}}{p_{rs}}$$

Fonte: Schartinger; Rammer; Fröhlinh (2002)

Dada a descrição das principais tentativas empíricas para medir as dimensões da proximidade, notam-se algumas formas de representação que podem ser classificadas a fim de melhor avaliarem as vantagens e desvantagens para uso desses indicadores propostos. A generalização das formas de medir a proximidade facilita a discussão metodológica e a comparação entre esses indicadores, o que pode favorecer a iniciativa de novas abordagens. Assim, o próximo capítulo apresenta uma proposta para uma classificação que organize as formas de mensuração, representação e análise do conceito proximidade.

3 QUADRO PARA MEDIÇÃO, REPRESENTAÇÃO E ANÁLISE DAS DIMENSÕES DA PROXIMIDADE

Como visto nos dois capítulos anteriores, estudos se preocuparam conceitualmente com as dimensões da proximidade (TORRE; GILLY, 2000; BOSCHMA, 2005; KNOBEN; OERLEMANS, 2006) e com a sua mensuração tanto para analisar os seus inter-relacionamentos quanto para avaliar a sua influência conjunta em outras variáveis (JAFPE, 1989; BRESCHI; LISSONI; MALERBA, 2003; BOSCHMA; IAMMARINO, 2009; BOSCHMA; FRENKEN, 2011).

A partir das definições conceituais e estudos empíricos com as principais tentativas de medição das dimensões da proximidade, nota-se a necessidade de renovar os esforços para abrir novas possibilidades de medir e representar essas dimensões. Para tanto, a criação de uma classificação para as formas de medir a proximidade poderia apoiar a escolha desses indicadores, além de formas para interpretação e análise facilitando a discussão metodológica para a iniciativa e uso de novas abordagens.

O maior detalhamento sobre os índices apresentados é crucial para melhor compreensão dos seus efeitos nas variáveis de interesse das pesquisas. Assim, este capítulo se apresenta em seis seções. As próximas três seções intituladas argumentos I, II e III apresentam a lógica e as justificativas para o quadro resumo que será apresentado na seção 3.6, esse resumo metodológico dos modelos identifica duas grandes abordagens baseadas em regras de distância e de projeção, respectivamente. Para as duas abordagens seguem detalhes das bases teóricas e de um esquema com as respectivas técnicas analíticas especificadas nas seções 3.4 e 3.5, seções que detalham essas técnicas a partir das perspectivas algébrica e geométrica.

3.1 Argumento I: O que medir? Sobre a natureza da distância

O conceito de distância é básico na experiência humana, usualmente significa algum grau de proximidade entre dois objetos físicos ou ideias e o termo métrico é utilizado como um padrão para uma medição. No sentido formal o conceito de distância

métrica, como uma função que satisfaz axiomas, e de espaço métrico foram originados por Maurice Fréchet (1878-1973) e Felix Hausdorff (1868-1942) e tornou-se uma ferramenta essencial para muitas áreas científicas (DEZA; DEZA, 2009). Assim, o primeiro argumento para elaborar uma classificação lida com o conceito de proximidade e distância, que podem ser vistas como medidas de similaridade e dissimilaridade, respectivamente. Para tanto, esta seção tem o objetivo de abordar alguns aspectos relevantes como a formalização matemática da proximidade e de seu aspecto geométrico, além da proposição de classes (famílias) de medidas a partir do uso desses conceitos em diferentes áreas.

Considera-se distância (ou dissimilaridade) sobre um conjunto X uma função, se para todo (x, y) de X ocorre:

a) não negatividade: $d(x, y) \geq 0$;

b) simetria: $d(x, y) = d(y, x)$;

c) reflexividade: $d(x, x) = 0$.

Uma função é denominada similaridade (ou proximidade) sobre o conjunto X se for não negativa, simétrica e $s(x, y) \leq s(x, x)$ para todo (x, y) de X , com igualdade se e somente se $x = y$.

Assim, nota-se a relação entre os conceitos de proximidade e de distância, sendo as principais transformações para obter distância (limitada por 1) a partir da similaridade:

$$d = (1-s); d = [(1-s) / s]; d = \sqrt{(1-s)}; d = \sqrt{2(1-s^2)}; d = \arccos(s) \text{ e } d = -\ln(s)$$

Os conceitos de distância semi-métrica e métrica podem ser definidos a partir de uma função d . Assim, define-se d como uma distância semi-métrica se além de a), b) e c) também atende a desigualdade triangular: $d(x, y) \leq d(x, z) + d(z, y)$ e, de forma geral, será considerada uma distância métrica sobre X se para todo x, y e z de X :

a) não negatividade: $d(x, y) \geq 0$;

b) identidade de indiscerníveis: $d(x, y) = 0$ se e somente se $x = y$;

c) simetria: $d(x, y) = d(y, x)$;

d) desigualdade triangular: $d(x, y) \leq d(x, z) + d(z, y)$.

Assim, para definir distância consideram-se três requisitos e um quarto item para definir uma subcategoria da distância: a distância métrica (WELLER-FAHY; BORGHETTI; SODEMANN, 2014), o adjetivo métrico é que indica a ligação a uma medida.

Para além das representações numéricas, algumas possíveis generalizações para as distâncias e os espaços métricos incluem a noção de espaço métrico probabilístico (distâncias como distribuições de probabilidades), espaço métrico Booleano, espaço métrico Fuzzy e espaço métrico sobre uma álgebra (espaço métrico com uma estrutura geométrica projetiva considerando, por exemplo, o produto interno). Em relação ao aspecto geométrico, as distâncias em geometria podem ser vistas nas perspectivas geodésica (caracteriza uma curva, generalizando a noção de linha reta), projetiva (que estuda as propriedades das figuras sobre projeção, incluem a geometria euclidiana e, de forma geral, a Minkowskiana), e não euclidianas (hiperbólica e elíptica, também chamada Riemannianas), além de outras que consideram superfícies, por exemplo. Para cada uma delas é definida a distância no respectivo espaço métrico.

Dada a possibilidade do desenvolvimento de distâncias similares em diferentes contextos torna-se tarefa quase impossível levantar e discutir de forma abrangente todas disciplinas e suas subáreas nas quais os coeficientes de similaridade e de dissimilaridade têm utilização e importância (WELLER-FAHY; BORGHETTI; SODEMANN, 2014). Para este trabalho foi considerada uma revisão sobre índices nas áreas transversais (Matemática e Estatística) e na área tema específica (Geografia Econômica), como mostrado a seguir nos itens a) e b):

a) indiretamente - Foram consideradas duas áreas de conhecimento transversais a Matemática e a Estatística. A matemática por conter aplicações em teoria dos grafos que consideram medidas a partir de interações em redes como na Internet, Web, Redes sociais e outras e a estatística, para a qual é corrente o uso de variadas formas de distância e que se estendem por diversas técnicas, entre elas: Escalonamento

Multidimensional (EMD), Análise de Agrupamento e Classificação e Reconhecimento de padrões;

b) diretamente - Considerou-se as áreas sociais aplicadas, em especial, a econômica e em específico a geografia econômica. Nesta disciplina, considera-se a proximidade geográfica e outros diversos tipos não espaciais de proximidade como as já levantadas (cognitiva, organizacional, social e institucional) para os quais as proximidades se baseiam em geral no índice cosseno.

Há diversas medidas de similaridade e de dissimilaridade e, apesar do seu uso nas diferentes disciplinas, a escolha depende da natureza dos dados que estão disponíveis e das possíveis formas de generalização dessas medidas. Estudos específicos sobre o relacionamento entre as medidas (CHA, 2007) fornecem uma categorização sintática e semântica das medidas aplicadas às funções de distribuição de probabilidade, a partir da correlação entre diferentes medidas usando análise de agrupamento, mas sem conseguir ser abrangente. Estudos mais abrangentes (DEZA; DEZA, 2009) fornecem uma enumeração das principais medidas de distância utilizadas em uma variedade de áreas, porém sendo uma lista extensa e interdisciplinar é útil quando se tenta identificar medidas específicas utilizadas em trabalhos publicados (WELLER-FAHY; BORGHETTI; SODEMANN, 2014).

Sendo a função de distribuição de probabilidade (FDP) uma das mais populares formas de representar padrões, há duas abordagens que considera medir distâncias com FDP: vetor e probabilística. Na primeira, um histograma ou uma FDP podem ser considerados como um vetor, isto é, um ponto no espaço euclidiano ou um sistema de coordenadas cartesiano e uma distância geométrica podem ser aplicadas para comparar duas FDP (por exemplo P e Q) pela sobreposição entre elas. Já a versão probabilística é baseada no fato de que um histograma de uma medição fornece a base para uma estimativa empírica da FDP.

Sugere-se neste ponto a categorização das distâncias em famílias de modo a alcançar certa abrangência inicial e reunir as principais medidas que possibilitem o desenvolvimento de pesquisas futuras. Uma proposta de agrupamento para os diversos tipos de medidas existentes é apresentada a seguir em três famílias, com as

respectivas características. Quanto ao atendimento dos axiomas, observa-se que somente o primeiro grupo é especificado como distância métrica no sentido exposto (WELLER-FAHY; BORGHETTI; SODEMANN, 2014; CHA, 2007).

As medidas consideradas como família 1 são definidas com base em lei de potência (WELLER-FAHY; BORGHETTI; SODEMANN, 2014) e podem ser consideradas como distâncias métricas (CHA, 2007). As medidas são apresentadas no quadro 6, sendo as principais representantes desta classe: quarteirão (ou *City Block*), euclidiana e a generalização denominada por Minkowski. Outras medidas que podem ser consideradas nesta família são as variações da medida Quarteirão (família L_1) e da *Intersection family* que consideram a intersecção entre duas FDP, ambas apresentadas nas tabelas 2 e 3 em Cha (2007).

Na escolha de uma medida, deve-se considerar determinadas características, por exemplo, com a medida Lorentzian, considera o logaritmo para garantir a propriedade de não negatividade. A maioria das medidas relacionadas a *Intersection family* são similaridades que podem ser transformadas em medidas baseadas em distância do tipo Quarteirão, apresenta-se a medida Wave Hedges nessas duas formas como exemplo.

Quadro 6 – Família 1 (com base em lei de potência)

Denominação da Medida	Função para cálculo
Quarteirão	$\sum_{i=1}^d P_i - Q_i $
Euclidiana	$\sqrt{\sum_{i=1}^d P_i - Q_i ^2}$
Minkowski	$\sqrt[p]{\sum_{i=1}^d P_i - Q_i ^p}$
Tchebychev	$\max_i P_i - Q_i $
Lorentz	$\sum_{i=1}^d \ln(1 + P_i - Q_i)$
Wave Hedges	$\sum_{i=1}^d \left(1 - \frac{\min(P_i, Q_i)}{\max(P_i, Q_i)}\right) = \sum_{i=1}^d \frac{ P_i - Q_i }{\max(P_i, Q_i)}$

Fonte: Elaboração própria a partir de Cha (2007).

Considera-se como família 2 as medidas definidas com base no produto interno, também denominadas semelhanças de correlação (WELLER-FAHY; BORGHETTI; SODEMANN, 2014). O quadro 7 inclui algumas das formas a partir de Cha (2007). Os requisitos para serem consideradas como uma distância não necessariamente se aplicam, assim são tidas como medida de similaridade. Sendo o índice cosseno o representante mais utilizado e, como um tipo mais simplista de similaridade de correlação, mede o grau de similaridade entre dois vetores em termos de orientação (WELLER-FAHY; BORGHETTI; SODEMANN, 2014). Outras medidas nessa família apresentadas são a Média Harmônica e outras versões de produto interno normalizado denominadas Kumar e Hassebrook, Jaccard e Dice.

Quadro 7 – Família 2 (com base no produto interno)

Denominação da Medida	Função para cálculo
Produto Interno	$P \bullet Q = \sum_{i=1}^d P_i Q_i$
Média Harmônica	$2 \sum_{i=1}^d \frac{P_i Q_i}{P_i + Q_i}$
Cosseno	$\frac{\sum_{i=1}^d P_i Q_i}{\sqrt{\sum_{i=1}^d P_i^2} \sqrt{\sum_{i=1}^d Q_i^2}}$
Kumar e Hassebrook	$\frac{\sum_{i=1}^d P_i Q_i}{\sum_{i=1}^d P_i^2 + \sum_{i=1}^d Q_i^2 - \sum_{i=1}^d P_i Q_i}$
Jaccard	$\frac{\sum_{i=1}^d P_i Q_i}{\sum_{i=1}^d P_i^2 + \sum_{i=1}^d Q_i^2 - \sum_{i=1}^d P_i Q_i}$
Dice	$\frac{2 \sum_{i=1}^d P_i Q_i}{\sum_{i=1}^d P_i^2 + \sum_{i=1}^d Q_i^2}$

Fonte: Elaboração própria a partir de Cha (2007).

Por fim, a família 3 denominada com base em lei de distribuição, inclui as formas de entropia e de probabilidade condicional. O quadro 8 apresenta algumas de suas

representantes as quais somente algumas podem ser qualificadas como distâncias, pois frequentemente falham quanto a desigualdade triangular e simetria (WELLER-FAHY; BORGHETTI; SODEMANN, 2014). Nesta família considerou-se variações das medidas fidelidade, qui-quadrado e entropia, respectivamente apresentadas nas tabelas 5, 6 e 7 em Cha (2007). Novamente a escolha de uma entre as medidas deve-se pautar em características desejadas, sendo a distância qui-quadrado útil para medir distância entre linhas e colunas em uma tabela. Entre as formas qui-quadrado, a medida de Pearson (qui-quadrado) pode não ser preferível por ser assimétrica, sendo utilizada a medida Aditiva simétrica (qui-quadrado), que com um método aditivo obtém tal simetria. Quanto a maioria das medidas relacionadas a Entropia, dado o conceito de Shannon de incerteza probabilística, apresenta-se a medida Kullback and Leibler assimétrica e uma forma simétrica obtida pelo método aditivo (Jeffreys).

Quadro 8 – Família 3 (com base em lei de distribuição)

Denominação da Medida	Função para cálculo
Fidelidade	$\sum_{i=1}^d \sqrt{P_i Q_i}$
Pearson (qui-quadrado)	$\sum_{i=1}^d \frac{(P_i - Q_i)^2}{Q_i}$
Neyman (qui-quadrado)	$\sum_{i=1}^d \frac{(P_i - Q_i)^2}{P_i}$
Aditiva simétrica (qui-quadrado)	$\sum_{i=1}^b \frac{(P_i - Q_i)^2 (P_i + Q_i)}{P_i Q_i}$
Kullback and Leibler	$\sum_{i=1}^d P_i \ln \frac{P_i}{Q_i}$
Jeffreys	$\sum_{i=1}^d (P_i - Q_i) \ln \frac{P_i}{Q_i}$

Fonte: Elaboração própria a partir de Cha (2007).

Nesta seção foi abordada a natureza da proximidade e alguns dos seus aspectos relevantes para a proposição de classe das medidas em famílias. O grande uso dos produtos internos nos trabalhos pode estar ligado ao fato de conceitualmente serem relacionados com as distâncias euclidianas (podendo ser convertidos em distâncias), sendo frequentemente usados como um índice para a similaridade de um par de vetores, considerando as variantes apresentadas nesta seção e como visto na prática na seção 2.2.1. Soma-se a isso que os produtos internos têm propriedades matemáticas convenientes e, portanto, embora não haja nenhuma maneira direta de coletar juízos de produtos internos, eles são fáceis de serem calculados.

Por fim, um produto interno é uma medida mais complexa do que uma distância, pois não envolve apenas os dois pontos (extremidades), mas também um terceiro ponto que serve como a origem. Assim, se a origem do sistema de coordenadas é deslocada no espaço, então os produtos internos também mudarão, enquanto as distâncias permanecerão as mesmas.

3.2 Argumento II: Como medir? Sobre a natureza da informação nos dados categorizados

A partir das possibilidades do que será medido pode-se estabelecer a forma de medir. Assim, o segundo argumento para elaborar uma classificação lida com a limitação quanto à quantidade a ser medida nos contextos especificados e a forma para representar as informações em proximidades. A perspectiva apresentada para medir o conceito proximidade será a partir das características teóricas definidas pelas classificações nas quais as unidades de análise ou indivíduos (pessoas, empresas, etc.) podem ser classificados e, posteriormente, comparados entre eles.

Primeiramente, nota-se que a literatura apresentada nos capítulos iniciais 1 e 2 diferencia a proximidade em espacial (ou geográfica) e não espacial, em parte por questões teóricas como visto nos estudos iniciais em geografia econômica, e por outro lado pela possibilidade da proximidade espacial ser medida de forma objetiva. Diferente das dimensões não espaciais, a proximidade espacial possui escala objetiva, por exemplo, ela pode ser medida em quilômetros ou em horas para percorrer

uma determinada distância. Assim, mesmo havendo autores que ressaltam a subjetividade na percepção da proximidade geográfica ela não apresenta as dificuldades de medição e representação se comparadas às dimensões não espaciais da proximidade.

Para pensar em medir (e representar) as formas não espaciais da proximidade entre indivíduos é preciso indicar inicialmente quais seriam as características dos indivíduos que representariam uma quantidade comparável entre eles. Desta forma, a ideia principal é compará-los ou associá-los a partir da definição das características relacionadas a cada dimensão. Considera-se, então, que se pode comparar pessoas e categorias entre si, ou associá-las a outras categorias, cujo resultado define uma relação. Isto posto já que indivíduos podem ser contados, classificados ou ranqueados, mas não podem ser medidos. Medir é a comparação numérica com uma unidade (HIESER, 2004).

Pelo exposto, apesar das diversas possibilidades de coleta de informações, entre elas as experimentações, os estudos observacionais e os levantamentos do tipo *survey*, e dos diferentes tipos de escala de medida: nominal, ordinal, intervalar e razão (STEVENS, 1946) as características utilizadas para as dimensões da proximidade serão informações do tipo “nominal” ou “ordinal” que agrupados são denominados dados categorizados. Dados categorizados desempenham papel importante em muitas áreas do conhecimento, em especial quando há limitações quanto à possibilidade de se obter medidas numéricas. Uma variável categorizada consiste num conjunto de categorias que não se sobrepõem e, portanto, só é possível obter contagens ou frequências de ocorrência de cada categoria da variável e são representadas na forma de uma tabela de contingência (AGRESTI, 2002; KATERI, 2014).

Dizer que há uma medida em uma "escala nominal" ou em uma "escala ordinal" (STEVENS, 1946) é um abuso de linguagem, porque o conceito de medição implica comparação numérica com uma unidade e para escalas nominais ou ordinais não há tal unidade. Porém, pode-se definir uma unidade de medida estatística medindo a posição relativa de um indivíduo em alguma população, determinando quantas unidades de dispersão ele está longe do centro dessa população. Essa abordagem,

fundamentada na psicometria, utiliza uma unidade de medida estatística (HIESER, 2004), já que não será possível medir estritamente no sentido de comparação com uma unidade objetiva, como por exemplo o metro. A perspicácia dos psicometristas e biometristas foi perceber que se pode medir a força de uma relação determinando a redução relativa da incerteza sobre a posição de um indivíduo em uma variável, já que se conhece a sua posição na outra variável.

O uso de uma unidade estatística para a medição de relações entre variáveis foi o que Galton, Yule e Spearman fizeram respectivamente para: as relações lineares entre as variáveis numéricas (GALTON, 1888), as relações entre as variáveis binárias (YULE, 1900) e para as relações entre as variáveis ordinais (SPEARMAN, 1906). Portanto, embora não seja possível medir diretamente os indivíduos, há um caminho indireto para medir: Pode-se medir a força da associação entre categorias em amostras ou populações (HEISER, 2004).

3.3 Argumento III: Como representar e analisar? Sobre as perspectivas algébricas e geométricas e respectivas abordagens analíticas

Com a possibilidade de medir de maneira indireta a proximidade entre os indivíduos pela força da associação entre categorias para as quais eles podem ser classificados surge a necessidade de levantar a variedade de técnicas para a análise de dados categorizados. Porém, considera-se uma base comum para apresentação e classificação de tais técnicas para além de seu desenvolvimento algébrico de forma mais geral para incluir outra perspectiva, sendo as categorias tratadas e analisadas tanto no aspecto algébrico dos modelos quanto ao aspecto geométrico para uma representação espacial. Assim, outro elemento da classificação proposta foi considerar como terceiro argumento as abordagens analíticas nas respectivas representações algébricas e geométricas.

As duas próximas seções 3.4 e 3.5 trazem detalhes das abordagens analíticas dada as duas perspectivas. Com a perspectiva algébrica se quer a estimação de modelos com objetivo de possibilitar a atribuição de escores para as categorias das variáveis contidas em uma análise para a qual haja interesse na relação entre linhas e colunas

de uma tabela de contingência, tabela que resume a informação das variáveis categorizadas. Já em relação a representação geométrica apresenta-se uma base comum de métodos que lidam com categorias do lado do espaço e considera-se uma correspondência simples entre os dados e a respectiva representação geométrica que, por vezes, opta-se pela perda de informação na representação para que se ganhe na interpretação e compreensão dos dados.

3.4 Perspectiva Algébrica: Análise Clássica de Dados Categorizados e desenvolvimentos posteriores

Esta seção apresenta métodos para a análise de dados categorizados, abordagem utilizada quando todas as características observadas em um estudo são representadas por categorias e a maneira mais comum de apresentar os dados é uma tabela de contingência produzida pela classificação cruzada dos atributos.

A informação dada por uma tabela de contingência é tradicionalmente resumida através de medidas apropriadas (medidas de associação), que diferenciam de acordo com a natureza das variáveis de classificação subjacentes (nominal ou ordinal). As subseções seguintes apresentam o desenvolvimento do modelo linear generalizado, introduzido por Nelder e Wedderburn (1972), que unifica vários modelos estendendo suas opções e contribuindo para o surgimento de outros. Segue com os modelos loglineares, para os quais se utiliza a informação adicional de ordem e atribui-se escores a eles e, por fim, a última seção traz os modelos “*Distance-Association*” que consideram a inclusão da distância.

3.4.1 Modelos Lineares Generalizados

A generalização dos Modelos Lineares Generalizados (MLG) incide essencialmente sobre dois aspectos:

a) a distribuição de probabilidades associada à variável resposta aleatória Y que não se restringe à distribuição Normal, podendo ser qualquer distribuição numa classe designada de família exponencial de distribuições, que engloba outras várias

distribuições, como a normal, gama e normal inversa para dados contínuos e binomial, poisson e binomial negativa para contagens;

b) a relação entre a combinação linear das variáveis preditoras (componente sistemático) e a variável resposta (componente aleatório) pode ser mais geral do que no modelo linear (função de ligação). O componente sistemático é estabelecido durante o planejamento do experimento, resultando em modelos de regressão, de análise de variância ou outros.

3.4.2 Modelos Loglineares

Considere uma tabela de contingência que classifica de forma cruzada uma amostra de acordo com duas variáveis categorizadas. As probabilidades das células são denotadas por π_{ij} e as frequências esperadas por $\mu_{ij} = n\pi_{ij}$. Sendo n_{ij} as contagens observadas para cada célula com os dados cruzados. Assim, um modelo sob independência estatística para frequência esperada μ_{ij} de uma tabela de frequências I por J, com parâmetros: α_i e β_j para as marginais, seria:

$$\mu_{ij} = \mu \alpha_i \beta_j \quad \text{para } (i = 1, 2, 3, \dots, I) \text{ e } j = 1, 2, 3, \dots, J).$$

E, denotando a variável de linha por X e a de coluna por Y tem-se o modelo de independência:

$$\mu_{ij} = \mu \alpha_i \beta_j \Rightarrow \log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y$$

Sendo os efeitos de X e Y captados por: $\lambda_i^X = \log(\alpha_i)$ e $\lambda_j^Y = \log(\beta_j)$

No contexto dos modelos loglineares clássicos apresentam-se duas opções, além do modelo de independência parcimonioso, mas restritivo há também o modelo saturado, que consideraria X e Y não independentes. O que acrescentaria o termo de interação entre X e Y:

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$$

Uma revisão dos MLG e em especial dos modelos loglineares pode ser vista em vários textos (MCCULLAGH; NELDER, 1989; KATERI, 2014) assim como uma visão mais aprofundada dos modelos loglineares a partir da análise de dados categorizados (AGRESTI, 2002).

3.4.3 Modelos de Associação

Esta seção apresenta como os modelos de associação podem ser generalizados a partir dos modelos loglineares. Os modelos de associação preenchem a lacuna entre esses dois casos extremos apresentados (modelos independente e saturado), ou seja, impõe uma estrutura especial sobre a associação, mas reduz o número de parâmetros de interação, proporcionando modelos intermediários de dependência. Embora muito poderosos e de grande valor interpretativo, não são muito populares no uso, principalmente porque eles não são prontamente fornecidos como opções de modelo em softwares estatísticos padrão (KATERI, 2014).

Seja um modelo multiplicativo para frequência esperada μ_{ij} de uma tabela de frequências I por J com parâmetros: α_i e β_j para as marginais e, θ_{ij} para as associações: $\mu_{ij} = \mu \alpha_i \beta_j \theta_{ij}$ em ($i=1, 2, 3, \dots, I$ e $j = 1, 2, 3, \dots, J$).

Assim, os diferentes modelos de associação são obtidos por diferentes parametrizações em relação as restrições quanto ao conhecimento dos parâmetros envolvidos $\theta_{ij} = \lambda_{ij}^{XY} = \phi \mu_i \nu_j$.

Por fim, um modelo generalizado denominado por Goodman (GOODMAN, 1979) como RC(M) é definido por:

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \sum_m \phi_m \mu_{im} \nu_{jm}$$

Na próxima seção será apresentada uma extensão para a qual foram reparametrizadas as associações em termos de distâncias (ROOIJ; HEISER, 2005).

3.4.4 Modelos “Distance-Association”

Esta seção apresenta a proposta de reparametrização para incluir a distância nos modelos de associação. As seções seguintes apresentam mais detalhes sobre os dois modelos “distance-association”: *One-Mode Distance-Association* (OMDA) e *Two-Mode Distance-Association* (TMDA) sugeridos por Rooij e Heiser (ROOIJ; HEISER, 2005).

Seja a frequência esperada μ_{ij} de uma tabela de frequências I por J e, seja essa frequência para a célula ij dessa tabela inversamente relacionada com a distância entre pontos no espaço euclidiano representando as categorias i e j, de modo que quanto mais frequentemente uma combinação ocorre, mais semelhantes as categorias são e mais próximas estarão as categorias correspondentes no espaço psicológico. Assim, para especificar as relações entre as frequências na tabela e as distâncias, utiliza-se uma função decrescente monótona, ψ , mapeando distâncias com os parâmetros de associação:

$$\theta_{ij} = \psi(d_{ij})$$

Onde d_{ij} é a distância que satisfaz os axiomas métricos.

Da Psicofísica, ramo da psicologia que trata das relações entre estímulos físicos e fenômenos mentais, frequentemente utilizam-se duas funções para ψ , Decaimento exponencial – (SHEPARD, 1957) ou Gaussiana – (THURSTONE, 1927) e alguma da métrica da família Minkowski para especificar as distâncias:

$$\theta_{ij} = \exp(-d_{ij}(\mathbf{X})),$$

$$\theta_{ij} = \exp(-d_{ij}^2(\mathbf{X})).$$

$$d_{ij}(\mathbf{X}) = \left[\sum_{m=1}^M |x_{im} - x_{jm}|^r \right]^{1/r}, \quad r \geq 1,$$

Onde x_{im} é a coordenada da categoria i na dimensão m ($m = 1, \dots, M$), coletada em uma matriz K por M .

Assim, Rooij e Heiser (2005) propõem um novo modelo em termos de distâncias que resulta em uma interpretação de distância que é, sem dúvida, mais direta do que a parametrização usual do modelo RC(M). Os dois modelos desenvolvidos pelos autores foram:

- a) *One-Mode Distance-Association* (OMDA): Para tabelas de contingência quadradas assumindo um padrão de associação simétrica;
- b) *Two-Mode Distance-Association* (TMDA): Para tabelas retangulares onde a suposição de uma associação simétrica é descartada.

Os autores também discutem uma representação de razão de chances em termos de distâncias e os dois modelos em comparação com outros modelos relacionados para tabelas de contingência, levantando as indeterminações nos modelos, uma maneira de identificá-los e uma discussão geral dos modelos relacionados e suas conexões.

Em relação ao modelo OMDA considera-se uma tabela quadrada com duas variáveis ($K = 2$) e a distância euclidiana ($r = 2$) em conjunto com uma função de transformação Gaussiana o modelo “*One-Mode Distance-Association*” teria a forma:

$$\mu_{ij} = \mu \alpha_i \beta_j \exp(-d_{ij}^2(X)) \Rightarrow \log(\mu_{ij}) = \lambda + \lambda_i^R + \lambda_j^C - d_{ij}^2(X) \Rightarrow$$

$$\Rightarrow \log(\mu_{ij}) = \lambda + \lambda_i^R + \lambda_j^C - \sum_m (x_{im} - x_{jm})^2$$

$$\text{Onde: } \lambda = \log(\mu), \lambda_i^R = \log(\alpha_i) \text{ e } \lambda_j^C = \log(\beta_j)$$

Uma versão em termos de probabilidades condicionais foi proposta por Nosofsky (1985) e uma discussão mais geral sobre esses tipos de modelos para dados de reconhecimento de estímulos consultar Takane e Shibayama (1992).

Uma vez especificado o modelo, uma função de verossimilhança pode ser otimizada sob distribuições amostrais de Poisson independente, Multinomial, ou Produto Multinomial - detalhes são fornecidos nos apêndices de Rooij e Heiser (2005).

O modelo TMDA expande a ideia apresentada para tabelas retangulares $I \times J$, nesse caso são necessários dois conjuntos de coordenadas, uma para os pontos da linha (X) e outra para os pontos da coluna (Y). A distância é então definida por:

$$d_{ij}(X; Y) = \left[\sum_{m=1}^M (x_{im} - y_{jm})^2 \right]^{1/2}$$

Distâncias entre pontos de um conjunto não são relacionados com as observações, mas podem ser interpretados como distâncias entre as pontuações dos modelos RC(M). Escrito em termos do modelo loglinear, ele pode ser definido por:

$$\log(\mu_{ij}) = \lambda + \lambda_i^R + \lambda_j^C - d_{ij}^2(X; Y) = \lambda + \lambda_i^R + \lambda_j^C - \sum_m (x_{im} - y_{jm})^2, \text{ onde: } \lambda = \log(\mu),$$

$$\lambda_i^R = \log(\alpha_i) \text{ e } \lambda_j^C = \log(\beta_j)$$

O modelo *Ideal Point Discriminant Analysis* (IPDA) proposto por Takane (1987) é uma versão similar do TMDA em termos de probabilidades condicionais. Da mesma forma, uma vez especificado o modelo, ele é estimado por uma função de verossimilhança, detalhes são fornecidos nos apêndices de Rooij e Heiser (2005).

3.5 Perspectiva Geométrica

Diferente das variáveis numéricas ou escalares, as associações para variáveis categorizadas possuem as características de positividade e exaustividade que levam a uma representação geométrica específica (HEISER, 2004). Esta seção apresenta uma base comum dos métodos que lidam com categorias pela perspectiva espacial. O interesse está em uma forma que relaciona as distâncias na representação geométrica com as frequências em que as categorias ocorrem e no levantamento dos principais conceitos nessa representação, como distância, massa, perfis, vértices, centro de gravidade (baricentro) e a inércia, para os quais a origem está na física.

Assim, considera-se uma correspondência simples entre os dados e a representação geométrica e, por vezes, opta-se pela perda de informação na representação para que se ganhe na interpretação e compreensão dos dados. Em princípio, pode não ser necessária a redução da quantidade de informação. Porém, em algumas aplicações,

especialmente aquelas com mais de quatro categorias de colunas, a dimensionalidade pode ser muito alta para que a representação geométrica seja útil, e normalmente há de resolver o problema confiando em uma aproximação (HEISER, 2004).

As duas próximas seções apresentam dois conjuntos de técnicas com a perspectiva geométrica. O Escalonamento Multidimensional, que já foi utilizado em estudo anterior (ENGELSMAN; RAAN, 1994) em uma tentativa de criar um mapa de ciência e tecnologia combinando patente com publicações em uma análise temporal e a Análise de Correspondência, técnica utilizada em um trabalho recente para avaliar o tamanho e as características de agentes quanto a distância cognitiva.

3.5.1 Escalonamento Multidimensional

O Escalonamento Multidimensional (EMD) definido de forma ampla compreende uma família de modelos geométricos para a representação multidimensional de dados e um conjunto correspondente de métodos para ajustar tais modelos a dados reais. Apesar de uma definição mais restrita limitar o termo a modelos de distância espacial para semelhanças, dissimilaridades ou outros dados de proximidade, Carrol e Arabie (1980) inclui modelos que também se aplicam para:

- a) dados não proximais, por exemplo, dados de preferência;
- b) dados não-espaciais, por exemplo, modelos geométricos discretos, tais como estruturas em árvore;
- c) dados não distanciados métricos, por exemplo, utilizando o produto interno ou projeção (CARROL; ARABIE, 1980).

O primeiro método prático disponível para EMD foi devido à Torgerson (1952), Torgerson (1958) e Gower (1966) baseado em teoremas de Eckart e Young (1936) e Young e Householder (1938). A ideia básica do escalonamento clássico é assumir que as dissimilaridades são distâncias e então encontrar coordenadas que as expliquem. Pode-se dizer que os modelos representam medidas de similaridade ou dissimilaridade entre pares de objetos como distâncias entre pontos em um espaço

multidimensional de dimensão menor. Assim, pode-se destacar quatro objetivos principais (BORG; GROENEN, 1997):

- 1) inspeção visual e exploração dos dados para conhecer a estrutura dos dados;
- 2) testar hipóteses estruturais: como determinados critérios pelos quais se pode distinguir objetos de interesse são espelhados em correspondentes diferenças empíricas;
- 3) para descobrir as dimensões subjacentes aos julgamentos de similaridade ou de dissimilaridade;
- 4) como modelo psicológico para explicar os julgamentos de dissimilaridade em termos de uma regra que imita um tipo particular de função de distância.

Os modelos são definidos especificando como dados de similaridade ou dissimilaridade, as proximidades p_{ij} , são mapeadas em distâncias de uma configuração m -dimensional X . O mapeamento é especificado por uma função f que especifica como a proximidade devem estar relacionadas às distâncias $f: p_{ij} = d_{ij}(X)$. Na prática, se busca uma configuração (em uma dada dimensionalidade) cujas distâncias satisfazem f tanto quanto possível em termos de uma medida de ajuste ou função de perda que agrega os erros de representação sobre todos os pares (i, j) : $e_{ij} = f(p_{ij}) - d_{ij}(X)$. Uma soma de quadrados normalizada desses erros define Stress, a função de perda mais comum em EMD.

Assim, EMD irá representar as proximidades pelas distâncias entre os pontos em um espaço definido por coordenadas no qual as distâncias são medidas por uma regra. Uma particular escolha de $f: f(p_{ij}) = d_{ij}(X)$ especifica um modelo EMD, sendo as distâncias sempre desconhecidas e a função pode ser completamente especificada ou pode ser restrita a vir de uma determinada classe de funções (SHEPARD, 1957 e THURSTONE, 1927). Se f não vier de razões teóricas será baseada no nível escalar das proximidades (métrico e não métrico). Outra maneira mais óbvia de gerar tais modelos é especificar f de diferentes maneiras, por exemplo, pode-se dividir as proximidades em subconjuntos e especificar f 's independentes ou mesmo f 's

diferentes para cada subconjunto em vez de apenas um único f para todas as proximidades.

Uma discussão chave para o desenvolvimento de uma aplicação envolve abordagens para obter as proximidades, que usualmente não são baseadas diretamente de julgamentos de similaridade, mas de índices calculados a partir de outras informações, entre as formas pode-se destacar:

- 1) coletar julgamentos de similaridade diretamente entre os objetos, por exemplo, da comparação dois a dois ou por classificação (*rank*);
- 2) derivar medidas de similaridade a partir dos dados usuais de casos por variáveis, por exemplo, a partir da correlação, que expressa a tendência de padrões similares nos dados ou dos perfis dos atributos (família Minkowski);
- 3) converter as medidas de não-semelhança em medidas de similaridade, a partir de considerações teóricas, por exemplo, com o uso de probabilidades ou da relação contida no modelo gravitacional;
- 4) usar medidas de similaridade definidas para os dados de co-ocorrência, considerando as possibilidades de índices e normalizações.

Na prática, frequentemente apenas informações de ordem das dissimilaridades (ou proximidades) estão disponíveis e, nesses casos utiliza-se EMD não métrico e, se for possível especificar funções paramétricas utiliza-se EMD métrico. Diversos avanços foram desenvolvidos, em especial cita-se os modelos Desdobramento Multidimensional (*Multidimensional Unfolding*) que podem considerar dois conjuntos de objetos ou duas unidades na análise (BUSING; VAN DEUN, 2005).

3.5.2 Análise de Correspondência

A análise de correspondência (AC) procura uma representação gráfica das linhas e colunas de uma matriz não negativa, tipicamente, uma matriz de contagens ou frequências em uma tabela de contingências. Linhas e colunas como pontos (perfis) em um espaço de baixa dimensão tal que as distâncias entre pontos de linha / coluna

refletem a semelhança de "perfil" de linha / coluna da matriz. O método pode, em princípio, ser aplicado a qualquer tabela retangular com medida de similaridade não negativa (BORG; GROENEN, 1997; BEH, 2004).

A Análise de Correspondência (AC) é conhecida sob diferentes nomes, entre eles *Reciprocal Averaging*, *Dual Scaling*, *Canonical Correlation Analysis* (aplicada a dados qualitativos), pois foi descoberta de forma independente em diferentes áreas (RICHARDSON; KUDER, 1933; HIRSCHFELD, 1935). Guttman (1941) apresentou um tratamento algébrico, mas a ênfase gráfica e geométrica que popularizou a técnica tem sido em grande parte devido a Benzecri (1973) que cunhou o termo em francês. Há uma vasta literatura sobre CA: (NISHISATO, 1980; NISHISATO, 1994; LEBART; MARINEAU; WARWICK, 1984; GREENACRE, 1984; GREENACRE, 1994; GIFI, 1990; GREENACRE; BLASIUS, 1994). Para uma discussão sobre a relação de AC com EMD consultar Heiser e Meulman (1983) e Borg e Groenen (1997).

Assim, o objetivo da análise é determinar escores que descrevem o quanto as respostas para variáveis são semelhantes ou diferentes. Para tanto, considera-se medir a força de associação entre as linhas e as colunas em relação a completa independência a partir de uma medida denominada Inércia, que é proporcional a estatística qui-quadrado. Ao decompor a inércia total, identifica-se importantes fontes de informação que ajudam a descrever essa associação. O uso de diferentes decomposições resultará em interpretações diferentes dessa associação e resultará em diferentes representações gráficas. A decomposição mais comumente utilizada é a decomposição de valores singulares (ECKART; YOUNG, 1936) que particiona a informação (inércia) contida na tabela em componentes, de modo a identificar como as categorias de linhas e de colunas (perfis) contribuem para os eixos principais, sendo os eixos ordenados por ordem de importância.

Para o desenvolvimento, aplicação e interpretação dessa técnica outros conceitos relevantes envolvem:

a) um sistema de coordenadas, cuja obtenção é útil para visualizar as associações entre as categorias de linhas e de colunas, categorias que serão representadas

graficamente nesse sistema (gráfico Biplot). O centro do gráfico (centróide) é o ponto no qual se representa a completa independência;

b) distâncias: representadas a partir da origem, ou seja, desvios a partir da independência. Deste modo, por construção são distâncias euclidianas ponderadas pela massa e denominadas distância qui-quadrado. Além disso, esses resultados levam à conclusão de que quando dois perfis de linha, ou dois perfis de coluna, são semelhantes, então eles serão posicionados próximos um do outro no gráfico de correspondência e se dois perfis forem diferentes, então eles serão posicionados a uma distância um do outro, porém nada se pode concluir a respeito da distância entre perfis de linha e de coluna (BORG; GROENEN, 1997; BEH, 2004);

A respeito da distância entre perfis de linha e de coluna denominada *inter-point distance*, em uma série de artigos discutiu-se uma maneira de medir essas distâncias recodificando a tabela de contingência de duas entradas para ter a forma de uma matriz de indicadores (CARROL, GREEN; SCHAFFER, 1986 e 1987 e 1989). No entanto, Greenacre (1989) demonstrou que as alegações feitas por esses autores são falhas. Beh (2004) relata que o erro de Carrol, Green e Schaffer foi tentar impor características de Escalonamento Multidimensional para serem aplicáveis à análise de correspondência. Assim, resume que as distâncias na análise de correspondência simples devem ser interpretadas dentro de uma variável escolhida e conclusões sobre as distâncias entre categorias de variáveis diferentes só devem ser usadas como um guia informal.

Por fim, ressalta-se o desenvolvimento da técnica que considera ainda outras extensões como: Análise de Correspondência Múltipla, Análise de Correspondência Conjunta de (GREENACRE, 1988 e 1990 e 1991) e Análise de Correspondência Canônica de (TER BRAAK, 1986, 1987). Além dessas, outras tentativas surgem também para ajustar o método para que ele possa atender a problemas interdisciplinares, como a Análise de Correspondência “*Detrended*” de Hill e Gauch (1980). Outra discussão envolve as similares teóricas como, por exemplo, as relações entre AC e os modelos loglineares (BEH, 2004).

3.6 Quadro geral e discussão

A partir dos três argumentos descritos, apresentou-se a lógica para a proposta do quadro geral como um esquema que indique o que medir, como medir e as técnicas analíticas disponíveis a partir de duas perspectivas, uma algébrica e outra geométrica. A proposta de classificação organiza a perspectiva multidimensional da proximidade, indicando novas formas de medição, de representação e de análise das múltiplas dimensões da proximidade.

Em relação às formas de medição, foi apresentada na proposta a categorização das distâncias em famílias, de modo a alcançar certa abrangência inicial e reunir as principais medidas que possibilitem o desenvolvimento de pesquisas futuras (WELLER-FAHY; BORGHETTI; SODEMANN, 2014; CHA, 2007), com as respectivas características e quanto ao atendimento dos axiomas. Somente o primeiro grupo (família 1) é especificado como distância métrica no sentido exposto na seção 3.1.

A fim de considerar uma classificação para as técnicas analíticas, considera-se a classificação para os modelos analíticos (FAUST; WASSERMAN, 1993) como:

- a) Modelos de Associação;
- b) Modelos de Correlação.

Os modelos de associação podem se basear em projeção e em distância, como apresentados nas seções 3.4.3 e 3.4.4, e os modelos de correlação apenas em projeção, indicados na seção 3.5.2. De acordo com Faust e Wasserman (1993), ambos os modelos partem da noção de independência na tabela de contingência e diferem na forma de medir a força e a natureza da relação entre linhas e colunas, seja pela associação, como nos modelos loglineares que utilizam a razão de chances, ou pela correlação, como na Análise de Correspondência e na Correlação Canônica. Porém, caso se considere somente essa classificação, não seriam inclusas algumas abordagens com base em uma regra de distância.

Para complementar a classificação apresentada, observa-se também a discussão apresentada em Rooij e Heiser (2005), os quais destacam duas classes bem definidas de modelos analíticos:

a) Tipo I, em que a relação entre os dois conjuntos é descrita por uma regra de distância, para a qual as categorias de ambos os conjuntos devem ser representadas por pontos no espaço euclidiano, considerando que a distância entre eles descreve a relação entre as categorias dos dois conjuntos;

b) Tipo II, em que a relação é descrita por uma regra de produto interno ou projeção; neste, pelo menos um conjunto deve ser desenhado usando vetores, enquanto os pontos do outro conjunto são projetados sobre eles para representar o relacionamento.

Na proposta, Rooij e Heiser (2005) apresentam os modelos loglineares como exemplo de técnica do Tipo I; em relação ao Tipo II, consideram os modelos RC(M), Correlação Canônica, Análise de Correspondência. Desse modo, sugere-se mesclar ambas as classificações, considerando os modelos de associação e de correlação (FAUST; WASSERMAN, 1993) como Tipo I ou do Tipo II.

Além disso, os modelos do Tipo I (ROOIJ; HEISER, 2005) podem ser complementados, acrescentando-se a ele as técnicas de Escalonamento Multidimensional. Considerou-se isso, pois, de modo similar há uma regra de distância nos modelos EMD¹ (CARROL; ARABIE, 1980).

Assim, denomina-se, como Tipo I os modelos baseados na regra de distância (ROOIJ; HEISER, 2005; FAUST; WASSERMAN, 1993) em conjunto com os modelos de EMD (CARROL; ARABIE, 1980); uma segunda classe como Tipo II os modelos baseados na regra de projeção, similar à classe de Rooij e Heiser (2005), que consideram os modelos de associação sem regra de distância; e os modelos de correlação de Faust e Wasserman (1993).

Em geral, a interpretação de distância é mais intuitiva do que a de produto interno, entretanto a representação por produto interno pode ser útil, dadas outras

¹ Para uma tipologia mais completa e específica sobre escalonamento multidimensional, consultar Carrol e Arabie (1980).

características desejadas. A classificação proposta é apresentada no quadro 9, a seguir.

Quadro 9 – Proposta de classificação para mensuração, representação e análise da proximidade

	Técnicas analíticas segundo abordagem (TIPO)	
	TIPO I	TIPO II
Perspectiva Algébrica (modelos loglineares)	<i>Distance-Association</i> (OMDA e TMDA) (ROOIJ; HEISER, 2005), entre outros	Modelos de Associação (LL, R, C e RC) (GOODMAN, 1979), entre outros
Perspectiva Geométrica	Escalonamento Multidimensional (CARROL; ARABIE, 1980), entre outros	Análise de Correspondência (GREENACRE, 1984), entre outros
Regra	Regra de Distância	Regra de Projeção

Fonte: Elaboração própria.

Há muitas vantagens com a elaboração de uma classificação, como: o direcionamento de novos estudos e a redução de possíveis caminhos falhos ou de interpretações incorretas no desenvolvimento de pesquisas. Muitas vezes, gráficos do Tipo II são interpretados como do Tipo I². Rooij e Heiser (2005) acreditam que qualquer cenário com categorias de ambos os conjuntos representados como pontos no espaço euclidiano (representação gráfica de Tipo I) é intuitivamente interpretado por uma regra de distância e não por projeção.

Em AC, somente as distâncias dentro do conjunto de pontos de linha ou dentro do conjunto de pontos de coluna podem ser interpretadas diretamente se a normalização apropriada for escolhida. A relação entre os pontos de linha e os pontos de coluna só pode ser avaliada por projeção (GREENACRE, 1984, p. 119). Uma modificação do

² Cf. CLOGG, ELIASON e WAHL (1990), os quais interpretam uma representação de produto interno com distâncias.

modelo de AC proposto (CARROL; GREEN; SCHAFFER, 1986, 1987 e 1989) indica uma representação para a distância entre linhas e colunas, contestada pela curiosa definição e pela fraca representação dessas distâncias graficamente. Em modelos do Tipo I, as distâncias entre todos os pontos podem ser interpretadas diretamente³.

Uma comparação em termos de representação espacial e interpretação é apresentada no Gráfico 1, de Rooij e Heiser (2005), no qual se apresentam uma categoria de linhas, a_1 , e duas categorias de colunas, B_1 e B_2 ; supõe-se que o gráfico seja um modelo de Tipo II. A associação de a_1 com B_1 é dada pelo comprimento do vetor B_1 vezes o comprimento do vetor a_1 , vezes o cosseno do ângulo entre os dois vetores. Todos os pontos na linha pontilhada têm o mesmo valor para a associação entre a_1 e B_1 ; todos os pontos no lado inferior esquerdo dessa linha têm uma associação menor; e todos os pontos no lado direito superior têm um valor maior. O ponto B_2 está no lado direito superior da linha pontilhada, de modo que a associação de a_1 com B_2 é maior. Supondo que o gráfico seja do Tipo I, todos os pontos no círculo teriam o mesmo valor para a associação com a_1 que B_1 tem. Pontos fora do círculo têm um valor menor para a associação; dentro do círculo, eles têm um valor mais alto do que a_1 tem com B_1 . Assim, no modelo do Tipo I, a associação entre a_1 e B_2 é menor do que entre a_1 e B_1 . As conclusões derivadas de ambos os modelos espaciais são contraditórias; além disso, aqueles do Tipo I são intuitivamente mais claros.

³ Para uma discussão detalhada sobre as relações entre EMD e AC, ver Heiser e Meulman (1983) e Rooij e Heiser (2005).

Gráfico 1 – Exemplo 1 de interpretação para modelos Tipo I e Tipo II

Fonte: (ROOIJ; HEISER, 2005)

Uma segunda comparação dada é apresentada no Gráfico 2. Observam-se dois pontos de linha a_1 e a_2 e um ponto de coluna B_1 . Em um modelo de Tipo II, a projeção de B_1 no vetor de a_1 ou a_2 é a mesma, e a associação entre os pontos de linha e coluna é agora determinada pelo comprimento dos vetores. Como o comprimento do vetor a_2 é maior do que o comprimento do vetor a_1 , a associação de B_1 com a_2 é maior. Em uma interpretação de Tipo I, a distância de a_1 a B_1 é menor, indicando uma associação maior. Se uma representação é de Tipo I ou II gera uma diferença significativa, portanto é importante verificar uma distinção clara entre os dois. Um gráfico do Tipo I pode e deve ser feito utilizando somente pontos em espaço multidimensional, enquanto um do Tipo II deve ser desenhado por meio do uso de vetores para pelo menos um conjunto.

Gráfico 2 – Exemplo 2 de interpretação para modelos Tipo I e Tipo II

Fonte: (ROOIJ; HEISER, 2005)

É importante ressaltar que o quadro com as técnicas apresentadas pode abrir novos caminhos e direcionar pesquisas que considerem a aplicação e as comparações entre as diferentes abordagens, de modo a apoiar a escolha de uma técnica em detrimento de outra. Nessa linha, Rooij e Heiser (2005) mostraram que os modelos “*Distance-association*” podem ser considerados como uma reparametrização do modelo RC(M) e que os primeiros fornecem uma ferramenta útil para a análise da associação.

Assim, sugerem-se, como novas linhas de análise, as duas abordagens em forma de regras:

- a) regra de distância;
- b) regra de projeção.

Consideram-se, para as duas regras, as técnicas levantadas dos modelos Loglineares, com possibilidade de aplicação para as duas abordagens, e o

Escalonamento Multidimensional e Análise de Correspondência para cada regra de distância e de projeção, respectivamente.

Embora as técnicas EMD e AC tenham várias propriedades em comum, visto que ambas exibem graficamente os objetos como pontos em um espaço de baixa dimensão, diferem em outros aspectos: EMD é uma técnica de um modo (apenas um conjunto de objetos é analisado), enquanto AC é uma técnica de dois modos. Porém, há uma extensão para a técnica EMD, de modo a considerar objetos de linha e de coluna, técnica denominada Desdobramento Multidimensional (DMD).

Os dados em AC são restritos a serem não negativos, enquanto que mesmo uma ordem de classificação pode ser analisada por EMD. Para tabelas de contingência, o tipo de dados mais utilizado para AC são as frequências entre duas variáveis categóricas que utilizam a distância χ^2 como uma medida de dissimilaridade, enquanto a EMD pode aceitar quaisquer medidas de dissimilaridade. Em EMD, as distâncias entre pontos podem ser diretamente interpretadas, enquanto que, em AC, a relação entre pontos de linha e de coluna só pode ser avaliada por projeção.

4 APLICAÇÃO EMPÍRICA

Este capítulo apresenta o uso dos conceitos abordados na classificação (Quadro 9), de modo a ilustrar como o esquema pode apoiar pesquisas em relação à medição, à representação e à análise das dimensões da proximidade. Para considerar o uso das técnicas analíticas disponíveis para a representação das dimensões, foi escolhida a perspectiva geométrica.

A aplicação tem foco no problema da influência das dimensões da proximidade na interação universidade e empresa, a fim de captar o potencial papel dos efeitos de transbordamentos. Assim, a análise avaliou os limites do ganho com a proximidade e os mecanismos de complementariedade e de substituição entre as dimensões geográfica e cognitiva nessas relações, conforme os tópicos revisados respectivamente nas seções 1.2.1 e 1.3.1., com a premissa de que a proximidade geográfica combinada com algum nível da cognitiva seria suficiente para promover a aprendizagem pela interação (BOSCHMA, 2005).

O foco na dimensão cognitiva favorece a comparação entre diferentes formas de medição, resumidas e apresentadas na proposta de categorização das distâncias em famílias (seção 3.1), para posterior verificação empírica da importância da proximidade na significância dos parâmetros em modelos estimados com o objetivo de explicar a quantidade de interações entre universidades e empresas no Brasil.

4.1 Interação Universidade-Empresa

O papel da universidade na sociedade ultrapassa a formação de pessoal, sendo também reconhecida por sua importância para a inovação. Esse modelo de instituição obtém importância nos estudos relacionados a inovação, em especial estudos sobre a colaboração entre universidades e empresas, e sobre as características e as motivações que afetam essas interações (KLEVORICK *et al.*, 1995; MOWERY *et al.*, 1998).

As interações entre universidade e empresa já foram avaliadas por diversos estudos, e diversos autores identificaram os benefícios da proximidade nessas interações,

como Jaffe (1989), Audrestch e Feldman (1996), Mansfield e Lee (1996), Arundel e Geuna (2004), Laursen *et al.* (2011) e D'Este e Iammarino, (2010). A proximidade geográfica é um fator importante, mas não suficiente, visto ser necessário considerar outras dimensões da proximidade como complementares.

Além da proximidade, outras dimensões de análise foram estudadas para avaliar as interações entre universidade e empresas. Do ponto de vista da universidade, ressaltam-se: a qualidade da universidade (MANSFIELD; LEE, 1996; D'ESTE; IAMMARINO, 2010); o tamanho do departamento ou do grupo de pesquisa (MANSFIELD; LEE, 1996; DE FUENTES; DUTRÉNIT, 2016) e as diferentes áreas da pesquisa acadêmica (METCALFE, 2003), já que a acumulação de conhecimento varia entre as diferentes disciplinas.

Pela perspectiva das empresas, considera-se a capacidade de absorção, pois uma maior capacidade representa maior habilidade para avaliar, assimilar e explorar o conhecimento externo disponível (COHEN; LEVINTHAL, 1990; BISHOP *et al.*, 2011). Por fim, há a estrutura produtiva da região, que também pode influenciar as interações, como nos estudos de Storper e Venables (2004) e Duranton e Puga (2000), nos quais regiões urbanas e diversificadas reúnem condições mais adequadas para a troca de conhecimentos entre os agentes e para a geração de inovações.

No Brasil, alguns trabalhos relacionados ao tema estudaram a colaboração entre agentes, particularmente entre universidade e empresa. Esses estudos evidenciam a desigualdade regional, haja vista a concentração geográfica da colaboração entre os agentes. Para além da importância da proximidade geográfica nas relações estabelecidas, outras variáveis e técnicas, como análise de redes sociais, são consideradas como em Ruffoni *et al.* (2017). Alguns trabalhos, como os de Rapini *et al.* (2009), Suzigan *et al.* (2009); Garcia *et al.* (2011) e (COSTA; RUFFONI; PUFFAL, 2011), utilizaram a base de dados do Diretório dos Grupos de Pesquisa da base Lattes do CNPq (DGP) com o intuito de avaliar o papel e a importância dessas interações e identificar algumas das formas de transferência de conhecimento da universidade para as empresas (GARCIA *et al.*, 2014).

Deste modo e como levantado na seção 1.4, ainda há espaço para estudar os efeitos da proximidade nas interações e neste capítulo apresenta-se uma aplicação empírica na qual estuda-se os efeitos conjuntos das dimensões na forma de sinergias e complementariedades entre proximidades e os efeitos de não linearidade das proximidades.

4.2 Definição da Unidade de Análise e Mensuração e Representação da Dimensão Espacial

A principal fonte de dados das interações realizadas pelos grupos de pesquisa brasileiros com empresas é o DGP (RAPINI *et al.*, 2009; SUZIGAN *et al.*, 2009; GARCIA *et al.*, 2011), em razão de ser a mais ampla fonte de informações sobre as atividades dos grupos de pesquisa no Brasil, pois reúne os dados por meio da coleta de informações junto aos líderes dos grupos. Dentre essas informações, há os fatores que caracterizam os grupos e as características das interações.

Para o presente estudo, foi definido como período de referência o ano de 2010 do DGP e todas as áreas. Com isso, foram identificados 2.915 grupos de pesquisas que declararam possuir interações com 4.740 empresas, o que perfaz um total de 7.071 interações.

Os dados foram considerados de forma agregada por mesorregiões, de modo a reduzir o problema da dificuldade ou da falta de informações disponíveis, por exemplo quando consideram como unidade de análise as interações desagregadas entre empresas e os grupos de pesquisa. Porém, ao considerar todas as 137 mesorregiões chega-se a um total de 18.769 combinações possíveis (137^2), o que reduziria a capacidade de explicação do fenômeno, já que consideraria um percentual elevado de regiões em que não ocorram as interações entre si.

Deste modo, do ponto de vista de uma melhor representação e simplicidade das análises que se seguem, foram escolhidas as mesorregiões com população, em 2010, maior do que 1,5 milhão de habitantes. Essa escolha justifica-se pela premissa de que regiões densamente povoadas são mais propensas a produzirem externalidades a partir da aglomeração. Deste modo, espera-se uma configuração espacial favorável

para captar as peculiaridades do fenômeno em estudo por também não considerar uma medida endógena como a própria quantidade de interações.

Esse recorte resultou em uma lista com 36 mesorregiões, definidas como as unidades de análise do estudo, apresentadas na tabela 1, junto com a representatividade de cada uma em relação à sua população e à quantidade de municípios. Somadas as 36 mesorregiões, chega-se a 63,2% da população (120.596.296 de pessoas) e 42% dos municípios do Brasil (2.344 municípios).

O interesse está na quantidade de interações entre as mesorregiões. A premissa é de que a escolha para colaborar parte da empresa, ou seja, a frequência de interações ocorridas entre empresas de uma dada mesorregião com grupos de pesquisa acadêmicos de cada mesorregião, o que pode ser representado por uma matriz com 36 linhas (mesorregiões das empresas) e por 36 colunas (mesorregiões dos grupos de pesquisa), com as contagens ou co-ocorrências, como apresentado na tabela 2.

Tabela 1 – População e quantidade de municípios por mesorregiões

Abreviação	Mesorregião	UF	Total População (Censo 2010)	Quantidade de Municípios
1-AgresPern	Agreste Pernambucano	PE	2.217.600	71
2-Campinas	Campinas	SP	3.785.620	49
3-CentralEspSan	Central Espírito-santense	ES	1.987.103	24
4-CentroAmazon	Centro Amazonense	AM	2.726.732	30
5-CentroGoiano	Centro Goiano	GO	3.056.794	82
6-CentroNorBaia	Centro Norte Baiano	BA	2.226.300	80
7-CentroSulBaia	Centro Sul Baiano	BA	2.478.787	118
8-DF	Distrito Federal	DF	2.570.160	1
9-LesteAlagoia	Leste Alagoano	AL	2.064.525	52
10-LestePotiguar	Leste Potiguar	RN	1.532.717	25
11-MacroMetroPaulis	Macro Metropolitana Paulista	SP	2.644.519	36
12-MetroBelem	Metropolitana de Belém	PA	2.437.297	11
13-MetroBH	Metropolitana de Belo Horizonte	MG	6.236.117	105
14-MetroCuritiba	Metropolitana de Curitiba	PR	3.493.742	37
15-Metrofortaleza	Metropolitana de Fortaleza	CE	3.468.137	11
16-MetroPorAleg	Metropolitana de Porto Alegre	RS	4.742.302	98
17-MetroRecife	Metropolitana de Recife	PE	3.693.177	15
18-MetroSalvador	Metropolitana de Salvador	BA	4.210.499	38
19-MetroSP	Metropolitana de São Paulo	SP	21.154.988	45
20-MetroRJ	Metropolitana do Rio de Janeiro	RJ	12.578.485	30
21-NordesBaiano	Nordeste Baiano	BA	1.545.922	60
22-NordesParaense	Nordeste Paraense	PA	1.789.387	49
23-NoroeRioGrand	Noroeste Rio-grandense	RS	1.946.510	216
24-NortCentrParanae	Norte Central Paranaense	PR	2.037.183	79
25-NorteMinas	Norte de Minas	MG	1.610.413	89
26-NorteMaraense	Norte Maranhense	MA	2.605.412	60
27-RibeiraoPreto	Ribeirão Preto	SP	2.376.360	66
28-SaoJoseRioPreto	São José do Rio Preto	SP	1.569.220	109
29-SudesParaense	Sudeste Paraense	PR	1.647.514	39
30-SulBaiano	Sul Baiano	BA	2.011.740	70
31-SulSudeMinas	Sul/Sudoeste de Minas	MG	2.438.611	146
32-TriangMineiro	Triângulo Mineiro/Alto Paranaíba	MG	2.144.482	66
33-ValeItajai	Vale do Itajaí	SC	1.508.980	54
34-ValeParaibaPaul	Vale do Paraíba Paulista	SP	2.264.594	39
35-ValerioDoce	Vale do Rio Doce	MG	1.620.993	102
36-ZonaMata	Zona da Mata	MG	2.173.374	142
Total			120.596.296	2344

Fonte: Elaboração própria

Nota-se que das 1.296 possibilidades de combinações entre as 36 mesorregiões (36^2), em 391 (ou 30,2%) ocorreram pelo menos uma interação entre empresas e grupos de pesquisa, sendo em grande parte realizada na mesma mesorregião (2.710 interações ou 62,4%). A fim de facilitar a comparação entre a distribuição de interações na matriz, foi aplicada uma formatação, para a qual a maior quantidade de interações entre duas mesorregiões é representada em caselas (cruzamentos entre linhas e colunas das mesorregiões) mais escuras, o que pode ser notado pela diagonal que indica as interações na mesma mesorregião.

Em relação à dimensão espacial, foi utilizada a distância geográfica, considerando a distância (em quilômetros) em linha entre a sede dos municípios mais populosos de cada par de mesorregião a partir de suas coordenadas georreferenciadas, latitude e longitude (Tabela 3)⁴. Por fim, os cálculos para obter as distâncias, dados dois pontos na superfície terrestre a partir das coordenadas georreferenciadas, foram feitos considerando ajuste para a curvatura da Terra (CASSI; PLUNKET, 2014), como indicados na tabela 4.

⁴ As coordenadas foram obtidas em: <<https://mydasdata.larc.nasa.gov/latitudelongitude-finder/>>

Tabela 2 – Quantidade de Interações entre as mesorregiões das empresas (linhas) e mesorregiões dos grupos de pesquisa (coluna)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
1					1												6																			
2		81			1			2		3	6	2	3	8	3	12	7	1	31	14				3			3			1	6		1	4		5
3			24					1									1		2	4																3
4		3		61								1	1			1	1		5	7													3			
5		3			97			4								2			2	4				1							1	4				2
6		1				6												6	1																	
7							7	1										1																		
8		10		1	4	1		84	2	2	1	11	10	9	7	23	11	9	17	64			1	3			6				1	1	4	3		15
9		1						18	1			1					6			3																
10		1						1		46					1		4	1		2																
11		12						3			3	1	1	1		2		2	17	2											2	2	1	2		2
12										1		70	2	1	1	2	1		2	2							1				1					1
13		4			2	2		2		1		1	174	4	1	5	1	5	6	17				1	3			2			2	9	1	2	1	24
14		2	1		1		1	2		1		2		177	1	5	2	1	6	6				7								1	4	1		4
15								1		2		1		1	83	1			1	7											1					1
16		5	1					3	2				2	4		272		3	5	9			8	1			1	1					6	1		4
17	1	2			1	1				5	1	2	1	2	1		165	2	4	6																
18		5				2	2		1	1			3	1	1	2	3	141	6	5								1		2			1			1
19		72		1	4			10		2	5	13	14	31	6	24	13	14	364	70				9			12	1		1	3	12	8	44		37
20		12	2		1	1		6	1	6	1	6	17	13	8	36	17	8	29	405				1		1	1				2		5	10	1	11
21		1				2												1			1															
22												9		1						1																
23		1						1								4			1				34	2												2
24												1		9		4		3	3					74			1	1				1				2
25		1						2				1													1											1
26														1	2			1		2						8		1				1				
27		7			1			1					3	4	1	2			6	1			2				61	1				5	1	1		2
28		2			1										1				3								4	7				1				1
29												4	1	1					1											1						
30						2										1		4	1	1								1			11					3
31		1											2			3			4	2											26					6
32		1			1			4					6	1						1							3	1		1		39				10
33														2		1			3	1													39			
34		5						3					2	1	1	1			4	10											4	3		57		1
35													3				2															1			5	8
36					1					1			4		1		1		2	7														1		68

Fonte: Elaboração própria

Tabela 3 – Lista dos municípios mais populosos para as mesorregiões e respectivas coordenadas georeferenciadas

Mesorregiões	Município mais populoso	Latitude	Longitude
1-AgresPern	Caruaru	-8,2850	-35,9702
2-Campinas	Campinas	-22,9099	-47,0626
3-CentralEspSan	Vila Velha	-20,3478	-40,2950
4-CentroAmazon	Manaus	-3,1190	-60,0217
5-CentroGoiano	Goiânia	-16,6869	-49,2648
6-CentroNorBaia	Feira de Santana	-12,2597	-38,9647
7-CentroSulBaia	Vitória da Conquista	-14,8619	-40,8445
8-DF	Brasília	-15,7942	-47,8825
9-LesteAlagoia	Maceió	-9,6498	-35,7089
10-LestePotiguar	Natal	-5,7793	-35,2009
11-MacroMetroPaulis	Sorocaba	-23,5015	-47,4526
12-MetroBelem	Belém	-1,4558	-48,4902
13-MetroBH	Belo Horizonte	-19,9245	-43,9352
14-MetroCuritiba	Curitiba	-25,4290	-49,2671
15-Metrofortaleza	Fortaleza	-3,7364	-38,5184
16-MetroPorAleg	Porto Alegre	-30,0346	-51,2177
17-MetroRecife	Recife	-8,0476	-34,8770
18-MetroSalvador	Salvador	-12,9730	-38,5023
19-MetroSP	São Paulo	-23,5505	-46,6333
20-MetroRJ	Rio de Janeiro	-22,9068	-43,1729
21-NordesBaiano	Alagoinhas	-12,1340	-38,4211
22-NordesParaense	Abaetetuba	-1,7222	-48,8792
23-NoroeRioGrand	Passo Fundo	-28,2588	-52,4160
24-NortCentrParanae	Londrina	-23,3045	-51,1696
25-NorteMinas	Montes Claros	-16,7286	-43,8582
26-NorteMaraense	São Luís	-2,5391	-44,2829
27-RibeiraoPreto	Ribeirão Preto	-21,1705	-47,8102
28-SaoJoseRioPreto	São José do Rio Preto	-20,8118	-49,3762
29-SudesParaense	Marabá	-5,3811	-49,1331
30-SulBaiano	Itabuna	-14,7880	-39,2784
31-SulSudeMinas	Poços de Caldas	-21,7854	-46,5619
32-TriangMineiro	Uberlândia	-18,9146	-48,2754
33-ValeItajai	Blumenau	-26,9166	-49,0717
34-ValeParaibaPaul	São José dos Campos	-23,2237	-45,9009
35-ValerioDoce	Governador Valadares	-18,8549	-41,9559
36-ZonaMata	Juiz de Fora	-21,7624	-43,3434

Fonte: Elaboração própria

Tabela 4 – Distância Geográfica entre as mesorregiões (km)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
1	0	2011	1420	2721	1718	550	904	1540	154	291	2088	1581	1552	2370	579	2893	123	590	2042	1798	505	1605	2812	2327	1270	1120	1915	2003	1488	808	1881	1777	2489	1970	1341	1694
2	2011	0	755	2606	729	1462	1108	796	1906	2290	77	2391	464	358	2322	893	2103	1428	84	398	1507	2364	801	422	765	2285	208	334	1962	1219	135	462	489	124	696	403
3	1420	755	0	2871	1029	910	613	948	1287	1711	817	2282	383	1078	1857	1537	1487	842	744	412	935	2271	1509	1169	550	2027	787	947	1919	628	670	851	1154	661	240	353
4	2721	2606	2871	0	1912	2531	2475	1937	2781	2767	2639	1294	2557	2734	2388	3133	2835	2607	2689	2852	2580	1248	2909	2438	2325	1749	2403	2280	1233	2618	2534	2172	2891	2703	2631	2746
5	1718	729	1029	1912	0	1213	923	178	1662	1953	781	1696	668	972	1857	1497	1833	1228	811	940	1272	1665	1327	762	576	1665	521	459	1257	1089	634	269	1138	807	811	839
6	550	1462	910	2531	1213	0	354	1039	459	831	1538	1596	1004	1820	949	2346	648	94	1494	1265	61	1602	2262	1781	724	1229	1366	1460	1353	283	1331	1241	1941	1423	800	1154
7	904	1108	613	2475	923	354	0	762	804	1183	1185	1711	651	1466	1263	1993	998	329	1141	927	401	1707	1908	1434	383	1421	1017	1119	1390	169	978	910	1588	1071	459	811
8	1540	796	948	1937	178	1039	762	0	1486	1776	858	1596	621	1081	1687	1620	1655	1058	872	932	1099	1569	1462	903	442	1526	598	580	1166	930	681	349	1243	852	715	818
9	154	1906	1287	2781	1662	459	804	1486	0	434	1982	1682	1444	2264	727	2780	200	479	1933	1675	405	1702	2707	2237	1182	1234	1821	1919	1554	691	1779	1700	2378	1858	1225	1574
10	291	2290	1711	2767	1953	831	1183	1776	434	0	2367	1551	1834	2649	432	3176	255	878	2324	2088	790	1583	3089	2593	1540	1070	2185	2265	1542	1096	2159	2035	2772	2253	1628	1981
11	2088	77	817	2639	781	1538	1185	858	1982	2367	0	2454	539	282	2398	817	2179	1504	84	442	1584	2427	725	380	841	2356	262	359	2023	1294	212	517	413	161	769	464
12	1581	2391	2282	1294	1696	1596	1711	1596	1682	1551	2454	0	2113	2667	1136	3191	1676	1688	2465	2453	1625	52	3009	2447	1772	483	2193	2154	442	1795	2270	1941	2832	2437	2062	2326
13	1552	464	383	2557	668	1004	651	621	1444	1834	539	2113	0	821	1893	1342	1642	966	490	341	1047	2094	1264	837	355	1934	427	576	1712	755	342	469	937	419	239	213
14	2370	358	1078	2734	972	1820	1466	1081	2264	2649	282	2667	821	0	2671	547	2462	1786	339	679	1866	2636	443	305	1118	2601	496	514	2229	1576	490	732	167	420	1049	728
15	579	2322	1857	2388	1857	949	1263	1687	727	432	2398	1136	1893	2671	0	3215	626	1027	2369	2190	934	1172	3098	2566	1558	654	2183	2233	1191	1232	2187	1993	2811	2308	1722	2071
16	2893	893	1537	3133	1497	2346	1993	1620	2780	3176	817	3191	1342	547	3215	0	2979	2304	852	1126	2388	3158	229	748	1658	3144	1043	1042	2750	2090	1028	1272	405	923	1555	1210
17	123	2103	1487	2835	1833	648	998	1655	200	255	2179	1676	1642	2462	626	2979	0	676	2131	1875	597	1703	2904	2428	1372	1208	2014	2107	1602	890	1975	1884	2577	2057	1424	1774
18	590	1428	842	2607	1228	94	329	1058	479	878	1504	1688	966	1786	1027	2304	676	0	1455	1210	94	1694	2229	1762	711	1323	1344	1448	1440	219	1300	1236	1901	1381	751	1104
19	2042	84	744	2689	811	1494	1141	872	1933	2324	84	2465	490	339	2369	852	2131	1455	0	361	1537	2439	780	464	812	2350	291	415	2038	1243	196	543	447	83	712	392
20	1798	398	412	2852	940	1265	927	932	1675	2088	442	2453	341	679	2190	1126	1875	1210	361	0	1299	2435	1101	819	691	2268	515	681	2051	991	370	691	743	281	468	128
21	505	1507	935	2580	1272	61	401	1099	405	790	1584	1625	1047	1866	934	2388	597	94	1537	1299	0	1634	2308	1832	777	1247	1417	1514	1396	309	1378	1297	1984	1465	838	1191
22	1605	2364	2271	1248	1665	1602	1707	1569	1702	1583	2427	52	2094	2636	1172	3158	1703	1694	2439	2435	1634	0	2975	2413	1757	519	2166	2123	408	1795	2245	1913	2802	2412	2049	2308
23	2812	801	1509	2909	1327	2262	1908	1462	2707	3089	725	3009	1264	443	3098	229	2904	2229	780	1101	2308	2975	0	565	1553	2987	915	883	2567	2019	930	1121	362	859	1492	1164
24	2327	422	1169	2438	762	1781	1434	903	2237	2593	380	2447	837	305	2566	748	2428	1762	464	819	1832	2413	565	0	1057	2425	419	333	2005	1567	502	573	454	538	1076	822
25	1270	765	550	2325	576	724	383	442	1182	1540	841	1772	355	1118	1558	1658	1372	711	812	691	777	1757	1553	1057	0	1578	645	737	1386	535	630	527	1254	753	311	562
26	1120	2285	2027	1749	1665	1229	1421	1526	1234	1070	2356	483	1934	2601	654	3144	1208	1323	2350	2268	1247	519	2987	2425	1578	0	2107	2105	624	1468	2154	1872	2758	2307	1832	2140
27	1915	208	787	2403	521	1366	1017	598	1821	2185	262	2193	427	496	2183	1043	2014	1344	291	515	1417	2166	915	419	645	2107	0	167	1761	1147	146	256	652	301	664	467
28	2003	334	947	2280	459	1460	1119	580	1919	2265	359	2154	576	514	2233	1042	2107	1448	415	681	1514	2123	883	333	737	2105	167	0	1716	1261	311	240	680	447	806	634
29	1488	1962	1919	1233	1257	1353	1390	1166	1554	1542	2023	442	1712	2229	1191	2750	1602	1440	2038	2051	1396	408	2567	2005	1386	624	1761	1716	0	1502	1845	1508	2395	2014	1688	1925
30	808	1219	628	2618	1089	283	169	930	691	1096	1294	1795	755	1576	1232	2090	890	219	1243	991	309	1795	2019	1567	535	1468	1147	1261	1502	0	1093	1061	1688	1168	534	886
31	1881	135	670	2534	634	1331	978	681	1779	2159	212	2270	342	490	2187	1028	1975	1300	196	370	1378	2245	930	502	630	2154	146	311	1845	1093	0	366	625	174	580	332
32	1777	462	851	2172	269	1241	910	349	1700	2035	517	1941	469	732	1993	1272	1884	1236	543	691	1297	1913	1121	573	527	1872	256	240	1508	1061	366	0	893	539	665	604
33	2489	489	1154	2891	1138	1941	1588	1243	2378	2772	413	2832	937	167	2811	405	2577	1901	447	743	1984	2802	362	454	1254	2758	652	680	2395	1688	625	893	0	520	1155	815
34	1970	124	661	2703	807	1423	1071	852	1858	2253	161	2437	419	420	2308	923	2057	1381	83	281	1465	2412	859	538	753	2307	301	447	2014	1168	174	539	520	0	635	309
35	1341	696	240	2631	811	800	459	715	1225	1628	769	2062	239	1049	1722	1555	1424	751	712	468	838	2049	1492	1076	311	1832	664	806	1688	534	580	665	1155	635	0	354
36	1694	403	353	2746	839	1154	811	818	1574	1981	464	2326	213	728	2071	1210	1774	1104	392	128	1191	2308	1164	822	562	2140	467	634	1925	886	332	604	815	309	354	0

Fonte: Elaboração própria

Como primeiro uso de técnicas na perspectiva geométrica organizadas na classificação (quadro 9), apresenta-se a aplicação do EMD, por meio do qual representam-se as distâncias geográficas calculadas em um mapa bidimensional (Gráfico 3), a partir das coordenadas que reflitam essas distâncias nos dois eixos, como indicado na seção 3.5.1. A utilização do EMD busca alcançar a inspeção visual e a exploração dos dados, de modo a conhecer a sua estrutura quanto à descoberta das dimensões subjacentes à similaridade apresentada.

Nesse caso, considerando as distâncias geográficas entre os municípios mais populosos de cada mesorregião, a representação reflete as distâncias do mapa político do Brasil. Por exemplo, a localização das mesorregiões – mesma ordem numérica apresentada na tabela 1 – são: Centro Amazonense (R4), Metropolitana de Porto Alegre (R16) e Metropolitana de Recife (R17), apenas com ressalva da orientação dos eixos, que rotacionada em 90° à esquerda resultaria no sentido Norte/Sul e Leste/Oeste usual apresentado nos mapas.

Gráfico 3 – Escalonamento Multidimensional para Distância Geográfica entre as mesorregiões (km)

Fonte: Elaboração Própria

As figuras 1 e 2 apresentadas a seguir mostram as 36 entre as 137 mesorregiões segundo divisão indicada na tabela 1 no mapa do Brasil. Uma formatação foi feita a fim de que, na figura 1, a tonalidade da cor varie proporcionalmente conforme a quantidade de interações de origem nas empresas da mesorregião (o total pode ser obtido pela soma por linhas na tabela 2). Já na figura 2, a tonalidade da cor varia conforme a quantidade de interações de destino nos grupos de pesquisa nas respectivas mesorregiões (da mesma forma o total pode ser obtido pela soma por colunas na tabela 2).

Figura 1 – Mapa das Mesorregiões segundo Quantidade de Interações de origem nas empresas

Fonte: Elaboração Própria

Figura 2 – Mesorregiões segundo Quantidade de Interações de destino nos grupos de pesquisa

Fonte: Elaboração Própria

O primeiro uso do EMD com a distância geográfica ilustra como a técnica pode apoiar a representação das dimensões da proximidade. No caso apresentado, pode-se interpretar as dimensões espaciais comparando a representação com os mapas políticos existentes; a interpretação é facilitada devido à representação ser direta. Deste modo, abrem-se novas possibilidades de representação para as dimensões não espaciais da proximidade, ainda não realizadas na literatura.

Na próxima seção, a dimensão cognitiva foi mensurada de duas formas (euclidiana e cosseno) e sua representação foi explorada com uso do EMD, de modo a apoiar a sua interpretação.

4.3 Mensuração e Representação da Dimensão Cognitiva

Para o exercício de mensuração da dimensão não espacial, especificamente da dimensão cognitiva, considera-se os argumentos I e II apresentados nas seções 3.1 e 3.2, ou seja, a escolha de uma das possibilidades de índices nas três famílias de medidas apresentadas e a informação a ser utilizada como *proxy* para o cálculo da medida.

Considerando a perspectiva apresentada para medir a proximidade cognitiva, foi utilizado como informação de apoio o pessoal ocupado técnico-científico, a partir da Relação Anual de Informações Sociais do Ministério do Trabalho e Emprego (RAIS), considerando a Classificação Brasileira de Ocupações (CBO). Em específico, foram utilizadas 11 categorias, que retratam quatro grupos ocupacionais potencialmente empregados em atividades de ciência e tecnologia e de pesquisa e desenvolvimento (P&D) segundo Araújo, Cavalcante e Alves (2009):

- a) Pesquisadores: (CBO 203);
- b) Engenheiros: engenheiros mecatrônicos (CBO 202), engenheiros civis (CBO 214), engenheiros agrônomos e de pesca (CBO 222);
- c) Diretores e gerentes de P&D: diretores de P&D (CBO 1237) e gerentes de P&D (CBO 1426);
- d) Profissionais “científicos”: biotecnologistas, geneticistas, pesquisadores em metrologia e especialistas em calibrações meteorológicas (CBO 201); matemáticos, estatísticos e afins (CBO 211); profissionais de informática (CBO 212); físicos, químicos e afins (CBO 213); biólogos e afins (CBO 221).

A partir das contagens dos totais de empregados ativos nessas classificações para as unidades de análise (mesorregiões), pode-se comparar as respectivas distribuições de profissionais nas 11 classes e obter a distância cognitiva entre as mesorregiões. Como levantadas na seção 3.2, há diversas medidas que poderiam ser consideradas;

para este estudo, foram consideradas três delas, uma de cada família, para efeito de comparação. Para isso, iniciou-se com os cálculos para a euclidiana (quadro 6) e para o cosseno (quadro 7), cujo resultados são apresentados nas tabelas 5 e 6.

O índice cosseno é uma medida de similaridade, pois seu valor é maior quanto maior for o grau de similaridade entre os dois vetores, dadas as distribuições da variável de apoio utilizadas no cálculo. Assim, para possibilitar a comparação direta com a euclidiana e posterior uso de uma distância cosseno (d), o índice cosseno usual foi transformado em distância a partir do cálculo de: $d = 1 - s$, conforme uma das formas apresentadas na seção 3.1.

Tabela 5 – Distância Cognitiva (Euclidiana) entre as mesorregiões

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
1	0,00	0,34	0,29	0,27	0,30	0,20	0,29	0,29	0,21	0,35	0,21	0,21	0,42	0,24	0,17	0,25	0,26	0,29	0,26	0,27	0,23	0,26	0,30	0,27	0,24	0,27	0,27	0,26	0,28	0,29	0,26	0,34	0,18	0,32	0,30	0,34
2	0,34	0,00	0,17	0,18	0,09	0,26	0,30	0,19	0,19	0,08	0,28	0,18	0,11	0,30	0,21	0,16	0,24	0,12	0,26	0,41	0,31	0,23	0,09	0,29	0,18	0,14	0,24	0,18	0,38	0,34	0,17	0,44	0,21	0,08	0,39	0,44
3	0,29	0,17	0,00	0,21	0,13	0,19	0,24	0,16	0,16	0,17	0,22	0,13	0,19	0,23	0,16	0,16	0,21	0,10	0,19	0,28	0,24	0,16	0,15	0,22	0,12	0,11	0,18	0,09	0,32	0,23	0,14	0,36	0,16	0,14	0,30	0,35
4	0,27	0,18	0,21	0,00	0,21	0,28	0,35	0,27	0,24	0,15	0,29	0,23	0,25	0,30	0,18	0,12	0,17	0,22	0,30	0,39	0,32	0,29	0,19	0,34	0,23	0,17	0,30	0,24	0,23	0,37	0,20	0,47	0,15	0,23	0,42	0,47
5	0,30	0,09	0,13	0,21	0,00	0,17	0,22	0,10	0,13	0,13	0,21	0,12	0,16	0,23	0,16	0,15	0,21	0,05	0,18	0,33	0,23	0,14	0,08	0,21	0,10	0,10	0,16	0,10	0,37	0,25	0,10	0,36	0,18	0,05	0,31	0,35
6	0,20	0,26	0,19	0,28	0,17	0,00	0,10	0,11	0,09	0,27	0,06	0,11	0,33	0,10	0,12	0,21	0,21	0,15	0,08	0,18	0,08	0,07	0,22	0,08	0,09	0,16	0,08	0,11	0,35	0,13	0,13	0,21	0,17	0,20	0,16	0,20
7	0,29	0,30	0,24	0,35	0,22	0,10	0,00	0,12	0,15	0,32	0,08	0,18	0,37	0,08	0,18	0,27	0,24	0,19	0,06	0,17	0,07	0,11	0,28	0,04	0,15	0,20	0,06	0,16	0,43	0,09	0,16	0,16	0,24	0,24	0,11	0,14
8	0,29	0,19	0,16	0,27	0,10	0,11	0,12	0,00	0,09	0,21	0,13	0,11	0,25	0,15	0,15	0,20	0,21	0,08	0,09	0,25	0,15	0,07	0,17	0,11	0,07	0,12	0,06	0,08	0,40	0,17	0,08	0,27	0,20	0,13	0,22	0,25
9	0,21	0,19	0,16	0,24	0,13	0,09	0,15	0,09	0,00	0,22	0,11	0,05	0,27	0,14	0,09	0,19	0,20	0,11	0,11	0,24	0,14	0,09	0,18	0,13	0,08	0,13	0,10	0,09	0,37	0,18	0,11	0,27	0,15	0,14	0,22	0,26
10	0,35	0,08	0,17	0,15	0,13	0,27	0,32	0,21	0,22	0,00	0,29	0,22	0,13	0,30	0,21	0,12	0,20	0,15	0,27	0,40	0,32	0,25	0,11	0,31	0,19	0,13	0,26	0,19	0,33	0,35	0,16	0,46	0,20	0,14	0,41	0,45
11	0,21	0,28	0,22	0,29	0,21	0,06	0,08	0,13	0,11	0,29	0,00	0,14	0,36	0,04	0,12	0,23	0,19	0,19	0,07	0,16	0,04	0,11	0,26	0,08	0,12	0,17	0,08	0,15	0,37	0,12	0,13	0,19	0,19	0,24	0,14	0,18
12	0,21	0,18	0,13	0,23	0,12	0,11	0,18	0,11	0,05	0,22	0,14	0,00	0,25	0,17	0,10	0,18	0,21	0,10	0,13	0,25	0,16	0,10	0,16	0,15	0,09	0,13	0,12	0,08	0,35	0,18	0,12	0,29	0,14	0,14	0,23	0,28
13	0,42	0,11	0,19	0,25	0,16	0,33	0,37	0,25	0,27	0,13	0,36	0,25	0,00	0,38	0,29	0,23	0,31	0,18	0,33	0,46	0,38	0,29	0,15	0,36	0,25	0,21	0,31	0,23	0,42	0,39	0,24	0,51	0,28	0,14	0,45	0,50
14	0,24	0,30	0,23	0,30	0,23	0,10	0,08	0,15	0,14	0,30	0,04	0,17	0,38	0,00	0,14	0,24	0,18	0,21	0,07	0,14	0,04	0,14	0,28	0,09	0,15	0,18	0,09	0,17	0,37	0,12	0,14	0,18	0,21	0,26	0,14	0,17
15	0,17	0,21	0,16	0,18	0,16	0,12	0,18	0,15	0,09	0,21	0,12	0,10	0,29	0,14	0,00	0,15	0,13	0,15	0,14	0,23	0,15	0,15	0,19	0,17	0,11	0,11	0,14	0,13	0,29	0,20	0,11	0,30	0,10	0,19	0,25	0,29
16	0,25	0,16	0,16	0,12	0,15	0,21	0,27	0,20	0,19	0,12	0,23	0,18	0,23	0,24	0,15	0,00	0,14	0,16	0,23	0,33	0,26	0,21	0,12	0,27	0,15	0,12	0,22	0,16	0,23	0,30	0,13	0,40	0,12	0,17	0,36	0,40
17	0,26	0,24	0,21	0,17	0,21	0,21	0,24	0,21	0,20	0,20	0,19	0,21	0,31	0,18	0,13	0,14	0,00	0,21	0,20	0,27	0,21	0,23	0,24	0,24	0,18	0,12	0,20	0,20	0,26	0,26	0,13	0,35	0,16	0,25	0,31	0,35
18	0,29	0,12	0,10	0,22	0,05	0,15	0,19	0,08	0,11	0,15	0,19	0,10	0,18	0,21	0,15	0,16	0,21	0,00	0,15	0,30	0,21	0,11	0,11	0,18	0,08	0,10	0,13	0,06	0,37	0,22	0,09	0,33	0,17	0,07	0,28	0,32
19	0,26	0,26	0,19	0,30	0,18	0,08	0,06	0,09	0,11	0,27	0,07	0,13	0,33	0,07	0,14	0,23	0,20	0,15	0,00	0,17	0,07	0,08	0,25	0,04	0,10	0,15	0,03	0,12	0,39	0,09	0,12	0,19	0,20	0,21	0,14	0,18
20	0,27	0,41	0,28	0,39	0,33	0,18	0,17	0,25	0,24	0,40	0,16	0,25	0,46	0,14	0,23	0,33	0,27	0,30	0,17	0,00	0,14	0,21	0,38	0,16	0,24	0,28	0,19	0,25	0,39	0,12	0,26	0,15	0,28	0,36	0,13	0,15
21	0,23	0,31	0,24	0,32	0,23	0,08	0,07	0,15	0,14	0,32	0,04	0,16	0,38	0,04	0,15	0,26	0,21	0,21	0,07	0,14	0,00	0,12	0,28	0,06	0,14	0,20	0,09	0,17	0,38	0,10	0,16	0,16	0,21	0,26	0,11	0,15
22	0,26	0,23	0,16	0,29	0,14	0,07	0,11	0,07	0,09	0,25	0,11	0,10	0,29	0,14	0,15	0,21	0,23	0,11	0,08	0,21	0,12	0,00	0,20	0,09	0,08	0,15	0,06	0,08	0,39	0,13	0,12	0,23	0,19	0,16	0,18	0,22
23	0,30	0,09	0,15	0,19	0,08	0,22	0,28	0,17	0,18	0,11	0,26	0,16	0,15	0,28	0,19	0,12	0,24	0,11	0,25	0,38	0,28	0,20	0,00	0,27	0,15	0,15	0,23	0,15	0,34	0,31	0,15	0,42	0,17	0,08	0,37	0,41
24	0,27	0,29	0,22	0,34	0,21	0,08	0,04	0,11	0,13	0,31	0,08	0,15	0,36	0,09	0,17	0,27	0,24	0,18	0,04	0,16	0,06	0,09	0,27	0,00	0,13	0,19	0,05	0,14	0,42	0,07	0,15	0,16	0,23	0,23	0,10	0,14
25	0,24	0,18	0,12	0,23	0,10	0,09	0,15	0,07	0,08	0,19	0,12	0,09	0,25	0,15	0,11	0,15	0,18	0,08	0,10	0,24	0,14	0,08	0,15	0,13	0,00	0,09	0,09	0,05	0,33	0,17	0,06	0,28	0,14	0,13	0,23	0,27
26	0,27	0,14	0,11	0,17	0,10	0,16	0,20	0,12	0,13	0,13	0,17	0,13	0,21	0,18	0,11	0,12	0,12	0,10	0,15	0,28	0,20	0,15	0,15	0,19	0,09	0,00	0,14	0,10	0,31	0,23	0,05	0,33	0,14	0,14	0,29	0,33
27	0,27	0,24	0,18	0,30	0,16	0,08	0,06	0,06	0,10	0,26	0,08	0,12	0,31	0,09	0,14	0,22	0,20	0,13	0,03	0,19	0,09	0,06	0,23	0,05	0,09	0,14	0,00	0,10	0,40	0,11	0,10	0,21	0,20	0,19	0,16	0,20
28	0,26	0,18	0,09	0,24	0,10	0,11	0,16	0,08	0,09	0,19	0,15	0,08	0,23	0,17	0,13	0,16	0,20	0,06	0,12	0,25	0,17	0,08	0,15	0,14	0,05	0,10	0,10	0,00	0,35	0,17	0,09	0,29	0,14	0,13	0,24	0,28
29	0,28	0,38	0,32	0,23	0,37	0,35	0,43	0,40	0,37	0,33	0,37	0,35	0,42	0,37	0,29	0,23	0,26	0,37	0,39	0,39	0,38	0,39	0,34	0,42	0,33	0,31	0,40	0,35	0,00	0,42	0,33	0,50	0,24	0,39	0,47	0,51
30	0,29	0,34	0,23	0,37	0,25	0,13	0,09	0,17	0,18	0,35	0,12	0,18	0,39	0,12	0,20	0,30	0,26	0,22	0,09	0,12	0,10	0,13	0,31	0,07	0,17	0,23	0,11	0,17	0,42	0,00	0,20	0,13	0,25	0,28	0,09	0,13
31	0,26	0,17	0,14	0,20	0,10	0,13	0,16	0,08	0,11	0,16	0,13	0,12	0,24	0,14	0,11	0,13	0,13	0,09	0,12	0,26	0,16	0,12	0,15	0,15	0,06	0,05	0,10	0,09	0,33	0,20	0,00	0,30	0,15	0,14	0,25	0,29
32	0,34	0,44	0,36	0,47	0,36	0,21	0,16	0,27	0,27	0,46	0,19	0,29	0,51	0,18	0,30	0,40	0,35	0,33	0,19	0,15	0,16	0,23	0,42	0,16	0,28	0,33	0,21	0,29	0,50	0,13	0,30	0,00	0,35	0,39	0,08	0,06
33	0,18	0,21	0,16	0,15	0,18	0,17	0,24	0,20	0,15	0,20	0,19	0,14	0,28	0,21	0,10	0,12	0,16	0,17	0,20	0,28	0,21	0,19	0,17	0,23	0,14	0,14	0,20	0,14	0,24	0,25	0,15	0,35	0,00	0,20	0,31	0,35
34	0,32	0,08	0,14	0,23	0,05	0,20	0,24	0,13	0,14	0,14	0,24	0,14	0,14	0,26	0,19	0,17	0,25	0,07	0,21	0,36	0,26	0,16	0,08	0,23	0,13	0,14	0,19	0,13	0,39	0,28	0,14	0,39	0,20	0,00	0,33	0,37
35	0,30	0,39	0,30	0,42	0,31	0,16	0,11	0,22	0,22	0,41	0,14	0,23	0,45	0,14	0,25	0,36	0,31	0,28	0,14	0,13	0,11	0,18	0,37	0,10	0,23	0,29	0,16	0,24	0,47	0,09	0,25	0,08	0,31	0,33	0,00	0,05
36	0,34	0,44	0,35	0,47	0,35	0,20	0,14	0,25	0,26	0,45	0,18	0,28	0,50	0,17	0,29	0,40	0,35	0,32	0,18	0,15	0,15	0,22	0,41	0,14	0,27	0,33	0,20	0,28	0,51</							

Tabela 6 – Distância Cognitiva (Cosseno) entre as mesorregiões

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
1	0,00	0,15	0,14	0,13	0,12	0,06	0,10	0,10	0,06	0,16	0,06	0,07	0,21	0,08	0,05	0,10	0,11	0,11	0,09	0,12	0,07	0,09	0,12	0,09	0,09	0,11	0,09	0,11	0,17	0,12	0,10	0,15	0,07	0,12	0,11	0,13
2	0,15	0,00	0,03	0,04	0,01	0,08	0,11	0,04	0,04	0,01	0,10	0,04	0,01	0,12	0,05	0,03	0,07	0,02	0,09	0,22	0,12	0,07	0,01	0,11	0,04	0,02	0,07	0,04	0,19	0,14	0,03	0,24	0,04	0,01	0,19	0,23
3	0,14	0,03	0,00	0,07	0,02	0,05	0,07	0,03	0,04	0,03	0,07	0,03	0,03	0,08	0,04	0,04	0,07	0,01	0,05	0,13	0,08	0,04	0,03	0,07	0,02	0,02	0,04	0,01	0,18	0,08	0,03	0,17	0,04	0,02	0,13	0,15
4	0,13	0,04	0,07	0,00	0,06	0,13	0,17	0,10	0,09	0,02	0,13	0,09	0,06	0,14	0,06	0,02	0,04	0,07	0,13	0,25	0,16	0,13	0,05	0,17	0,08	0,04	0,12	0,09	0,09	0,20	0,06	0,31	0,04	0,06	0,25	0,29
5	0,12	0,01	0,02	0,06	0,00	0,04	0,06	0,01	0,02	0,02	0,06	0,02	0,03	0,07	0,03	0,02	0,06	0,00	0,04	0,15	0,07	0,03	0,01	0,06	0,01	0,01	0,03	0,01	0,19	0,08	0,01	0,16	0,02	0,00	0,12	0,15
6	0,06	0,08	0,05	0,13	0,04	0,00	0,01	0,01	0,01	0,10	0,01	0,02	0,12	0,01	0,02	0,07	0,07	0,03	0,01	0,05	0,01	0,01	0,07	0,01	0,01	0,04	0,01	0,02	0,21	0,02	0,02	0,06	0,04	0,05	0,03	0,05
7	0,10	0,11	0,07	0,17	0,06	0,01	0,00	0,02	0,03	0,13	0,01	0,04	0,16	0,01	0,04	0,10	0,08	0,05	0,00	0,04	0,00	0,01	0,10	0,00	0,03	0,05	0,00	0,03	0,27	0,01	0,03	0,03	0,07	0,07	0,02	0,02
8	0,10	0,04	0,03	0,10	0,01	0,01	0,02	0,00	0,01	0,06	0,02	0,01	0,07	0,03	0,02	0,05	0,05	0,01	0,01	0,08	0,03	0,01	0,04	0,02	0,00	0,02	0,00	0,01	0,22	0,03	0,01	0,09	0,03	0,02	0,06	0,08
9	0,06	0,04	0,04	0,09	0,02	0,01	0,03	0,01	0,00	0,07	0,02	0,00	0,08	0,03	0,01	0,05	0,06	0,01	0,02	0,09	0,03	0,01	0,04	0,02	0,01	0,02	0,01	0,01	0,22	0,05	0,02	0,10	0,03	0,02	0,06	0,08
10	0,16	0,01	0,03	0,02	0,02	0,10	0,13	0,06	0,07	0,00	0,11	0,06	0,02	0,12	0,05	0,01	0,05	0,03	0,10	0,22	0,14	0,08	0,01	0,13	0,05	0,02	0,09	0,05	0,14	0,16	0,03	0,26	0,04	0,02	0,22	0,25
11	0,06	0,10	0,07	0,13	0,06	0,01	0,01	0,02	0,02	0,11	0,00	0,03	0,15	0,00	0,02	0,08	0,05	0,05	0,01	0,04	0,00	0,02	0,09	0,01	0,02	0,04	0,01	0,03	0,22	0,02	0,03	0,05	0,05	0,07	0,03	0,04
12	0,07	0,04	0,03	0,09	0,02	0,02	0,04	0,01	0,00	0,06	0,03	0,00	0,07	0,04	0,01	0,05	0,07	0,01	0,02	0,10	0,04	0,01	0,04	0,03	0,01	0,02	0,02	0,01	0,22	0,05	0,02	0,11	0,03	0,02	0,07	0,10
13	0,21	0,01	0,03	0,06	0,03	0,12	0,16	0,07	0,08	0,02	0,15	0,07	0,00	0,17	0,09	0,05	0,11	0,03	0,12	0,26	0,17	0,09	0,02	0,15	0,07	0,05	0,11	0,05	0,21	0,18	0,06	0,30	0,07	0,02	0,24	0,28
14	0,08	0,12	0,08	0,14	0,07	0,01	0,01	0,03	0,03	0,12	0,00	0,04	0,17	0,00	0,03	0,09	0,05	0,06	0,01	0,03	0,00	0,03	0,11	0,01	0,03	0,05	0,01	0,04	0,22	0,02	0,03	0,04	0,06	0,09	0,02	0,03
15	0,05	0,05	0,04	0,06	0,03	0,02	0,04	0,02	0,01	0,05	0,02	0,01	0,09	0,03	0,00	0,03	0,03	0,02	0,02	0,09	0,03	0,03	0,05	0,04	0,02	0,02	0,02	0,02	0,16	0,06	0,01	0,12	0,01	0,04	0,08	0,10
16	0,10	0,03	0,04	0,02	0,02	0,07	0,10	0,05	0,05	0,01	0,08	0,05	0,05	0,09	0,03	0,00	0,03	0,03	0,08	0,18	0,10	0,07	0,02	0,10	0,03	0,02	0,07	0,04	0,09	0,13	0,02	0,22	0,02	0,03	0,18	0,21
17	0,11	0,07	0,07	0,04	0,06	0,07	0,08	0,05	0,06	0,05	0,05	0,07	0,11	0,05	0,03	0,03	0,00	0,06	0,06	0,12	0,07	0,08	0,08	0,08	0,05	0,02	0,06	0,06	0,11	0,10	0,02	0,16	0,04	0,08	0,13	0,15
18	0,11	0,02	0,01	0,07	0,00	0,03	0,05	0,01	0,01	0,03	0,05	0,01	0,03	0,06	0,02	0,03	0,00	0,03	0,13	0,06	0,02	0,02	0,02	0,04	0,01	0,01	0,02	0,00	0,20	0,06	0,01	0,14	0,03	0,00	0,10	0,13
19	0,09	0,09	0,05	0,13	0,04	0,01	0,00	0,01	0,02	0,10	0,01	0,02	0,12	0,01	0,02	0,08	0,06	0,03	0,00	0,04	0,01	0,01	0,08	0,00	0,01	0,03	0,00	0,02	0,24	0,01	0,02	0,05	0,05	0,06	0,03	0,04
20	0,12	0,22	0,13	0,25	0,15	0,05	0,04	0,08	0,09	0,22	0,04	0,10	0,26	0,03	0,09	0,18	0,12	0,13	0,04	0,00	0,03	0,07	0,21	0,03	0,09	0,12	0,05	0,09	0,28	0,02	0,09	0,02	0,13	0,17	0,02	0,02
21	0,07	0,12	0,08	0,16	0,07	0,01	0,00	0,03	0,03	0,14	0,00	0,04	0,17	0,00	0,03	0,10	0,07	0,06	0,01	0,03	0,00	0,02	0,11	0,00	0,03	0,06	0,01	0,04	0,24	0,01	0,04	0,03	0,07	0,09	0,01	0,02
22	0,09	0,07	0,04	0,13	0,03	0,01	0,01	0,01	0,01	0,08	0,02	0,01	0,09	0,03	0,03	0,07	0,08	0,02	0,01	0,07	0,02	0,00	0,05	0,01	0,01	0,03	0,01	0,01	0,24	0,02	0,02	0,07	0,04	0,03	0,04	0,06
23	0,12	0,01	0,03	0,05	0,01	0,07	0,10	0,04	0,04	0,01	0,09	0,04	0,02	0,11	0,05	0,02	0,08	0,02	0,08	0,21	0,11	0,05	0,00	0,10	0,03	0,03	0,07	0,03	0,16	0,13	0,03	0,23	0,03	0,01	0,18	0,21
24	0,09	0,11	0,07	0,17	0,06	0,01	0,00	0,02	0,02	0,13	0,01	0,03	0,15	0,01	0,04	0,10	0,08	0,04	0,00	0,03	0,00	0,01	0,10	0,00	0,02	0,05	0,00	0,03	0,27	0,01	0,03	0,03	0,07	0,07	0,01	0,02
25	0,09	0,04	0,02	0,08	0,01	0,01	0,03	0,00	0,01	0,05	0,02	0,01	0,07	0,03	0,02	0,03	0,05	0,01	0,01	0,09	0,03	0,01	0,03	0,02	0,00	0,01	0,01	0,00	0,18	0,04	0,01	0,10	0,02	0,02	0,07	0,09
26	0,11	0,02	0,02	0,04	0,01	0,04	0,05	0,02	0,02	0,02	0,04	0,02	0,05	0,05	0,02	0,02	0,02	0,01	0,03	0,12	0,06	0,03	0,03	0,05	0,01	0,00	0,03	0,02	0,15	0,07	0,00	0,15	0,02	0,02	0,11	0,14
27	0,09	0,07	0,04	0,12	0,03	0,01	0,00	0,00	0,01	0,09	0,01	0,02	0,11	0,01	0,02	0,07	0,06	0,02	0,00	0,05	0,01	0,01	0,07	0,00	0,01	0,03	0,00	0,01	0,24	0,02	0,01	0,06	0,05	0,04	0,03	0,05
28	0,11	0,04	0,01	0,09	0,01	0,02	0,03	0,01	0,01	0,05	0,03	0,01	0,05	0,04	0,02	0,04	0,06	0,00	0,02	0,09	0,04	0,01	0,03	0,03	0,00	0,02	0,01	0,00	0,20	0,04	0,01	0,11	0,03	0,02	0,07	0,10
29	0,17	0,19	0,18	0,09	0,19	0,21	0,27	0,22	0,22	0,14	0,22	0,22	0,21	0,22	0,16	0,09	0,11	0,20	0,24	0,28	0,24	0,24	0,16	0,27	0,18	0,15	0,24	0,20	0,00	0,29	0,16	0,38	0,12	0,21	0,34	0,37
30	0,12	0,14	0,08	0,20	0,08	0,02	0,01	0,03	0,05	0,16	0,02	0,05	0,18	0,02	0,06	0,13	0,10	0,06	0,01	0,02	0,01	0,02	0,13	0,01	0,04	0,07	0,02	0,04	0,29	0,00	0,05	0,02	0,09	0,10	0,01	0,02
31	0,10	0,03	0,03	0,06	0,01	0,02	0,03	0,01	0,02	0,03	0,03	0,02	0,06	0,03	0,01	0,02	0,02	0,01	0,02	0,09	0,04	0,02	0,03	0,03	0,01	0,00	0,01	0,01	0,16	0,05	0,00	0,12	0,02	0,02	0,08	0,11
32	0,15	0,24	0,17	0,31	0,16	0,06	0,03	0,09	0,10	0,26	0,05	0,11	0,30	0,04	0,12	0,22	0,16	0,14	0,05	0,02	0,03	0,07	0,23	0,03	0,10	0,15	0,06	0,11	0,38	0,02	0,12	0,00	0,17	0,18	0,01	0,00
33	0,07	0,04	0,04	0,04	0,02	0,04	0,07	0,03	0,03	0,04	0,05	0,03	0,07	0,06	0,01	0,02	0,04	0,03	0,05	0,13	0,07	0,04	0,03	0,07	0,02	0,02	0,05	0,03	0,12	0,09	0,02	0,17	0,00	0,03	0,13	0,15
34	0,12	0,01	0,02	0,06	0,00	0,05	0,07	0,02	0,02	0,02	0,07	0,02	0,02	0,09	0,04	0,03	0,08	0,00	0,06	0,17	0,09	0,03	0,01	0,07	0,02	0,02	0,04	0,02	0,21	0,10	0,02	0,18	0,03	0,00	0,14	0,17
35	0,11	0,19	0,13	0,25	0,12	0,03	0,02	0,06	0,06	0,22	0,03	0,07	0,24	0,02	0,08	0,18	0,13	0,10	0,03	0,02	0,01	0,04	0,18	0,01	0,07	0,11	0,03	0,07	0,34	0,01	0,08	0,01	0,13	0,14	0,00	0,00
36	0,13	0,23	0,15	0,29	0,15	0,05	0,02	0,08	0,08	0,25	0,04	0,10	0,28	0,03	0,10	0,21	0,15	0,13	0,04	0,02	0,02	0,06	0,21	0,02	0,09	0,14	0,05	0,10	0,37</							

Para facilitar a visualização e a comparação relativa dos valores nas tabelas 5 e 6, foi aplicada uma formatação nas tonalidades da cor nas caselas, (célula que representa cada intersecção entre mesorregiões nas linhas e colunas) de modo que distâncias maiores entre duas mesorregiões são representadas com tom mais escuro. A princípio, embora as diferenças entre as duas distâncias nessas tabelas pareçam sutis, a tarefa de comparação carece de melhores instrumentos.

A distância euclidiana possui valores mais homogêneos e amplitude maior quando comparada à distância cosseno, haja vista que os valores máximos para essas distâncias são 0,512 e 0,381, respectivamente. Porém, essa análise requer cautela e outras técnicas que podem ser mais adequadas para apoiar essa reflexão. A partir das tabelas, nota-se, por exemplo, valores maiores para as mesorregiões 1 (Agreste Pernambucano) e 29 (Sudeste Paraense), o que pode indicar que essas mesorregiões apresentam peculiaridades quanto à variável utilizada para a definição das medidas de distância cognitiva; em outra instância, pode indicar a existência de pontos discrepantes que distorceriam a estimação de modelos estatísticos.

Para apoiar essa análise comparativa, utiliza-se o EMD, pois, como realizado com a distância geográfica, pode ajudar a melhor representar as medidas, assumindo que as dissimilaridades são distâncias e encontrando as coordenadas que as expliquem ou que melhor se ajustem. Assim, as distâncias entre pares de mesorregiões serão representadas entre pontos em um espaço multidimensional de mesma dimensão, e por vezes de dimensão menor, o que facilita a comparação entre as diferentes medidas.

Os mapas para as Distância Cognitiva (Euclidiana) e Distância Cognitiva (Cosseno), apresentados respectivamente nos gráficos 4 e 5, direcionam a uma interpretação específica para a distância cognitiva, valendo-se da sua representação gráfica e das variáveis definidas para sua composição.

Gráfico 4 – Escalonamento Multidimensional para Distância Cognitiva (Euclidiana) entre as mesorregiões

Fonte: Elaboração Própria

Gráfico 5 – Escalonamento Multidimensional para Distância Cognitiva (Cosseno) entre as mesorregiões

Fonte: Elaboração Própria

Devido aos indícios a partir da análise inicial das tabelas serem facilitados pela representação geométrica, o uso de EMD explicita a existência de mesorregiões que são mais distantes cognitivamente (mesorregiões 1 e 29) das demais, em especial quanto à distância cognitiva calculada pelo índice cosseno. Embora as dimensões não possuam significado direto, com uso da técnica EMD há a possibilidade de interpretação das dimensões, visto que elas podem fornecer informações e até conjecturas a partir da disposição e da natureza das mesorregiões.

O exercício para interpretar a dimensão 1 (eixo horizontal nos gráficos 4 e 5) pode ser feito observando a oposição entre as mesorregiões representadas pelos pontos R32, R36, R35 e R20 em relação às mesorregiões R4, R10, R13 e R29. Essa oposição em relação ao eixo horizontal pode dar intuição para inferir sobre os critérios que guiaram essa disposição dos pontos (mesorregiões) em relação a dada dimensão. A oposição entre as mesorregiões é direcionada dada as características utilizadas no cálculo das

distâncias. Como foram considerados grupos ocupacionais potencialmente empregados em atividades de ciência e tecnologia e de P&D, é natural que essa informação seja utilizada para caracterizar a dimensão 1.

Para as regiões destacadas e nomeadas, como: 4 (Centro Amazonense), 10 (Leste Potiguar), 13 (Metropolitana de Belo Horizonte) e 29 (Sudeste Paraense), em relação às mesorregiões em oposição 32 (Triângulo Mineiro/Alto Paranaíba), 36 (Zona da Mata), 35 (Vale do Rio Doce) e 20 (Metropolitana do Rio de Janeiro), apresenta-se a oposição pela representação relativa de dois grupos em especial: os engenheiros, em maior quantidade relativa para o primeiro grupo de mesorregiões, e os profissionais de informática, em maior quantidade no segundo grupo (gráfico 6).

Gráfico 6 – Distribuição dos grupos ocupacionais potencialmente empregados em atividades de ciência e tecnologia e de pesquisa e desenvolvimento (P&D) para mesorregiões selecionadas

Fonte: Elaboração Própria

A diferença relativa entre as categorias mostra que, com uso da EMD, pode-se dar significado exploratório para a distância cognitiva representada pelas distâncias nos gráficos segundo a informação de interesse utilizada na construção da medida. Além da EMD, o quadro proposto ainda apresenta, na perspectiva geométrica, a AC, que

também pode ser utilizada como técnica analítica para medição e representação das proximidades, aplicação que foi feita e apresentada na próxima seção.

4.4 Proposta de medição e representação da distância cognitiva com uso da Análise de Correspondência

A seção anterior explorou o uso da perspectiva geométrica da classificação proposta com uso das distâncias euclidiana e cosseno. Esta seção propõe uma forma de medição e de representação de uma distância cognitiva denominada distância cognitiva (qui-quadrado), cuja medida foi classificada como família 3, conforme apontada na seção 3.1, e obtida a partir das mesmas distribuições de profissionais nas 11 classes, a fim de ser comparada com as duas já apresentadas.

Obteve-se a forma da distância cognitiva (qui-quadrado) a partir do uso da técnica de AC, apresentada na seção 3.5.2. De forma objetiva, a AC procura uma representação gráfica das linhas de uma tabela de contingências, para que essas linhas possam ser representadas como pontos em um espaço de menor dimensão, representado pelas colunas. Para este trabalho, as linhas são as mesorregiões e as colunas, os grupos ocupacionais, de tal modo que as distâncias entre esses pontos refletem a semelhança de "perfil" entre as mesorregiões quanto às ocupações.

A mesma análise pode ser feita do ponto de vista da representação do perfil das colunas no espaço das linhas, porém, não é o escopo deste estudo, assim como não há interesse em verificar a associação entre as variáveis de linhas e colunas, análises possíveis com uso da AC.

O modo como foi desenvolvida a aplicação da AC possibilita a obtenção das distâncias, o que facilita a mensuração das dimensões não espaciais da proximidade. Além disso, a técnica também facilita a representação da distância, pois há a redução de dimensão intrínseca com o seu uso. Por exemplo, em relação a esta pesquisa, embora haja 11 dimensões (categorias ocupacionais) iniciais na tabela de contingência, a AC representará os pontos (mesorregiões) em um espaço reduzido.

Assim, com a escolha das duas primeiras dimensões, explica-se 71,18% da informação (medida pela inércia), ou seja, a representação das mesorregiões em 2 dimensões, ao invés das 11 dimensões iniciais, mantém boa parte da associação da tabela de contingência em relação à completa independência entre os perfis das mesorregiões, segundo a distribuição dos grupos ocupacionais que representam a dimensão cognitiva. O Apêndice A apresenta os resultados gerais da AC e, em especial, as coordenadas para as 36 mesorregiões, nomeadas como R1, R2, ..., R36, indicadas na segunda parte do quadro 11, intitulada "*Statistics for row and column categories in row norm. (x 1000)*".

O uso da decomposição da inércia total em eixos ordenados (dimensões ortogonais) pela sua importância resulta em uma representação gráfica dessas coordenadas no plano euclidiano e possibilita o cálculo da Distância Cognitiva (Qui-Quadrado). Por construção (uso de coordenadas padronizadas), a distância euclidiana entre os pontos representados no plano aproxima-se da distância qui-quadrado entre os perfis de linha correspondentes (mesorregiões). Deste modo, a tabela 7 apresenta as distâncias calculadas a partir das coordenadas obtidas com a AC.

Tabela 7 – Distância Cognitiva (Qui-Quadrado) entre as mesorregiões

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
1	0,00	0,37	0,05	0,16	0,06	0,23	0,73	0,46	0,33	0,27	0,11	0,30	0,05	0,19	0,26	0,22	0,09	0,11	0,47	0,35	0,90	0,93	0,81	0,36	0,69	0,47	0,21	0,24	0,64	0,48	0,31	0,42	0,32	0,35	0,24	0,36
2	0,37	0,00	0,32	0,47	0,38	0,59	0,99	0,19	0,68	0,61	0,42	0,62	0,39	0,27	0,13	0,15	0,39	0,47	0,10	0,39	1,21	1,24	1,05	0,40	0,97	0,79	0,53	0,51	1,00	0,77	0,57	0,53	0,06	0,64	0,59	0,52
3	0,05	0,32	0,00	0,20	0,07	0,28	0,75	0,41	0,37	0,32	0,15	0,33	0,09	0,15	0,21	0,17	0,12	0,16	0,42	0,34	0,93	0,96	0,83	0,33	0,72	0,50	0,24	0,25	0,69	0,50	0,32	0,41	0,27	0,39	0,29	0,35
4	0,16	0,47	0,20	0,00	0,21	0,24	0,81	0,60	0,35	0,15	0,06	0,38	0,11	0,35	0,39	0,33	0,09	0,11	0,57	0,29	0,94	0,97	0,90	0,52	0,76	0,51	0,31	0,36	0,60	0,55	0,42	0,57	0,43	0,19	0,14	0,50
5	0,06	0,38	0,07	0,21	0,00	0,22	0,68	0,45	0,30	0,30	0,17	0,26	0,11	0,16	0,26	0,24	0,15	0,13	0,47	0,40	0,86	0,89	0,76	0,31	0,64	0,43	0,16	0,18	0,62	0,43	0,26	0,36	0,32	0,40	0,26	0,30
6	0,23	0,59	0,28	0,24	0,22	0,00	0,58	0,66	0,11	0,19	0,25	0,17	0,23	0,37	0,47	0,45	0,26	0,15	0,69	0,52	0,70	0,73	0,67	0,46	0,52	0,27	0,14	0,21	0,41	0,33	0,24	0,45	0,54	0,34	0,16	0,37
7	0,73	0,99	0,75	0,81	0,68	0,58	0,00	0,93	0,48	0,76	0,80	0,44	0,77	0,73	0,85	0,88	0,80	0,70	1,06	1,08	0,28	0,31	0,09	0,62	0,09	0,33	0,52	0,50	0,53	0,26	0,43	0,49	0,93	0,91	0,73	0,47
8	0,46	0,19	0,41	0,60	0,45	0,66	0,93	0,00	0,72	0,73	0,55	0,63	0,50	0,29	0,21	0,28	0,52	0,57	0,19	0,57	1,18	1,21	0,98	0,31	0,94	0,80	0,56	0,52	1,05	0,75	0,56	0,45	0,20	0,79	0,70	0,47
9	0,33	0,68	0,37	0,35	0,30	0,11	0,48	0,72	0,00	0,28	0,36	0,12	0,34	0,43	0,55	0,54	0,36	0,26	0,77	0,63	0,59	0,62	0,57	0,48	0,42	0,16	0,17	0,22	0,33	0,23	0,22	0,44	0,62	0,43	0,25	0,36
10	0,27	0,61	0,32	0,15	0,30	0,19	0,76	0,73	0,28	0,00	0,20	0,35	0,23	0,45	0,52	0,47	0,22	0,17	0,71	0,42	0,85	0,87	0,85	0,59	0,69	0,44	0,31	0,38	0,47	0,51	0,43	0,62	0,57	0,15	0,04	0,53
11	0,11	0,42	0,15	0,06	0,17	0,25	0,80	0,55	0,36	0,20	0,00	0,37	0,06	0,29	0,33	0,28	0,03	0,10	0,51	0,28	0,95	0,98	0,89	0,47	0,75	0,51	0,29	0,33	0,63	0,55	0,40	0,53	0,38	0,25	0,18	0,46
12	0,30	0,62	0,33	0,38	0,26	0,17	0,44	0,63	0,12	0,35	0,37	0,00	0,33	0,36	0,48	0,48	0,36	0,27	0,71	0,64	0,60	0,63	0,52	0,37	0,39	0,18	0,09	0,12	0,43	0,18	0,10	0,32	0,56	0,51	0,32	0,23
13	0,05	0,39	0,09	0,11	0,11	0,23	0,77	0,50	0,34	0,23	0,06	0,33	0,00	0,24	0,29	0,24	0,04	0,09	0,48	0,31	0,92	0,96	0,85	0,41	0,72	0,49	0,25	0,28	0,64	0,51	0,35	0,48	0,34	0,30	0,21	0,41
14	0,19	0,27	0,15	0,35	0,16	0,37	0,73	0,29	0,43	0,45	0,29	0,36	0,24	0,00	0,13	0,16	0,26	0,29	0,35	0,45	0,94	0,97	0,79	0,20	0,71	0,53	0,27	0,24	0,77	0,50	0,31	0,31	0,21	0,54	0,42	0,27
15	0,26	0,13	0,21	0,39	0,26	0,47	0,85	0,21	0,55	0,52	0,33	0,48	0,29	0,13	0,00	0,07	0,30	0,36	0,22	0,39	1,07	1,10	0,92	0,28	0,84	0,66	0,39	0,37	0,88	0,63	0,44	0,41	0,08	0,57	0,49	0,40
16	0,22	0,15	0,17	0,33	0,24	0,45	0,88	0,28	0,54	0,47	0,28	0,48	0,24	0,16	0,07	0,00	0,25	0,32	0,24	0,32	1,08	1,11	0,95	0,34	0,86	0,66	0,39	0,38	0,86	0,65	0,45	0,46	0,10	0,51	0,45	0,43
17	0,09	0,39	0,12	0,09	0,15	0,26	0,80	0,52	0,36	0,22	0,03	0,36	0,04	0,26	0,30	0,25	0,00	0,11	0,49	0,28	0,95	0,98	0,88	0,44	0,75	0,52	0,28	0,32	0,65	0,54	0,39	0,51	0,35	0,28	0,20	0,45
18	0,11	0,47	0,16	0,11	0,13	0,15	0,70	0,57	0,26	0,17	0,10	0,27	0,09	0,29	0,36	0,32	0,11	0,00	0,57	0,38	0,85	0,88	0,79	0,43	0,65	0,41	0,19	0,25	0,55	0,45	0,31	0,47	0,42	0,28	0,14	0,39
19	0,47	0,10	0,42	0,57	0,47	0,69	1,06	0,19	0,77	0,71	0,51	0,71	0,48	0,35	0,22	0,24	0,49	0,57	0,00	0,45	1,29	1,32	1,12	0,46	1,05	0,88	0,62	0,60	1,10	0,85	0,66	0,59	0,15	0,73	0,69	0,60
20	0,35	0,39	0,34	0,29	0,40	0,52	1,08	0,57	0,63	0,42	0,28	0,64	0,31	0,45	0,39	0,32	0,28	0,38	0,45	0,00	1,22	1,25	1,16	0,64	1,03	0,79	0,55	0,58	0,89	0,82	0,66	0,74	0,38	0,35	0,42	0,69
21	0,90	1,21	0,93	0,94	0,86	0,70	0,28	1,18	0,59	0,85	0,95	0,60	0,92	0,94	1,07	1,08	0,95	0,85	1,29	1,22	0,00	0,03	0,30	0,87	0,25	0,43	0,70	0,70	0,47	0,44	0,63	0,75	1,15	0,99	0,83	0,71
22	0,93	1,24	0,96	0,97	0,89	0,73	0,31	1,21	0,62	0,87	0,98	0,63	0,96	0,97	1,10	1,11	0,98	0,88	1,32	1,25	0,03	0,00	0,32	0,90	0,28	0,47	0,73	0,73	0,49	0,47	0,66	0,78	1,18	1,02	0,86	0,74
23	0,81	1,05	0,83	0,90	0,76	0,67	0,09	0,98	0,57	0,85	0,89	0,52	0,85	0,79	0,92	0,95	0,88	0,79	1,12	1,16	0,30	0,32	0,00	0,67	0,17	0,42	0,60	0,58	0,62	0,35	0,50	0,54	0,99	1,01	0,82	0,52
24	0,36	0,40	0,33	0,52	0,31	0,46	0,62	0,31	0,48	0,59	0,47	0,37	0,41	0,20	0,28	0,34	0,44	0,43	0,46	0,64	0,87	0,90	0,67	0,00	0,63	0,52	0,32	0,26	0,80	0,45	0,28	0,14	0,35	0,71	0,56	0,16
25	0,69	0,97	0,72	0,76	0,64	0,52	0,09	0,94	0,42	0,69	0,75	0,39	0,72	0,71	0,84	0,86	0,75	0,65	1,05	1,03	0,25	0,28	0,17	0,63	0,00	0,26	0,48	0,47	0,45	0,21	0,40	0,50	0,92	0,85	0,67	0,46
26	0,47	0,79	0,50	0,51	0,43	0,27	0,33	0,80	0,16	0,44	0,51	0,18	0,49	0,53	0,66	0,66	0,52	0,41	0,88	0,79	0,43	0,47	0,42	0,52	0,26	0,00	0,27	0,29	0,30	0,12	0,24	0,44	0,74	0,59	0,41	0,37
27	0,21	0,53	0,24	0,31	0,16	0,14	0,52	0,56	0,17	0,31	0,29	0,09	0,25	0,27	0,39	0,39	0,28	0,19	0,62	0,55	0,70	0,73	0,60	0,32	0,48	0,27	0,00	0,07	0,50	0,27	0,12	0,31	0,47	0,46	0,28	0,22
28	0,24	0,51	0,25	0,36	0,18	0,21	0,50	0,52	0,22	0,38	0,33	0,12	0,28	0,24	0,37	0,38	0,32	0,25	0,60	0,58	0,70	0,73	0,58	0,26	0,47	0,29	0,07	0,00	0,55	0,26	0,07	0,24	0,45	0,52	0,34	0,16
29	0,64	1,00	0,69	0,60	0,62	0,41	0,53	1,05	0,33	0,47	0,63	0,43	0,64	0,77	0,88	0,86	0,65	0,55	1,10	0,89	0,47	0,49	0,62	0,80	0,45	0,30	0,50	0,55	0,00	0,41	0,52	0,74	0,95	0,58	0,47	0,66
30	0,48	0,77	0,50	0,55	0,43	0,33	0,26	0,75	0,23	0,51	0,55	0,18	0,51	0,50	0,63	0,65	0,54	0,45	0,85	0,82	0,44	0,47	0,35	0,45	0,21	0,12	0,27	0,26	0,41	0,00	0,19	0,35	0,71	0,66	0,48	0,29
31	0,31	0,57	0,32	0,42	0,26	0,24	0,43	0,56	0,22	0,43	0,40	0,10	0,35	0,31	0,44	0,45	0,39	0,31	0,66	0,66	0,63	0,66	0,50	0,28	0,40	0,24	0,12	0,07	0,52	0,19	0,00	0,23	0,52	0,57	0,39	0,14
32	0,42	0,53	0,41	0,57	0,36	0,45	0,49	0,45	0,44	0,62	0,53	0,32	0,48	0,31	0,41	0,46	0,51	0,47	0,59	0,74	0,75	0,78	0,54	0,14	0,50	0,44	0,31	0,24	0,74	0,35	0,23	0,00	0,48	0,75	0,58	0,09
33	0,32	0,06	0,27	0,43	0,32	0,54	0,93	0,20	0,62	0,57	0,38	0,56	0,34	0,21	0,08	0,10	0,35	0,42	0,15	0,38	1,15	1,18	0,99	0,35	0,92	0,74	0,47	0,45	0,95	0,71	0,52	0,48	0,00	0,61	0,55	0,47
34	0,35	0,64	0,39	0,19	0,40	0,34	0,91	0,79	0,43	0,15	0,25	0,51	0,30	0,54	0,57	0,51	0,28	0,28	0,73	0,35	0,99	1,02	1,01	0,71	0,85	0,59	0,46	0,52	0,58	0,66	0,57	0,75	0,61	0,00	0,19	0,67
35	0,24	0,59	0,29	0,14	0,26	0,16	0,73	0,70	0,25	0,04	0,18	0,32	0,21	0,42	0,49	0,45	0,20	0,14	0,69	0,42	0,83	0,86	0,82	0,56	0,67	0,41	0,28	0,34	0,47	0,48	0,39	0,58	0,55	0,19	0,00	0,50
36	0,36	0,52	0,35	0,50	0,30	0,37	0,47	0,47	0,36	0,53	0,46	0,23	0,41	0,27	0,40	0,43	0,45	0,39	0,60	0,69	0,71	0,74	0,52	0,16	0,46	0,37	0,22	0,16	0,6							

As vantagens do uso da AC estão relacionadas com a obtenção das coordenadas para as mesorregiões e para as categorias da variável de interesse, a qual especifica a informação utilizada para caracterizar a distância entre as mesorregiões. Assim, por construção, os perfis de linhas (das mesorregiões) são representados no espaço das “colunas”, nesse caso espaço dos grupos ocupacionais, indicado por meio dos pontos representados pelos triângulos (gráfico 7). Em outras palavras, a variável utilizada na coluna servirá de apoio para a interpretação da distância cognitiva, dado o uso do “mapa assimétrico”, com coordenadas “principais” para as linhas (mesorregiões) e com coordenadas “padrão” para as colunas – conforme a terminologia estatística.

A possibilidade de interpretação das dimensões de acordo com a característica definida permite elaborar conjecturas mais diretas. Por exemplo, para a dimensão 1, há, a partir da projeção das classes, a oposição entre diretores (CBO 1237), gerentes (CBO 1426), profissionais de informática (CBO 212), matemáticos, estatísticos e afins (CBO 211) em relação aos engenheiros e outros profissionais científicos (físicos, químicos e biólogos). Já a dimensão 2 mostra a nítida oposição entre engenheiros agrônomos (CBO 222) e biotecnologistas (CBO 201), o que caracterizaria as dimensões e consequentemente a distância cognitiva facilitando a sua interpretação.

Caracterizando a dimensão 1 como uma escala para as mesorregiões, os perfis na esquerda do gráfico representam-se como mais voltado para gestão se comparados aos da direita, mais técnicos. Já a dimensão 2 indica uma escala, por exemplo, com perfis mais focados em laboratório abaixo (biotecnologia, alimentícia, farmacêutica e similares) para perfis caracterizados pelo campo acima (relação com a agroindústria e similares).

Gráfico 7 – Análise de Correspondência para obtenção da Distância Cognitiva (Qui-Quadrado) entre as mesorregiões – coordenadas “principais” para as linhas (mesorregiões) e as coordenadas “padrão” para as colunas

Fonte: Elaboração Própria

Dadas as distâncias entre as mesorregiões, representadas pelos espaços entre os pontos no gráfico 7, observa-se que elas estarão mais próximas de acordo com a maior semelhança entre os seus perfis, segundo as distribuições da variável de interesse escolhida (no caso, os grupos ocupacionais). Além disso, a distância euclidiana entre os pontos que representam as mesorregiões e a origem do gráfico (centroide) aproxima-se da distância qui-quadrado entre seus respectivos perfis e o perfil geral da população (perfil sob independência).

No gráfico 8, foi suprimida a representação dos grupos ocupacionais (pontos representados pelos triângulos no gráfico 7) para melhor visualização e comparação entre as posições das mesorregiões. Assim, em relação à dimensão 1, as mesorregiões: 19 (Metropolitana de São Paulo) 8 (Distrito Federal), 2 (Campinas), 33 (Vale do Itajaí), 15 (Metropolitana de Fortaleza) e 16 (Metropolitana de Porto Alegre)

opõem-se às mesorregiões 22 (Nordeste Paraense), 21 (Nordeste Baiano) e 29 (Sudeste Paraense). Em relação à dimensão 2, as mesorregiões 20 (Metropolitana do Rio de Janeiro) e 34 (Vale do Paraíba Paulista) seriam mais relacionadas com biotecnologia em oposição a mesorregião 23 (Noroeste Rio-grandense) mais ligada à engenharias agrônomas e de pesca.

Gráfico 8 – Análise de Correspondência para obtenção da Distância Cognitiva (Qui-Quadrado) entre as mesorregiões — coordenadas “principais” para as linhas (mesorregiões)

Fonte: Elaboração Própria

O uso do EMD para a distância geográfica ilustrou como essa técnica pode apoiar a representação das dimensões da proximidade; desta forma, nesta seção, foi possível

com uso da AC interpretar as dimensões no gráfico obtido para melhor representação da medida para a proximidade cognitiva, facilitando a comparação de diferentes definições da variável de interesse que caracterizaria a proximidade, dado que as representações para as formas não espaciais da proximidade não possuem mapas de referência como os mapas políticos para a distância geográfica.

Por fim, novamente se ressalta que, por meio da AC, abrem-se novas possibilidades de representação para medir as dimensões não espaciais da proximidade, ainda não realizadas na literatura. A próxima seção apresenta a verificação empírica das possíveis diferenças na inferência de modelos estatísticos, considerando as distintas formas de medir a proximidade.

4.5 Influência da medição da proximidade não espacial na análise inferencial

Estudos recentes procuraram captar a influência das dimensões da proximidade no número de colaborações e sugerem que a proximidade espacial e as não espaciais têm papel importante nas decisões relativas à escolha de parceiros para colaboração, seja em parcerias científicas (CAPELLO; CARAGLIU, 2015; (FERNÁNDEZ; FERRÁNDIZ; LEÓN, 2016), em patenteamento (MARROCU; PACI; USAI, 2013) ou em outros projetos conjuntos (MAREK *et al.*, 2016).

Deste modo, esta seção apresenta uma verificação empírica da importância da proximidade, testando se as distintas formas de medir uma distância não espacial podem resultar em diferenças na inferência quanto aos modelos estimados, para explicar a quantidade de interações entre universidades e empresas no Brasil. A análise também avaliou os limites do ganho com a proximidade e os mecanismos de complementariedade e de substituição entre as dimensões geográfica e cognitiva nessas relações, respectivamente nas seções 4.5.1 e 4.5.2.

Para tanto, foram utilizadas as mesmas unidades de análise apresentadas na seção 4.2, que consideram dados agregados em 36 mesorregiões (tabela 1), recorte esse que perfaz a distribuição de 4.342 interações. Dado isso, a quantidade de interesse ou variável a ser explicada será a quantidade de interações computadas entre cada par de mesorregião. Assim, para construção da base a ser utilizada nos modelos,

considera-se a premissa de que as relações são unilaterais, ou seja, ocorrem entre empresas de uma dada mesorregião com grupos de pesquisa acadêmicos de cada mesorregião, portanto há 1.296 possibilidades de interações entre as 36 mesorregiões conforme mostrado na tabela 2.

O modelo que considera uma distribuição Binomial Negativa é preferível, tanto pela quantidade de interações ser dados de contagem quanto por essa quantidade possuir elevada dispersão quando comparada com a média (variância igual a 432,14 e média igual a 3,35). Além disso, em quase 70% dos pares possíveis de mesorregiões, não ocorreu interação; a grande quantidade de zeros justifica a utilização de um modelo Binomial Negativo Inflado de Zeros (CAMERON; TRIVEDI, 2009; CAMERON; TRIVEDI, 2013).

A abordagem com modelo Binomial Negativa foi utilizada por muitos autores (MAREK *et al.*, 2016; CAPELLO; CARAGLIU, 2015; FERNÁNDEZ; FERRÁNDIZ; LEÓN, 2016; CASSI; PLUNKET, 2014); em resumo, consiste em um processo em dois estágios: uma regressão Binomial Negativa, que estima a quantidade de interações entre um par de mesorregiões, dado que cada uma tem pelo menos uma interação; uma regressão logística que estima a probabilidade de não ocorrer interação, deste modo dois coeficientes são obtidos para cada preditor. Assim, o modelo Binomial Negativo Inflado de Zeros considera duas fontes de zeros, a saber: o zero verdadeiro, que faz parte da distribuição amostral, e o excesso de zeros, pois parte das interações podem não ocorrer porque ambas mesorregiões não interagiram, e outra parte por não haver interesse na interação (FERNÁNDEZ; FERRÁNDIZ; LEÓN, 2016).

O conjunto de variáveis selecionadas como preditoras para explicar a quantidade de interação entre as mesorregiões organizam-se em dois grupos: características internas e externas às regiões, similar a abordagem de Capello e Caragliu (2015).

Para o primeiro grupo, que considera os fatores internos da mesorregião, obtêm-se informações da mesorregião de origem da empresa (capacidade de absorção, quantidade de empresas) e da mesorregião de origem do grupo (quantidade de artigos), como motivadores de uma maior quantidade de interações.

Em relação ao grupo externo de informações, consideram-se, além das distâncias (espacial e não espacial), as diferenças estruturais entre as regiões, as quais representariam o interesse de a empresa de uma determinada mesorregião interagir com grupos de pesquisa localizados em mesorregiões com menor “*gap*” entre a estrutura produtiva (PIB *per capita*) e a capacidade de aplicação do conhecimento científico (Patentes por empregados) (CAPELLO; CARAGLIU, 2015).

Além de detalhes sobre a descrição para todas as variáveis, também é possível notar no quadro 10 variáveis derivadas para uso nos modelos, obtidas a partir da distância espacial (geográfica) e da distância não espacial (cognitiva), esta última definida nas três diferentes maneiras, conforme apresentadas (euclidiana, cosseno e qui-quadrado), para fins de comparação de seus resultados de inferência.

As variáveis que representam o quadrado de cada distância indicada, como: *distgeo2*, *distnaogoeuclideana2*, *distnaogeocosseno2* e *distnaogeoquiq2*, serão utilizadas na próxima seção 4.5.1, a fim de se avaliarem os limites do ganho com a proximidade. Por fim, cada variável que representa a interação entre a distância geográfica e cada distância cognitiva (*Int_distnaogoeuclideana*, *Int_distnaogeocosseno* e *Int_distnaogeoquiq*) foram úteis para avaliar, na seção 4.5.2, os mecanismos de complementariedade e de substituição entre as dimensões geográfica e cognitiva.

De maneira geral, as demais variáveis representam motivadores do interesse da empresa em interagir com grupos de pesquisa. Pretende-se captar se a capacidade da empresa em interagir com grupos localizados em outras mesorregiões (*capabsoremp*) tem efeito na quantidade de interações entre os pares de mesorregiões. Duas outras variáveis representam as “massas” na mesorregião de origem da empresa pela quantidade de empresas (*Inquantempresas*), e na mesorregião de origem do grupo de pesquisa pela quantidade de artigos (*Inquantartigosuniv*), de modo a medir o quanto o volume de empresas presentes na mesorregião de origem incrementa o volume de interações, assim como uma medida de qualidade científica da mesorregião de origem do grupo de pesquisa possibilita um interesse da empresa em interagir mais.

Por fim, outras variáveis que representam as características externas à mesorregião na qual a empresa está localizada podem explicar a quantidade de interações pelo menor *gap* entre a estrutura produtiva, representada pela diferença entre PIB *per capita* (pibpercapitadiferenca) e entre a capacidade de aplicação do conhecimento científico, medida pela diferença no patenteamento (denspatentesdiferenca). A tabela 8 apresenta as estatísticas descritivas para todas as variáveis do estudo.

Tabela 8 – Resumo de Estatísticas Descritivas para as variáveis do estudo

Variável	Média	Desvio Padrão	Mínimo	Máximo
quantinteracao	3,35	20,79	0	405
distgeo	1309	801	0	3215
distnaogeoEuclidean	0,20	0,10	0	0,51
distnaogeocosseno	0,07	0,07	0	0,38
distnaogeoqui	0,47	0,27	0	1,32
distgeo2	2354364	2383217	0	10300000
distnaogeoEuclidean2	0,05	0,05	0	0,26
distnaogeocosseno2	0,01	0,02	0	0,15
distnaogeoqui2	0,29	0,31	0	1,75
Int_distnaogeoEuclidean	273,23	223,42	0	1281,59
Int_distnaogeocosseno	95,10	120,22	0	796,46
Int_distnaogeoqui	649,10	596,58	0	3519,22
capabsoremp	0,14	0,04	0,08	0,26
Inquantempresas	10,65	0,86	8,52	12,95
Inquantartigossuniv	8,50	1,71	4,16	11,14
pibpercapitadiferenca	11.530,30	9.653,98	0,00	51.112,31
denspatentesdiferenca	0,55	0,42	0	1,86

Fonte: Elaboração Própria

Quadro 10 – Descrição das variáveis do estudo

Variável	Descrição
Quantinteracao	Quantidade de interações entre cada par de mesorregiões a partir do Diretório dos Grupos de Pesquisa da base Lattes do CNPq (DGP), considerando período de referência 2010
distgeo	Distância em linha (quilômetros) entre a sede dos municípios mais populosos de cada par de mesorregião a partir de suas coordenadas georeferenciadas (latitude e longitude)
distnaogeo euclideana	Distância cognitiva calculada conforme distância euclidiana entre distribuições de pessoal ocupado técnico-científico a partir da RAIS 2009, considerando 11 classes (CBO)
distnaogeo cosseno	Distância cognitiva calculada conforme índice cosseno entre distribuições de pessoal ocupado técnico-científico a partir da RAIS 2009, considerando 11 classes (CBO)
distnaogeo quiq	Distância cognitiva calculada com as coordenadas obtidas na Análise de Correspondência com a tabela de contingência, formada pela distribuição de pessoal ocupado técnico-científico a partir da RAIS 2009 considerando 11 classes (CBO)
distgeo2	Quadrado da distância geográfica
distnaogeo euclideana2	Quadrado da distância cognitiva (euclidiana)
distnaogeo cosseno2	Quadrado da distância cognitiva (cosseno)
distnaogeo quiq2	Quadrado da distância cognitiva (qui-quadrado)
Int_distnaogeo euclideana	Interação entre a distância geográfica e a distância cognitiva (euclidiana)
Int_distnaogeo cosseno	Interação entre a distância geográfica e a distância cognitiva (cosseno)
Int_distnaogeo quiq	Interação entre a distância geográfica e a distância cognitiva (qui-quadrado)
capabsoremp	Razão entre o total de empregados com graduação, mestrado ou doutorado e o total de empregados com vínculos ativos em dezembro de 2009 na RAIS/MCTI
Inquantempresas	Logaritmo natural da quantidade de empresas na mesorregião da empresa que interage a partir da RAIS 2009
Inquantartigosuniv	Logaritmo natural da quantidade de artigos na mesorregião do grupo de pesquisa que interage a partir do DGP, considerando período de referência 2010
pibpercapitadiferenca	Diferença entre Produto Interno Bruto <i>per capita</i> da mesorregião da empresa e da mesorregião do grupo de pesquisa (IBGE, 2009)
denspatentesdiferenca	Diferença entre Densidade de patentes de invenção por mil empregados entre as mesorregiões da empresa e do grupo de pesquisa (localização do depositante - INPI, 2009)

Fonte: Elaboração Própria

Inicialmente, foram observadas as correlações, utilizando o coeficiente de Pearson entre a quantidade de interações (Y) e as variáveis preditoras, as quais fornecem indícios às próximas seções para guiar a especificação dos modelos a serem estimados. De forma geral, apresentam-se nos gráficos 9 e 10, a dispersão e os respectivos testes de correlação. No gráfico 9 as quatro distâncias e no 10 as demais variáveis.

Entre as principais variáveis preditoras, as distâncias espacial e não espacial, nota-se que a partir das correlações (gráfico 9), todas as formas utilizadas para medir as distâncias estão positivamente correlacionadas entre si e negativamente correlacionadas com a quantidade de interações (Y). Porém, verifica-se uma ordem de importância quanto maior o coeficiente de Pearson: a distância geográfica seria a mais relevante com índice de -0,2, seguida pela distância cognitiva mensurada pela forma euclidiana e pela metodologia proposta (qui-quadrado) com índice decorrelação similar (-0,19). Apesar da alta correlação entre as formas euclidiana e cosseno (0,95), o índice usual cosseno apresenta-se como o de valor menos correlacionado com o Y do estudo.

Gráfico 9 – Gráficos de dispersão com respectivos testes de correlação de Pearson entre Y e as dimensões da proximidade Geográfica e Cognitiva (Euclidiana, Cosseno e Qui-Quadrado)

Fonte: Elaboração Própria

Gráfico 10 – Gráficos de dispersão com respectivos testes de correlação de Pearson entre Y e as variáveis que não representam distâncias

Fonte: Elaboração Própria

Os gráficos de dispersão mostram o relacionamento entre as demais variáveis preditoras (gráfico 10). Nota-se que as cinco variáveis estão razoavelmente correlacionadas com a quantidade de interação (Y), cujos indícios as caracterizam como preditoras individuais, que serão avaliadas em conjunto nas próximas seções. Esses indícios mostram que a capacidade de absorção das empresas, a quantidade de empresas na mesorregião de origem da empresa e a quantidade de artigos produzidos na mesorregião de origem do grupo favorecem o maior número de interações entre empresas e grupos de pesquisas desses pares de mesoregiões.

De modo contrário, a correlação negativa entre a quantidade de interações e cada uma das variáveis estruturais (PIB *per capita* e densidade de patentes por empregado) sugere que quanto menor o *gap* nestas quantidades, maior o interesse da empresa de uma determinada mesorregião interagir com grupos de pesquisa de outra, do mesmo modo que indicado por Capello e Caragliu (2015).

Dada a expectativa de inferência em relação às características e à quantidade de interações nas mesoregiões, a verificação empírica foi planejada em três conjuntos de modelos, todos estimados considerando erros robustos para lidar com a possível heterocedasticidade nos dados. A primeira parte da análise, objetivo desta seção, consiste em 16 modelos. Na tabela 9, apresentam-se oito modelos estimados: quatro deles consideram apenas as diferentes distâncias cognitivas (1 a 4) e para os outros quatro acrescentou-se a capacidade de absorção das empresas (5 a 8). Na tabela 10, incluíram-se as quantidades de empresas e de artigos para os quatro primeiros modelos (9 a 12), e o PIB *per capita* e a densidade de patentes nos outros quatro (13 a 16).

A segunda e terceira análise considerou os mesmos modelos, porém adaptados para avaliar:

- a) os limites do ganho com a proximidade (seção 4.5.1), sendo adicionadas as distâncias quadráticas em 16 modelos (17 a 32);
- b) os mecanismos de complementariedade e de substituição entre as dimensões geográfica e cognitiva, incluindo-se variáveis de interação entre a distância geográfica e cada uma das distâncias cognitivas, o que totaliza mais 12 modelos (33 a 44) apresentados na (seção 4.5.2).

Duas primeiras observações, relativas ao uso do modelo Binomial Negativo Inflado de Zeros, podem ser feitas:

- a) teste de superdispersão significativo em todos os modelos estimados, visto pelo logaritmo natural do parâmetro de dispersão α para modelos de contagem ($\ln \alpha$), caso contrário poderia ser mais apropriado modelos de Poisson;
- b) um segundo teste, denominado Vuong, que considera a necessidade do modelo inflado de zeros em relação ao padrão Binomial Negativo. Nesse caso, os modelos iniciais com poucas variáveis não foram condizentes com essa suposição.

As conclusões inferenciais dos primeiros modelos, apresentados nas tabelas 9 e 10, foram significantes e com sinal esperado, de acordo com a revisão levantada da literatura, em especial, em relação às distâncias e à capacidade de absorção. A distância geográfica, individualmente ou em conjunto com a cognitiva, foi significativa e com sinal negativo, o que mostra que quanto mais distantes as mesorregiões da empresa e do grupo de pesquisa, menos interações ocorrem. Já em relação à capacidade de absorção, conclui-se que as mesorregiões que apresentam empresas com maior capacidade podem ir mais longe, dada a maior quantidade de interações.

Ao incluir todas as variáveis do estudo (modelos 13 a 16), algumas conclusões de inferência mudam conforme a escolha da medida selecionada para mensurar a distância cognitiva. Considerando a distância cognitiva euclidiana, a densidade de patentes não apresenta a significância e, quando da utilização da distância cognitiva qui-quadrado, além da densidade de patentes, a variável PIB *per capita* também não. Essas mudanças na inferência mostram a possível sensibilidade da escolha do índice a ser utilizado para medir a distância cognitiva nos modelos. Um resumo dos sinais esperados nas inferências dos diferentes modelos estimados é apresentado no apêndice B, no qual o sinal (negativo ou positivo) retrata a inferência esperada pela literatura e a formatação das caselas representa a mudança na inferência.

Tabela 9 – Estimativas dos parâmetros para Quantidade de Interações considerando Binomial Negativa Inflada de Zeros – Capacidade de Absorção

	1	2	3	4	5	6	7	8
Binomial Negativa	distgeo	-0,001*** (0,0001)	-0,001*** (0,0001)	-0,001*** (0,0001)	-0,001*** (0,0001)	-0,001*** (0,0001)	-0,001*** (0,0001)	-0,001*** (0,0001)
	distnaogoeuclidean		-3,336*** (0,817)			-3,238*** (0,627)		
	distnaogeocosseno			-3,157** (1,478)			-2,380* (1,273)	
	distnaogeoqui				-2,833*** (0,445)			-2,827*** (0,316)
	capabsoremp				24,973*** (3,374)	23,487*** (2,798)	24,376*** (2,952)	20,297*** (2,231)
	Inquantempresas							
	Inquantartigossuniv							
	pibpercapitadiferenca							
	denspatentesdiferenca							
	Constante	2,213***	2,618***	2,326***	2,903***	-1,736***	-1,121**	-1,555***
Logística	distgeo	0,001*** (0,001)	0,001* (0,0002)***	0,002** (0,001)	0,007*** (0,002)	0,0003 (0,0003)	0,0006*** (0,0002)	0,0005** (0,0002)
	distnaogoeuclidean		-5,518 (4,017)			-3,003 (2,216)		
	distnaogeocosseno			-198,752 (146,729)			-6,927 (7,522)	
	distnaogeoqui				-0,255 (3,121)			-2,751*** (0,849)
	capabsoremp				-29,732*** (10,723)	-19,345** (8,323)	-26,713*** (8,834)	-28,376*** (7,581)
	Inquantempresas							
	Inquantartigossuniv							
	pibpercapitadiferenca							
	denspatentesdiferenca							
	Constante	-15,209***	-2,648**	-3,812***	-3,441**	0,133	0,351	0,512
	Teste superdispersão (ln alpha)	1,900***	1,769***	1,867***	1,657***	1,503***	1,361***	1,472***
	Teste Vuong	-0,16	0,43	0,62	0,42	0,93	0,86	0,85
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1								

Fonte: Elaboração Própria

Tabela 10 – Estimativas dos parâmetros para Quantidade de Interações considerando Binomial Negativa Inflada de Zeros – Demais Variáveis

	9	10	11	12	13	14	15	16
Binomial Negativa								
distgeo	-0,001*** (0,0001)	-0,001*** (0,0001)	-0,001*** (0,0001)	-0,001*** (0,0001)	-0,001*** (0,0001)	-0,001*** (0,0001)	-0,001*** (0,0001)	-0,001*** (0,0001)
distnaogoeuclidean		-5,087*** (0,615)				-4,499*** (0,615)		
distnaogecosseno			-4,834** (1,291)				-4,633*** (1,072)	
distnaogeoquiq				-2,655*** (0,336)				-2,175*** (0,477)
capabsoremp	14,414*** (2,553)	11,822*** (2,312)	13,069*** (2,659)	12,517*** (2,001)	15,966*** (2,357)	14,702*** (2,273)	15,324*** (2,325)	13,523*** (3,004)
Inquantempresas	0,305** (0,120)	0,456*** (0,117)	0,353*** (0,128)	0,372*** (0,123)	0,319*** (0,115)	0,409*** (0,109)	0,349*** (0,112)	0,375*** (0,133)
Inquantartigossuniv	0,548*** (0,0102)	0,656*** (0,087)	0,602*** (0,097)	0,579*** (0,077)	0,560*** (0,084)	0,654*** (0,077)	0,612*** (0,079)	0,581*** (0,079)
pibpercapitadiferenca					-0,00001*** (0,000001)	-0,00003*** (0,000001)	-0,00004*** (0,000001)	-0,00001 (0,00001)
denspatentesdiferenca					-0,589*** (0,220)	-0,333 (0,228)	-0,507** (0,222)	-0,222 (0,235)
Constante	-8,422***	-10,135***	-9,079***	-8,598***	-8,645***	-9,869***	-9,185***	-8,833***
Logística								
distgeo	0,001*** (0,0002)	0,001*** (0,0001)	0,001*** (0,0001)	0,001*** (0,0001)	0,001*** (0,0002)	0,001*** (0,0002)	0,001*** (0,0002)	0,001*** (0,0003)
distnaogoeuclidean		-1,439 (1,649)				-0,013 (1,766)		
distnaogecosseno			-2,784 (3,529)				-0,769 (2,955)	
distnaogeoquiq				-0,099 (0,833)				1,598 (1,699)
capabsoremp	-8,722 (5,394)	-11,303** (5,731)	-10,409* (5,755)	-9,739 (5,084)	-12,363** (5,777)	-12,206** (6,178)	-12,499** (5,828)	-11,169 (9,448)
Inquantempresas	-0,568** (0,235)	-0,385 (0,238)	-0,487** (0,246)	-0,491** (0,227)	-0,383* (0,226)	-0,326 (0,229)	-0,354 (0,223)	-0,288 (0,297)
Inquantartigossuniv	-1,030*** (0,238)	-0,861*** (0,227)	-0,948*** (0,240)	-0,913*** (0,236)	-0,951*** (0,214)	-0,850*** (0,213)	-0,889*** (0,212)	-0,993*** (0,245)
pibpercapitadiferenca					-0,000001 (0,00001)	-0,000002 (0,00001)	-0,000005 (0,00001)	-0,00001 (0,00004)
denspatentesdiferenca					-1,481** (0,643)	-1,118 (0,705)	-1,333** (0,653)	-1,607* (0,884)
Constante	14,780***	12,009***	13,621***	13,013***	12,801***	11,290***	12,052***	11,282***
Teste superdispersão (ln alpha)	0,801***	0,534***	0,723***	0,491***	0,642***	0,438***	0,573***	0,532***
Teste Vuong	1,80**	1,50*	1,50*	1,57	2,31*	1,79**	2,09**	1,94**

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Fonte: Elaboração Própria

As duas próximas seções mostram variações no modelo inicial para testar as duas hipóteses, uma sobre os limites nos ganhos com a proximidade e outra sobre a complementariedade ou substituição entre as proximidades geográfica e cognitiva.

4.5.1 Limites nos ganhos

Esta seção apresenta os resultados para os modelos com o acréscimo de termos quadráticos, de modo a medir efeitos de não linearidade, ou seja, se há limites no ganho com a proximidade (tabelas 11 e 12). A existência de um limite de ganho com a proximidade supõe a existência de um nível ótimo.

A primeira observação sobre as conclusões inferenciais dos modelos desta seção é de que a distância geográfica permanece significativa e, ao acrescentar o termo quadrático, ele é também significativo e positivo (modelos 17, 21 25 e 29). O valor positivo do termo quadrático implica que o efeito negativo da distância decai, isto é, quanto mais distante menos interação, sendo essa queda na interação menor do que proporcional. O modelo 29, ao incluir todas as variáveis, revela que o PIB não apresenta significância.

Considerando o efeito conjunto das distâncias geográfica e cognitiva na quantidade de interação, observa-se o mesmo resultado geral: além do efeito negativo da distância e de valores positivos para o termo quadrático, há a mesma indicação de que esse efeito negativo da distância decai. Porém, os resultados em relação à significância (inferência) variam quanto à escolha da forma de medir a distância cognitiva. Para exemplificar as diferenças, nota-se que os modelos que consideram a escolha da distância cognitiva euclidiana (modelos 18, 22, 26 e 30), em dois deles (18 e 26) o efeito quadrático da distância geográfica é não significativo. Já em relação à escolha das distâncias cognitivas que consideram cosseno ou qui-quadrado, o efeito quadrático que não apresenta significância é o da distância cognitiva.

Por fim, ao incluir todas as variáveis do estudo (modelos 29 a 32), as conclusões de inferência que mudariam, considerando a escolha da distância cognitiva, seriam o efeito dessa distância não ser significativo (modelo 31) e o PIB não ser significativo.

Novamente, há mudanças na inferência que mostram a sensibilidade da escolha do índice a ser utilizado para medir a distância cognitiva nos modelos.

Tabela 11 – Estimativas dos parâmetros para Quantidade de Interações considerando Binomial Negativa Inflada de Zeros – Capacidade de Absorção (com termo quadrático)

	17	18	19	20	21	22	23	24
Binomial Negativa								
distgeo	-0,004*** (0,001)	-0,001*** (0,0005)	-0,003*** (0,0004)	-0,002*** (0,0004)	-0,005*** (0,0003)	-0,004*** (0,0005)	-0,005*** (0,0004)	-0,003*** (0,0005)
distnaogoeuclidean		-13,684*** (3,134)				-10,191*** (2,984)		
distnaogecosseno			-7,623 (5,841)				-2,957 (3,756)	
distnaogeoquiq				-1,778 (1,084)				-1,783* (0,930)
distgeo2	0,000002** (0,0000005)	0,0000003 (0,0000002)	0,0000006** (0,0000002)	0,0000005** (0,0000002)	0,000001*** (0,0000001)	0,000001*** (0,0000002)	0,000002*** (0,0000002)	0,000001*** (0,0000002)
distnaogoeuclidean2		27,169*** (7,088)				23,541*** (6,554)		
distnaogecosseno2			30,212 (22,603)				21,739 (16,179)	
distnaogeoquiq2				-0,288 (0,956)				-0,578 (0,908)
capabsoremp					24,973*** (1,702)	24,765*** (1,684)	24,938*** (1,694)	20,899*** (2,019)
Inquantempresas								
Inquantartigossuniv								
pibpercapitadiferenca								
denspatentesdiferenca								
Constante	3,065***	3,443***	3,013***	3,276***	-0,795***	-0,431	-0,794***	0,176
distgeo	0,002 (0,003)	0,009*** (0,002)	0,008*** (0,002)	0,007*** (0,001)	0,333*** (0,109)	-0,003** (0,001)	0,015 (0,034)	0,001 (0,001)
distnaogoeuclidean		-20,858*** (7,239)				-16,921 (26,699)		
distnaogecosseno			-20,514* (12,105)				-8,401 (19,745)	
distnaogeoquiq				0,053 (2,971)				11,539** (4,793)
distgeo2	0,0000002 (0,0000002)	0,0000003*** (0,0000006)	-0,0000003*** (0,0000008)	-0,0000002*** (0,0000005)	-0,0000001 (0,0000001)	0,0000001 (0,0000009)	-0,0000002 (0,0000006)	0,0000001 (0,0000004)
distnaogoeuclidean2		41,375*** (14,553)				44,163 (33,544)		
distnaogecosseno2			88,939** (41,419)				79,665 (104,211)	
distnaogeoquiq2				0,591 (2,476)				-22,626** (9,356)
capabsoremp					-36,349* (19,012)	-5,653 (12,208)	-9,577 (8,679)	-29,174*** (8,597)
Inquantempresas								
Inquantartigossuniv								
pibpercapitadiferenca								
denspatentesdiferenca								
Constante	-4,119	-4,048***	-5,191***	-5,083**	-487,051***	-5,961	-25,121	0,497
Teste superdispersão (ln alpha)	1,554***	1,360***	1,473***	1,371***	1,147***	1,029***	1,096***	0,889***
Teste Vuong	1,27	2,82***	2,68***	2,03**	2,90***	1,99**	1,98**	1,99**
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1								

Fonte: Elaboração Própria

Tabela 12 – Estimativas dos parâmetros para Quantidade de Interações considerando Binomial Negativa Inflada de Zeros – Demais Variáveis (com termo quadrático)

	25	26	27	28	29	30	31	32	
Binomial Negativa	distgeo	-0,006*** (0,0003)	-0,004*** (0,0003)	-0,005*** (0,0002)	-0,004*** (0,0003)	-0,005*** (0,0003)	-0,004*** (0,0003)	-0,005*** (0,0003)	-0,004*** (0,0003)
	distnaogeoeludeana		-13,699*** (1,791)				-12,863*** (1,651)		
	distnaogeoecosseno			-11,945*** (2,778)				-9,758 (2,452)	
	distnaogeoquiq				-4,074 (0,796)				-4,067*** (0,930)
	distgeo2	0,000002** (0,0000001)	0,000001 (0,0000001)	0,000001** (0,0000001)	0,000001** (0,0000001)	0,000002*** (0,0000001)	0,000001*** (0,0000001)	0,000002*** (0,0000001)	0,000001*** (0,0000001)
	distnaogeoeludeana2		26,003*** (4,182)				24,353*** (3,644)		
	distnaogeoecosseno2			47,755*** (12,938)				35,352*** (9,573)	
	distnaogeoquiq2				2,823 (0,723)				2,683*** (0,766)
	capabsoremp	15,686*** (1,775)	14,949*** (1,666)	15,273*** (1,702)	15,307*** (1,809)	14,970*** (1,783)	13,847*** (1,915)	14,695*** (1,772)	13,060*** (1,690)
	Inquantempresas	0,556** (0,091)	0,633*** (0,093)	0,601*** (0,091)	0,583*** (0,104)	0,599*** (0,091)	0,665*** (0,100)	0,634*** (0,092)	0,652*** (0,104)
	Inquantartigosuniv	0,790*** (0,052)	0,814*** (0,049)	0,802*** (0,052)	0,789*** (0,047)	0,807*** (0,051)	0,812*** (0,060)	0,829*** (0,051)	0,799*** (0,046)
	pibpercapitadiferenca					0,0000002 (0,0000007)	0,000002 (0,0000007)	-0,000001 (0,0000007)	0,00001** (0,0000007)
	denspatentesdiferenca					-0,483*** (0,183)	-0,401** (0,198)	-0,510*** (0,177)	-0,273 (0,175)
	Constante	-12,742***	-13,118***	-13,149***	-12,548***	-13,166***	-13243***	-13,577***	-13,044
Logística	distgeo	0,041 (0,027)	0,031 (0,025)	0,024 (0,026)	0,047 (0,034)	0,056** (0,025)	0,001 (0,004)	0,046* (0,026)	0,061** (0,026)
	distnaogeoeludeana		-34,400 (67,147)				37,554 (28,109)		
	distnaogeoecosseno			-40,259 (42,862)				36,492 (34,808)	
	distnaogeoquiq				-0,856 (5,148)				4,366 (5,023)
	distgeo2	-0,000006 (0,000005)	-0,000004 (0,000004)	-0,000002 (0,000004)	-0,000007 (0,000006)	-0,000009** (0,000005)	0,000001 (0,000002)	-0,000007 (0,000005)	-0,000009** (0,000004)
	distnaogeoeludeana2		93,745 (145,649)				-55,772 (56,704)		
	distnaogeoecosseno2			224,424 (185,201)				-107,318 (129,950)	
	distnaogeoquiq2				0,289 (4,404)				-0,524 (5,043)
	capabsoremp	-30,683* (15,784)	-40,240* (20,683)	-40,758* (22,218)	-36,084* (18,732)	-41,196* (21,050)	-45,862 (54,017)	-52,209*** (20,099)	-51,671 (35,451)
	Inquantempresas	0,059 (0,665)	0,132 (0,745)	0,379 (0,823)	0,182 (0,661)	0,307 (0,667)	0,099 (1,262)	0,272 (0,614)	0,741 (0,945)
	Inquantartigosuniv	-1,917** (0,874)	-2,609 (1,717)	-3,282 (2,027)	-2,005** (0,881)	-1,337** (0,618)	-1,595*** (0,601)	-1,483** (0,637)	-1,598* (0,923)
	pibpercapitadiferenca					-0,00009 (0,00014)	-0,00002 (0,0003)	-0,0001 (0,0001)	-0,00001 (0,0002)
	denspatentesdiferenca					-2,486 (3,211)	-4,260 (7,408)	-3,219 (3,244)	-3,395 (3,780)
	Constante	-44,991	-25,859	-16,399	-52,385	-64,537*	6,852	-50,144	-75,560**
Teste superdispersão (ln alpha)	0,416***	1,945*	0,380***	0,222**	0,364***	0,151	0,311***	0,197*	
Teste Vuong	4,02***	4,07***	4,25***	3,65***	4,30***	4,10***	4,35***	4,11***	
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1									

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Fonte: Elaboração Própria

4.5.2 Complementariedade e Substituição

Esta seção defronta-se com o segundo ponto de análise, que considera as possíveis sinergias entre as distâncias geográfica e cognitiva, podendo ser realizadas com modelos que consideram termos de interação entre essas duas variáveis.

Inicialmente, observa-se que as conclusões inferenciais de todos os modelos apresentados nesta seção (33 a 44), conforme tabelas 13 e 14, são condizentes com a literatura e com os modelos sem os termos com as interações. Além disso, com a inclusão das interações, não há resultados diferentes em relação a cada definição de distância cognitiva. As distâncias geográfica e cognitiva são significantes e negativas e todas as demais variáveis permanecem significantes e com mesmo sinal: capacidade de absorção, quantidade de empresas e de grupos, além do PIB e da densidade de patentes.

Em relação à inclusão dos respectivos termos com a interação entre as distâncias geográfica e a cognitiva, observa-se a sua significância, as quais se revelam variáveis relevantes na explicação da quantidade de interações entre empresas e grupos de pesquisa das mesorregiões. Por fim, na ausência dos termos de interações nos modelos, é possível que os efeitos individuais de cada distância podem estar superestimados, já que parte de cada um deles é explicada, na verdade, em conjunto.

Além disso, dado o sinal positivo para os termos de interação, há a indicação de que os efeitos conjuntos das distâncias na quantidade de interações são de ordem complementar. Deste modo, além das mesorregiões mais próximas geograficamente interagirem mais, assim como as mais próximas cognitivamente, comprovam-se os indícios de que as duas variáveis relativamente a cada sentido (geográfico e cognitivo, nas diferentes formas de medição) se complementam para o alcance de mais interações.

Apesar das conclusões inferenciais nesta seção terem permanecidas as mesmas, independentemente da escolha da forma como a distância cognitiva foi medida, as diferenças nos resultados das seções anteriores mostram quão importante é a escolha da forma de medir e a seleção de meios de avaliar vantagens de cada abordagem, além da interpretação de cada índice, pois isto pode implicar em distintos resultados

e mudanças nas inferências esperadas, ou seja, na significância de algumas das variáveis.

Tabela 13 – Estimativas dos parâmetros para Quantidade de Interações considerando Binomial Negativa Inflada de Zeros – Capacidade de Absorção (com interações)

	33	34	35	36	37*	38
Binomial Negativa	distgeo	-0,002*** (0,0002)	-0,002*** (0,0002)	-0,001*** (0,0002)	-0,003*** (0,0003)	-0,002*** (0,0002)
	distnaogeoeludeana	-5,730*** (0,976)			-6,089*** (0,821)	
	distnaogeocosseno		-7,209*** (1,735)		-7,721*** (1,496)	
	distnaogeoquiq			-3,786*** (0,541)		-3,794*** (0,397)
	Int_distnaogeoeludeana	0,006*** (0,001)			0,007*** (0,001)	
	Int_distnaogeocosseno		0,006*** (0,002)			0,008*** (0,003)
	Int_distnaogeoquiq			0,002*** (0,0005)		0,001*** (0,0004)
	capabsoremp				23,725*** (3,077)	25,301*** (6,781)
	Inquantempresas					20,506*** (2,260)
	Inquantartigossuniv					
	pibpercapitadiferenca					
	denspatentesdiferenca					
	Constante	2,914**	2,512***	3,244***	-0,802	-1,497
		-0,0004 (0,0004)	0,0002 (0,0002)	0,0005 (0,0004)	-0,0004 (0,0004)	-0,0001 (0,0005)
		0,002 (0,001)				
Logística	distgeo	2,714* (1,533)			1,509 (1,439)	
	distnaogeoeludeana					
	distnaogeocosseno		5,672 (4,308)			1,689 (10,278)
	distnaogeoquiq			3,066* (1,588)		-0,009 (0,971)
	Int_distnaogeoeludeana	0,006*** (0,002)			0,005** (0,002)	
	Int_distnaogeocosseno		0,005** (0,003)			0,004 (0,003)
	Int_distnaogeoquiq			0,0004 (0,0008)		-0,003 (0,002)
	capabsoremp				-12,504 (13,659)	-3,929 (34,581)
	Inquantempresas					-51,234 (49,381)
	Inquantartigossuniv					
	pibpercapitadiferenca					
	denspatentesdiferenca					
	Constante	-1,121**	-4,154***	-4,221***	-1,378	-1,579
		1,670***	1,776***	1,554***	1,200***	1,386***
		1,47*	1,62*	1,13	1,39*	1,67**
		1,203***				1,40*

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Fonte: Elaboração Própria

Tabela 14 – Estimativas dos parâmetros para Quantidade de Interações considerando Binomial Negativa Inflada de Zeros – Demais variáveis (com interações)

	39	40	41	42	43	44		
Binomial Negativa	distgeo	-0,002*** (0,0002)	-0,002*** (0,0001)	-0,002*** (0,0001)	-0,002*** (0,0003)	-0,001*** (0,0001)	-0,002*** (0,0003)	
	distnaogeoeludeana	-8,062*** (0,873)			-7,907*** (1,211)			
	distnaogeoecosseno		-10,757*** (1,554)			-9,124*** (1,574)		
	distnaogeoquiq			-4,045*** (0,539)			-3,643*** (0,656)	
	Int_distnaogeoeludeana	0,007*** (0,001)			0,006*** (0,001)			
	Int_distnaogeoecosseno		0,008*** (0,001)			0,007*** (0,001)		
	Int_distnaogeoquiq			0,002*** (0,0005)			0,002*** (0,0005)	
	capabsoremp	13,320*** (2,016)	13,997*** (2,265)	12,785*** (1,896)	15,108*** (2,524)	15,788*** (2,426)	14,822*** (2,582)	
	Inquantempresas	0,533*** (0,112)	0,442*** (0,116)	0,444*** (0,143)	0,523*** (0,138)	0,394*** (0,117)	0,437*** (0,148)	
	Inquantartigossuniv	0,658*** (0,105)	0,581*** (0,088)	0,596*** (0,076)	0,749*** (0,123)	0,610*** (0,094)	0,634*** (0,122)	
	pibpercapitadiferenca				-0,00002*** (0,000007)	-0,00004*** (0,000007)	-0,00002** (0,000008)	
	denspatentesdiferenca				-0,003 (0,284)	-0,257 (0,229)	-0,227 (0,255)	
	Constante	-10,923**	-9,801***	-9,242	-11,907**	-9,675***	-9,831***	
	Binomial Negativa	distgeo	0,0002 (0,0003)	0,0004* (0,0002)	-0,001** (0,0003)	-0,002*** (0,003)	0,001** (0,0004)	0,001 (0,001)
		distnaogeoeludeana	4,428 (2,998)			11,903 (20,808)		
distnaogeoecosseno			1,512 (4,624)			3,920 (5,161)		
distnaogeoquiq				2,608 (1,601)			5,524 (5,058)	
Int_distnaogeoeludeana		0,001 (0,001)			-0,00001 (0,004)			
Int_distnaogeoecosseno			0,003 (0,002)			0,0006 (0,002)		
Int_distnaogeoquiq				-0,0006 (0,0006)			-0,0005 (0,0009)	
capabsoremp		-5,291 (7,211)	-6,319 (5,848)	-10,332* (5,301)	-18,489 (24,652)	-10,651 (7,270)	-6,549 (12,241)	
Inquantempresas		-0,644** (0,306)	-0,551** (0,244)	-0,421* (0,256)	-0,492 (0,628)	-0,416 (0,269)	-0,367 (0,351)	
Inquantartigossuniv		-0,978*** (0,256)	-1,015*** (0,246)	-1,028*** (0,252)	-1,069 (0,709)	-0,940*** (0,242)	-1,165*** (0,397)	
pibpercapitadiferenca					0,00002 (0,00005)	-0,000002 (0,00002)	-0,00006 (0,00007)	
denspatentesdiferenca					-2,773 (3,248)	-1,293 (0,934)	-1,563* (0,908)	
Constante		13,412***	13,955***	12,130***	11,379**	12,419***	10,981**	
Teste superdispersão (ln alpha)		0,412***	0,623***	0,333***	0,515***	0,564***	0,392**	
Teste Vuong		1,61*	1,70**	1,86**	2,26**	1,94**	2,83***	

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Fonte: Elaboração Própria

5 CONCLUSÃO

Este estudo teve o objetivo de propor uma classificação que organiza a perspectiva multidimensional da proximidade e realizar uma aplicação empírica. Para tanto, foram levantadas as definições conceituais para a proximidade e as principais tentativas empíricas de medir suas dimensões.

Alguns estudos avaliaram o conceito proximidade quanto à sua perspectiva multidimensional e à ambiguidade conceitual relacionada às formas não espaciais de proximidade. Uma das propostas mais importantes definiu quatro dimensões: cognitiva, organizacional, institucional e social (TORRE; GILLY, 2000; KIRAT; LUNG, 1999; TORRE *et al.*, 2005; BOSCHMA, 2005; KNOBEN; OERLEMANS, 2006).

Outros autores ocuparam-se com tentativas empíricas de medição das dimensões da proximidade, tanto para analisar os seus inter-relacionamentos, quanto para avaliar a sua influência conjunta em outras variáveis (JAFFE, 1989; BRESCHI; LISSONI; MALERBA, 2003; BOSCHMA; IAMMARINO, 2009; BOSCHMA; FRENKEN, 2011). Porém, ainda há campo para novos desenvolvimentos quanto a esses indicadores para medir as várias dimensões da proximidade.

A necessidade de se renovarem os esforços para abrir novas possibilidades de medir, representar e analisar as dimensões da proximidade fica exposta com as limitações de interpretação do índice cosseno (JAFFE, 1989; BRESCHI; LISSONI; MALERBA, 2003) e da variedade relacionada (FRENKEN *et al.*, 2007; BOSCHMA; IAMMARINO, 2009; BOSCHMA; FRENKEN, 2011) como distância.

Tanto o índice cosseno quanto a variedade relacionada não são diretamente interpretáveis como distância, pois não necessariamente atendem axiomas que definem a distância no sentido métrico conforme os autores desejariam. Além disso, a forma como as medidas são estabelecidas para a sua utilização nos modelos estatísticos é de fundamental relevância, já que podem ter influência direta nos resultados inferenciais obtidos.

Neste contexto, a criação de uma classificação direciona a escolha entre diferentes técnicas para apoiar a medição das dimensões da proximidade. Além disso, outras

formas para a interpretação e análise desses indicadores surgem para lidar com a sensibilidade dos resultados inferenciais obtidos com os modelos.

A elaboração da classificação foi justificada por três argumentos que apresentam a lógica para formação de um quadro resumo dos conceitos. Os argumentos levantam e direcionam questões mais básicas sobre o que e como medir, além da indagação de como representar e analisar as dimensões da proximidade.

O primeiro argumento trata do questionamento sobre o que medir e discute a natureza da distância no sentido mais formal como uma função em seus diferentes contextos. Esse argumento possibilitou uma revisão das medidas de similaridade e dissimilaridade e aspectos relevantes para propor classes de medidas em três famílias. Algumas medidas para cada família são apresentadas nos quadros 6, 7 e 8, respectivamente denominadas por: Família 1 (com base em lei de potência), Família 2 (com base no produto interno) e Família 3 (com base em lei de distribuição).

Considera-se que poucas das medidas levantadas e classificadas nessas três famílias foram exploradas pelos estudos da área, ao se avaliarem as propostas mais utilizadas na literatura. Além disso, quanto às restrições dos índices cosseno e variedade relacionada em relação à representação como distância métrica, nem sempre esses índices são utilizados com uma reflexão maior sobre a forma de cálculo e com as implicações quanto ao seu uso como instrumento de medição das dimensões não espaciais da proximidade.

O segundo argumento especifica a natureza dos dados com a questão de como medir a proximidade. Em suma, justifica-se a medição da proximidade com base nas características teóricas definidas pelas classificações (variáveis categorizadas), em que as unidades de análise do estudo (pessoas, empresas etc.) são classificadas. Consequentemente, revela-se como as unidades podem ser comparadas e, a partir disso, de que modo as medidas para cada dimensão da proximidade podem ser obtidas.

O terceiro argumento lida com a abordagem analítica, respondendo a questão sobre como representar e analisar as medidas obtidas para as dimensões da proximidade. O esquema indica o que e como medir e, por fim, as técnicas analíticas, com base em

duas perspectivas: algébrica e geométrica. Na abordagem algébrica, apresentam-se como opção analítica os modelos loglineares e, na geométrica, há a indicação de uso das técnicas multivariadas, como Escalonamento Multidimensional e Análise de Correspondência.

Uma aplicação dos conceitos e das técnicas apresentados na classificação foi realizada com dados reais. O foco de aplicação foi na literatura que estuda a influência das dimensões da proximidade na interação universidade e empresa, de modo a captar o potencial papel dos efeitos de transbordamentos. A análise avaliou tanto os limites do ganho com a proximidade quanto aos mecanismos de complementariedade e de substituição entre as dimensões geográfica e cognitiva.

O foco na dimensão cognitiva favoreceu a comparação entre diferentes formas para a sua medição. Em específico, a dimensão cognitiva para a proximidade foi medida por indicadores nas três famílias apresentadas; foram utilizadas, ainda, as técnicas analíticas levantadas na perspectiva geométrica: Escalonamento Multidimensional e a Análise de Correspondência. Com as medidas, foi feita a posterior verificação empírica da importância dessa definição da proximidade com a avaliação da significância dos parâmetros em modelos estimados.

Os resultados, observados pela comparação dos resultados nos modelos estimados, revelam a importância necessária à forma como as medidas são definidas e representadas. Consideram-se as mesorregiões como unidades de análise e as características internas e externas que motivam o interesse das empresas em interagir com grupos de pesquisa nas universidades como variáveis preditoras dessas interações conforme literatura (CAPELLO; CARAGLIU, 2015).

Ressalta-se que as expectativas de inferência estão de acordo com a literatura, exceto as conclusões em relação, especificamente, às três diferentes variações de medida para a dimensão cognitiva, o que revela a sensibilidade da escolha da forma como a proximidade é definida nos estudos já realizados sobre o tema.

Assim, espera-se que novos estudos possam ser direcionados com a utilização das diferentes medidas levantadas na classificação e das técnicas para representação e análise dessas medidas. As técnicas analíticas levantadas e classificadas na proposta

se mostram, ainda, pouco exploradas na literatura sobre o tema proximidade em geografia econômica, podendo ser consideradas como novas linhas para exploração em estudos futuros.

Inicialmente, as contribuições desse estudo são de caráter metodológico, pois ele apresenta uma reflexão sobre as técnicas específicas de medição, representação e análise. Porém, para além da disseminação das técnicas apresentadas, há, também, a contribuição prática de direcionar algumas linhas de pesquisa para estudos. Tal direcionamento é apresentado com base nas sugestões organizadas na forma de uma classificação.

Ainda do ponto de vista das possibilidades posteriores, sugere-se o uso da classificação proposta para aplicações que considerem a perspectiva multidimensional da proximidade. Para tanto, pode-se especificar quais seriam as características mais relevantes para serem consideradas como *proxies* para cada dimensão da proximidade e, com uma forma conjunta, comparar todas essas dimensões, tanto do ponto de vista da sua medição quanto da sua representação. O uso de técnicas analíticas reconhecidas e utilizadas em diversas áreas pode apoiar essa tarefa.

Por fim, a própria classificação apresentada pode ser adaptada para incluir outras técnicas analíticas que incorporem dados não espaciais, como análise de redes, de árvores e de agrupamentos, além de outras técnicas que podem lidar com dados de não proximidade.

Referências Bibliográficas⁵

ACOSTA, M.; CORONADO, D.; FERRÁNDIZ, E.; León, M. D. Factors affecting inter-regional academic scientific collaboration within Europe: The role of economic distance. **Scientometrics**, v. 87, n. 1, p. 63-74, 2011.

AGRAWAL, A.; KAPUR, D.; McHALE, J. How do spatial and social proximity influence knowledge flows? Evidence from patent data. **Journal of Urban Economics**, v. 64, n. 2, p. 258–269, 2008.

AGRESTI, A. **Categorical Data Analysis**, 2nd ed. Wiley, New York, 2002.

ARAÚJO, B. C.; CAVALCANTE, L. R.; ALVES, P. 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**, Brasília, n. 5, p. 16-21, 2009.

ARUNDEL A; GEUNA A. Proximity and the use of public science by innovative European firms. **Economics of Innovation and New Technologies**, n. 13, p. 559-580, 2004.

AUDRETSCH, D. **Corporate form and spatial form**, in CLARK G. L., FELDMAN M. P. and GERTLER M. S. (Eds) *The Oxford Handbook of Economic Geography*, 333–347. Oxford University Press, Oxford, 2000.

AUDRETSCH, D.; FELDMAN, M. R&D Spillovers and the Geography of Innovation and Production. **The American Economic Review**, v. 86, n. 3, p. 630-640, 1996.

AUTANT-BERNARD, C.; LESAGE J. Quantifying Knowledge Spillovers Using Spatial Econometric Models. **Journal of Regional Science**, v. 51, p. 471–96, 2011.

BEH, E. J. Simple correspondence analysis: A bibliographic review. **International Statistical Review**, v. 72, n. 2, p. 257–284, 2004.

BENZECRI, J. P. *L'Analyse des donnees*. Paris: Dunod, 1973.

BISHOP, K.; D'ESTE, P.; NEELY, A. Gaining from interactions with universities: Multiple methods for nurturing absorptive capacity. **Research Policy**, v. 40, n. 1, p. 30-40, 2011.

BORG, I.; GROENEN, P. J. F. **Modern Multidimensional Scaling: theory and applications**, New York: Springer, 1997.

⁵ De acordo com a Associação Brasileira de Normas Técnicas (ABNT NBR 6023).

BOSCHMA, R. Proximity and Innovation : A Critical Assessment. **Regional Studies**, v. 39, n. 1, p. 61–74, 2005.

BOSCHMA, R. A.; FRENKEN, K. The spatial evolution of innovation networks. A proximity perspective. In: Boschma RA, Martin R (eds) **Handbook of evolutionary economic geography**, chapter 5. Edward Elgar, Cheltenham, 2010.

BOSCHMA, R.; FRENKEN, K. The emerging empirics of evolutionary economic geography. **Journal of Economic Geography**, v. 11, n. 2, p. 295–307, 2011.

BOSCHMA, R.; IAMMARINO, S. Related variety, trade linkages, and regional growth in Italy. **Economic Geography**, v. 85, n. 3, p. 289–311, 2009.

BRESCHI, S.; LISSONI, F.; MALERBA, F. Knowledge-relatedness in firm technological diversification. **Research Policy**, v. 32, n. 1, p. 69–87, 2003.

BRESCHI, S.; LISSONI, F. Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. **Journal of Economic Geography**, v. 9, n. 4, p. 439–468, 2009.

BROEKEL, T.; BOSCHMA, R. Knowledge networks in the Dutch aviation industry: the proximity paradox. **Journal of Economic Geography**, v. 12, n. 2, p. 409–433, 2012.

BROWN, J. S.; DUGUID, P. Organizational learning and communities of practice: toward a unified view of working, learning and innovation. **Organization Science**, v. 2, n. 1, p. 40–57, 1991.

BURMEISTER, A.; COLLETIS-WAHL, K. Proximity in production networks: the circulatory dimension. **European Urban and Regional Studies**, v. 4, n. 3, p. 231–241, 1997.

BUSING, F. M. T. A.; VAN DEUN, K. **Unfolding Degeneracies' History**. In K. Van Deun, Degeneracies in multidimensional unfolding. Unpublished doctoral dissertation, Catholic University Leuven, 2005.

CAMERON, A.; TRIVEDI, P. **Microeconometrics using stata**. College Station, Texas: Stata Press, 2009.

CAMERON, A.; TRIVEDI, P. **Regression Analysis of count data**. Cambridge: Cambridge University Press, 2013.

CAPELLO, R. **Proximity and Regional Innovation Processes: Is There Space for New Reflections?** In Regional Development and Proximity Relations, edited by A. Torre and F. Wallet, 163–94. London, UK: Edward Elgar, 2014.

CAPELLO, R.; CARAGLIU, A. Proximities and the Intensity of Scientific Relations: Synergies and Non-linearities. **International Regional Science Review**, v. 1, p. 1–38, 2015.

CARROLL, J. D.; ARABIE, P. **Multidimensional Scaling**, Annual Review of Psychology, v. 31, p. 607-649, 1980.

CARROLL, J. D.; GREEN, P. E.; SCHAFFER, C.M. Interpoint distance comparison in correspondence analysis. **Journal of Marketing Research**, v. 23, p. 271–280, 1986.

CARROLL, J. D.; GREEN, P. E.; SCHAFFER, C. M. Comparing interpoint distance in correspondence analysis: A clarification. **Journal of Marketing Research**, v. 24, p. 445–450, 1987.

CARROLL, J. D.; GREEN, P. E.; SCHAFFER, C. M. Reply to Greenacre's commentary on the Carroll-Green-Schaffer scaling of two-way correspondence analysis solutions. **Journal of Marketing Research**, v. 26, p. 366–368, 1989.

CASSI, L.; PLUNKET, A. Proximity, network formation and inventive performance: in search of the proximity paradox. **The Annals of Regional Science**, p. 395–422, 2014.

CHA, S. -H., Comprehensive survey on distance/similarity measures between probability density functions. **International Journal of Mathematical Models and Methods in Applied Sciences**, v. 4, n. 1, p. 300–307, 2007.

CLOGG, C. C.; ELIASON, S. R.; WAHL, R. J. Labor-market experiences and labor force outcomes. **American Journal of Sociology**, v. 95, p. 1536–1576, 1990.

COENEN, L.; MOODYSSON, J.; ASHEIM, B. T. Nodes, networks and proximities: on the knowledge dynamics of the Medicon Valley biotech cluster. **European Planning Studies**, v. 12, n. 7, p. 1003–1018, 2004.

COHEN, W. M.; LEVINTHAL, D. A. Absorptive capacity: a new perspective on learning and innovation. **Administrative Science Quarterly**, v. 35, n. 1, p. 128–152, 1990.

COSTA, A. B. da; RUFFONI, J.; PUFFAL, D. Proximidade Geográfica e Interação Universidade-Empresa no Rio Grande do Sul. **Revista de Economia (Curitiba)**, v. 37, p. 213- 238, 2011.

De FUENTES C; DUTRÉNIT G. Geographic proximity and university–industry interaction: The case of Mexico. **The Journal of Technology Transfer**, v. 41, n. 2, p. 329-348, 2016.

DEZA, M. M.; DEZA, E. **Encyclopedia of Distances**. Berlin, Germany: Springer-Verlag, 2009.

D'ESTE, P.; IAMMARINO, S. The spatial profile of university-business research partnerships. **Papers in Regional Science**, v. 89, n. 2, p. 336-350, 2010.

DURANTON, G.; PUGA, D. Diversity and Specialisation in Cities: Why. Where and When Does it Matter? **Urban Studies**, v. 37, n. 3, p. 533-555, 2000.

ECKART, C.; YOUNG, G. Approximation of one matrix by another of lower rank. **Psychometrika**, v. 1, p. 211-218, 1936.

ENGELSMAN, E. C.; RAAN, A. F. J. VAN. Cartography of technology. **Research Policy**, v. 31, n. 1, p. 1–26, 1994.

ENKEL, E.; HEIL, S. Preparing for distant collaboration: Antecedents to potential absorptive capacity in cross-industry innovation. **Technovation**, v. 34, n. 4, p. 242–260, 2014.

FAFCHAMPS, M.; GOYAL, S.; VAN DER LEIJ, M. J. Matching and network effects. **Journal of the European Economic Association**, v. 8, n. 1, p. 203–231, 2010.

FAUST, K.; WASSERMAN, S. **Association and correlational models for studying measurements on ordinal relations**. In P.V. Marsden (Ed.), *Sociological methodology*, 177-215. Cambridge, MA: Basil Blackwell, 1993.

FERNÁNDEZ, A.; FERRÁNDIZ, E.; LEÓN, M. D. Proximity dimensions and scientific collaboration among academic institutions in Europe: The closer, the better? **Scientometrics**, v. 106, n. 3, p. 1073–1092, 2016.

FILIPPI, M.; TORRE, A. Local organisations and institutions. How can geographical proximity be activated by collective projects? **International Journal of Technology Management**, v. 26, n. 26, p. 386–400, 2003.

FRENKEN, K.; PONDS, R.; VAN OORT, F. The Citation Impact of Research Collaboration in Science-based Industries: A Spatial-institutional Analysis. *Papers in Regional Science*, p. 89:351–71, 2010.

FRENKEN K.; VAN OORT F.; VERBURG T. Related variety, unrelated variety and regional economic growth, **Regional Studies**, v. 41, p. 685–697, 2007.

FRITSCH M.; FRANKE G. Innovation, regional knowledge spillovers and R&D cooperation, **Research Policy**, v. 33, p. 245–255, 2004.

GALTON, F. Co-relations and their measurement, chiefly from anthropometric data. **Proceedings of the Royal Society of London**, v. 45, p. 135-145, 1888.

GARCIA, R.; ARAUJO, V.; MASCARINI, S.; SANTOS, E. Os efeitos da proximidade geográfica para o estímulo da interação universidade-empresa. **Revista de Economia**, Curitiba, v. 37, n. 4, p. 307-330, 2011.

GARCIA, R.; ARAUJO, V.; MASCARINI, S.; SANTOS, E. Efeitos da qualidade da pesquisa acadêmica sobre a distância geográfica das interações universidade-empresa. **Estudo Econômico**, São Paulo, v. 44, n. 1, p. 105-132, 2014.

GELDES, C.; FELZENSZTEIN, C; TURKINA, E; DURAND, A. How does proximity affect interfirm marketing cooperation? A study of an agribusiness cluster. **Journal of Business Research**, v. 68, n. 2, p. 263–272, 2015.

GERTLER, M. S. Being there – proximity, organization, and culture in the development and adoption of advanced manufacturing technologies. **Economic Geography**, v. 71, n. 1, p. 1–26, 1995.

GERTLER M. S. **Buzz without being there: communities of practice in context**, in AMIN A. and ROBERTS J. (Eds) *Communities of Practice: Community, Economic Creativity, and Organization*, pp. 203–226. Oxford University Press, Oxford, 2008.

GIFI, A. **Nonlinear Multivariate Analysis**. New York: Wiley, 1990.

GIULIANI, E; BELL, M. The Micro-Determinants of Meso-Level Learning and Innovation: evidence from a chilean wine cluster. **Research Policy**, v. 34, p. 47-68, 2005.

GLAESER, E. L.; KALLAL, H. D.; SCHEINKMAN, J. A.; SHLEIFER, A. Growth of cities. **Journal of Political Economy**, v. 100, n. 6, p. 1126-1152, 1992.

GONÇALVES, E.; FAJARDO, B. DE A. G. A influência da proximidade tecnológica e geográfica sobre a inovação regional no Brasil. **Revista de Economia Contemporânea**, v. 15, n. 1, p. 112–142, 2011.

GOODMAN, L. A. Simple Models for the Analysis of Association in Cross classifications Having Ordered Categories. **Journal of the American Statistical Association** v. 74, p. 537-52, 1979.

GOWER, J. C. Some distance properties of latent root and vector methods used in multivariate analysis. **Biometrika**, v. 53, p. 325-338, 1966.

GRANOVETTER, M. S. Economic action and social structure: the problem of embeddedness. **American Journal of Sociology**, v. 91, n. 3, p. 481–510, 1985

GREENACRE, M. J. **Theory and Application of Correspondence Analysis**. London: Academic Press, 1984.

GREENACRE, M. J. **Correspondence analysis in proctice**. London: Academic Press, 1994.

GREENACRE, M. J. Correspondence analysis of multivariate categorical data by weighted least-squares. **Biometrika**, v. 75, p. 457–467, 1988.

GREENACRE, M.; BLASIUS, J. (Eds.). **Correspondence analysis in the social sciences**. London: Academic Press, 1994.

GREENACRE, M. J. The Carroll–Green–Schaffer scaling in correspondence analysis: a theoretical and empirical appraisal. **Journal of Marketing Research**, v. 26, p. 358–365, 1989.

GREENACRE, M. J. Some limitations of multiple correspondence analysis. **Computational Statistics Quarterly**, v. 3, p. 249-256, 1990.

GREENACRE, M. J. Interpreting multiple correspondence analysis. **Applied Stochastic Models and Data Analysis**, v. 7, p. 195–210, 1991.

GUTTMAN, L. **The quantification of a class of attributes: A theory and method of scale construction**. In P. Horst *et al.* (Eds.), The prediction of personal adjustment, 319-348. New York: Social Science Research Council, 1941.

HANSEN, T. Substitution or Overlap? The Relations between Geographical and Non-spatial Proximity Dimensions in Collaborative Innovation Projects. **Regional Studies**, v. 49, n. 10, p. 1672–1684, 2014.

HEISER, W. J. Geometric representation of association between categories. **Psychometrika**, v. 69, n. 4, p. 513-545, 2004.

HEISER, W. J.; MEULMAN, J. J. Analyzing rectangular tables by joint and constrained MDS. **Journal of Econometrics**, v. 22, p. 193-167, 1983.

HILL, M. O.; GAUCH JR., H. G. Detrended correspondence analysis: an improved ordination technique. **Vegetatio**, v. 42, p. 47–58, 1980.

HIRSCHFELD, H. O. A connection between correlation and contingency. **Proceedings Cambridge Philosophical Society**, v. 31, p. 520-524, 1935.

HOWELLS, J. R. L. Tacit knowledge, innovation and economic geography. **Urban Studies**, v. 39, n. 5–6, p. 871– 884, 2002.

HUBER, F. On the Role and Interrelationship of Spatial, Social and Cognitive Proximity: Personal Knowledge Relationships of R&D Workers in the Cambridge Information Technology Cluster. **Regional Studies**, v. 46, n. 9, p. 1169–1182, 2012.

HWANG, K. International collaboration in multilayered center-Periphery in the globalization of science and technology. **Science Technology Human Values**, v. 33, p. 101–133, 2008.

JAFFE, A. B. Characterizing the “technological position” of firms, with application to quantifying technological opportunity and research spillovers. **Research Policy**, v. 18, n. 1, p. 87–97, 1989.

JAFFE, A.; TRAJTENBERG, M.; HENDERSON, R. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, **Quarterly Journal of Economics**, v. 108, p. 577-598, 1993.

KATERI, M. **Contingency Tables: Methods and Implementation Using R**, Springer, New York, 2014.

KIRAT, T.; LUNG, Y. Innovation and Proximity: Territories as Loci of Collective

Learning Processes. **European Urban and Regional Studies**, v. 6, n. 1, p. 27–38, 1999.

KLEVORICK, A.; LEVIN, R.; NELSON, R.; WINTER, S. On the sources and significance of inter-industry differences in technological opportunities. **Research Policy** v. 24, p. 185-205, 1995.

KNOBEN, J.; OERLEMANS, L. A. G. Proximity and inter-organizational collaboration: A literature review. **International Journal of Management Reviews**, v. 8, n. 2, p. 71–89, 2006.

KOURTIT, K. Super-Proximity and Spatial Development. **Investigaciones Regionales**, v. 36, n. 36, p. 215–231, 2016.

KRAFFT, J.; QUATRARO, F.; SAVIOTI, P. P. The Dynamics of Knowledge-intensive Sectors' Knowledge Base: Evidence from Biotechnology and Telecommunications. **Industry and Innovation**, v. 21, n. 3, p. 215-242, 2014.

LAURSEN K., REICHSTEIN T., SALTER, A. Exploring the effect of geographical proximity and university quality on university-industry collaboration in the United Kingdom. **Regional Studies**, v. 45, n. 4, p. 507–523, 2011.

LEBART L.; MARINEAU A.; WARWICK, K. M. **Multivariate descriptive statistical analysis**. New York: Wiley, 1984.

MAGGIONI, M.; NOSVELLI, M.; UBERTI, E. Space versus Networks in the Geography of Innovation: A European Analysis. **Papers in Regional Science**, v. 86, p. 471–93, 2007.

MANSFIELD E., LEE J. The modern university: contributor to industrial innovation and receipt of industrial R&D support. **Research Policy**, v. 25, p. 1047–1058, 1996.

MAREK, P.; TITZE, M.; FUHRMEISTER, C.; BLUM, U. R&D collaborations and the role of proximity. **Regional Studies**, p. 1-13, 2016.

MARROCU, E.; PACI, R.; USAI, S. Proximity, networking and knowledge production in Europe: What lessons for innovation policy? **Technological Forecasting and Social Change**, v. 80, n. 8, p. 1484–1498, 2013.

MARSHALL, A. **Principles of Economics**, The Royal Economic Society, Mac Millan, 9th ed., London, 1961.

MATTES, J. Dimensions of Proximity and Knowledge Bases: Innovation between Spatial and Non-spatial Factors. **Regional Studies**, v. 46, n. 8, p. 1085–1099, 2012.

MCCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**, 2nd ed. Chapman and Hall, New York, 1989.

METCALFE, S. Equilibrium and evolutionary foundations of competition and technology policy: new perspectives on the division of labour and the innovation process. **Revista Brasileira de Inovação**, v. 2, n. 1, p. 111-146, 2003.

MORA, T.; MORENO, R. Specialisation Changes in European Regions: The Role Played by Externalities across Regions. **Journal of Geographical Systems**, v. 12, p. 311-34, 2010.

MOWERY, D. C.; OXLEY, J. E.; SILVERMAN, B. S. Technological overlap and interfirm cooperation: implications for the resource-based view of the firm. *Res Policy*, v. 27, n. 5, p. 507-523, 1998.

MUSCIO, A.; POZZALI, A. The effects of cognitive distance in university-industry collaborations: Some evidence from Italian universities. **Journal of Technology Transfer**, v. 38, n. 4, p. 486-508, 2013.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized Linear Models. **Journal of the Royal Statistical Society. Series A (General)**, v. 135, n. 3, p. 370-384, 1972.

NISHISATO, S. **Analysis of categorical data: Dual scaling and its applications**. Toronto: University of Toronto Press, 1980.

NISHISATO, S. **Elements of dual scaling: An introduction to practical data analysis**. Hillsdale, NJ: Erlbaum, 1994.

NOOTEBOOM, B. Innovation and inter-firm linkages: new implications for policy. **Research Policy**, v. 28, n. 8, p. 793-805, 1999.

NOOTEBOOM, B. **Learning and Innovation in Organizations and Economies**. Oxford: Oxford University Press, 2000.

NORTH, D.C. Institutions. **Journal of Economic Perspectives**, v. 5, n. 1, p. 97-112, 1991.

NOSOFSKY, R. M. Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. **Perception & Psychophysics**, v. 38, p. 415-432, 1985.

OERLEMANS, L. A. G.; MEEUS, M. T. H. Do organisational and spatial proximity impact on firm performance? **Regional Studies**, v. 39, n. 1, p. 89-104 2005.

PACI, R.; MARROCU, E.; USAI, S. The Complementary Effects of Proximity Dimensions on Knowledge Spillovers. **Spatial Economic Analysis**, v. 9, n. 1, p. 9-30, 2014.

PRESUTTI, M.; BOARI, C.; MAJOCCHI, A. The Importance of Proximity for the Start-Ups' Knowledge Acquisition and Exploitation. **Journal of Small Business Management**, v. 49, n. 3, p. 361-389, 2011.

RALLET, A.; TORRE, A. Is geographical proximity necessary in the innovation networks in the era of global economy? **Geo Journal**, v. 49, n. 4, p. 373–380, 1999.

RAPINI, M.; ALBUQUERQUE, E.; CHAVES, C.; SILVA, L.; SOUZA, S.; RIGHI, H.; CRUZ, W. University-industry interactions in an immature system of innovation: evidence from Minas Gerais, Brazil. **Science and Public Policy** v. 36, p. 373-386, 2009.

RUFFONI, J.; TATSCH, A. S.; STEFANI, R.; SCHAEFFER, P. R.; GRINGS, L. G. Does the Geographical Proximity Matter in Knowledge and Information Flow? A study about wine cluster in the Southern region of Brazil. **Rev. Econ. Sociol. Rural** v. 55, n. 3, 2017.

RICHARDSON, M.; KUDER, G. F. Making a rating scale that measures. **Personnel Journal**, v. 12, p. 36-40, 1933.

ROOIJ, M.; HEISER, W. J. Graphical representations and odds ratios in a distance-association model for the analysis of cross-classified data. **Psychometrika**, v. 70, n. 1, p. 99-122, 2005.

SCHAMP, E. W.; RENTMEISTER, B.; LO, V. Dimensions of proximity in knowledge-based networks: the cases of investment banking and automobile design. **European Planning Studies**, v. 12, n. 5, p. 607–624, 2004.

SCHARTINGER, D.; RAMMER, C.; FRÖHLICH, J. Knowledge interactions between universities and industry in Austria: Sectoral patterns and determinants. **Research Policy**, v. 31, p. 303–328, 2002.

SHEPARD, R. N. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. **Psychometrika**, v. 22, p. 325–345, 1957.

SONNENWALD, D. H. Scientific collaboration: A synthesis of challenges and strategies. In B. Cronin (Ed.), **Annual Review of Information Science and Technology** (Vol. 41). Medford, NJ: Information Today Inc, 2007.

SORENSEN, O.; RIVKIN, J. W.; FLEMING, L. Complexity, networks and knowledge flows. **Research Policy**, v. 35, p. 994–1017, 2006.

SPEARMAN, C.E. 'Footrule' for measuring correlation. **British Journal of Psychology**, v. 2, p. 89-108, 1906.

STEVENS, S. S. On the theory of scales of measurement. **Science**, v. 103, p. 677-680, 1946.

STORPER, M.; VENABLES, A. J. Buzz: face-to-face contact and the urban economy. **Journal of economic geography**, v. 4, n. 4, p. 351-370, 2004.

SUZIGAN, W.; ALBUQUERQUE, E.; GARCIA, R.; RAPINI, M. University and industry linkages in Brazil: some preliminary and descriptive results. **Seoul Journal of Economics**, v. 22, p. 591-611, 2009.

TAKANE, Y. Analysis of contingency tables by ideal point discriminant analysis. **Psychometrika**, v. 52, p. 493-513, 1987.

TAKANE, Y.; SHIBAYAMA, T. **Structures in stimulus identification data**. In F. G., Ashby (Ed.), Probabilistic multidimensional models of perception and cognition, 335-362, Hillsdale, NJ: Earlbaum, 1992.

TER BRAAK, C. J. F. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. **Ecology**, v. 67, p. 1167–1179, 1986.

TER BRAAK, C. J. F. **Ordination**. In **Data Analysis in Community and Landscape Ecology**, Eds. R.H. Jongman, C.F.J. ter Braak and O.F.R. van Tongeren, pp. 91–173. Wageningen: Pudoc, 1987.

THURSTONE, L. L. A law of comparative judgment. *Psychological Review*, v. 34, p. 273-286, 1927.

TORGERSON, W. S. Multidimensional scaling: I. Theory and method. **Psychometrika**, v. 17, p. 401-409, 1952.

TORGERSON, W. S. **Theory and methods of scaling**. New York: Wiley, 1958.

TORRE, A. *et al.* Proximity and Localization. **Regional Studies**, v. 39, n. 1, p. 47–59, 2005.

TORRE, A.; GILLY, J.-P. On the Analytical Dimension of Proximity Dynamics. **Regional Studies**, v. 34, n. 2, p. 169–180, 2000

TORRE, A.; WALLET, F. **Regional Development and Proximity Relations**. Cheltenham, UK: Edward Elgar, 2014.

TUSHMAN, M. L.; ANDERSON, P. Technological discontinuities and organizational environments. **Administrative Science Quarterly**, v. 31, n. 3, p. 439 – 465, 1986.

VON THUNEN, J.H. Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie, Schumacher-Zarchlin H., **Wiegandt, Hempel und Parey**, 1826.

WELLER-FAHY, D.; BORGHETTI, B. J.; SODERMANN, A. A. A survey of distance and similarity measures used within network intrusion anomaly detection. **IEEE Communication Surveys & Tutorials**, v. 17, n. 1, p. 70–91, 2014.

WERKER, C.; OOMS, W.; CANIËLS, M. C. J. Personal and related kinds of proximity driving collaborations: a multi-case study of Dutch nanotechnology researchers. **SpringerPlus**, v. 5, n. 1, p. 1751, 2016.

WETERINGS, A.; BOSCHMA, R. Does spatial proximity to costumers matter for innovative performance? Evidence from the Dutch software sector. **Research Policy**, v. 38, n. 5, p. 746-755, 2009.

WUYTS, S. *et al.* Empirical tests of optimal cognitive distance. **Journal of Economic Behavior and Organization**, v. 58, n. 2, p. 277–302, 2005.

YOUNG, G.; HOUSEHOLDER, A. S. Discussion of a set of point in terms of their mutual distances. **Psychometrika**, v. 3, p. 19-22, 1938.

YULE, G.U. On the association of attributes in statistics. **Philosophical Transactions of the Royal Society of London (A)**, v. 194, p. 257-319, 1900.

Apêndice B – Resumo das inferências estimadas pelos modelos

	Tabela 8 e 9																Tabela 11 e 12 - Limites nos ganhos																Tabela 13 e 14 - Complementariedade e Substituição															
	Distâncias e características Univ. Emp.																Distâncias e características Univ. Emp.																Distâncias e características Univ. Emp.															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44				
distgeo	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
distnaogeoEuclidean	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
distnaogeoCosseno	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
distnaogeoQuiq	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
distgeo2																	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			
distnaogeoEuclidean2																	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			
distnaogeoCosseno2																	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			
distnaogeoQuiq2																	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			
Int_distnaogeoEuclidean																																																
Int_distnaogeoCosseno																																																
Int_distnaogeoQuiq																																																
capabsoremp					+	+	+	+	+	+	+	+	+	+	+	+					+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			
Inquantempresas									+	+	+	+	+	+	+	+					+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
Inquantartigosuniv									+	+	+	+	+	+	+	+					+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
pibpercapitadiferenca									-	-	-	-	-	-	-	-																																
denspatentesdiferenca									-	-	-	-	-	-	-	-																																

Fonte: Elaboração própria