

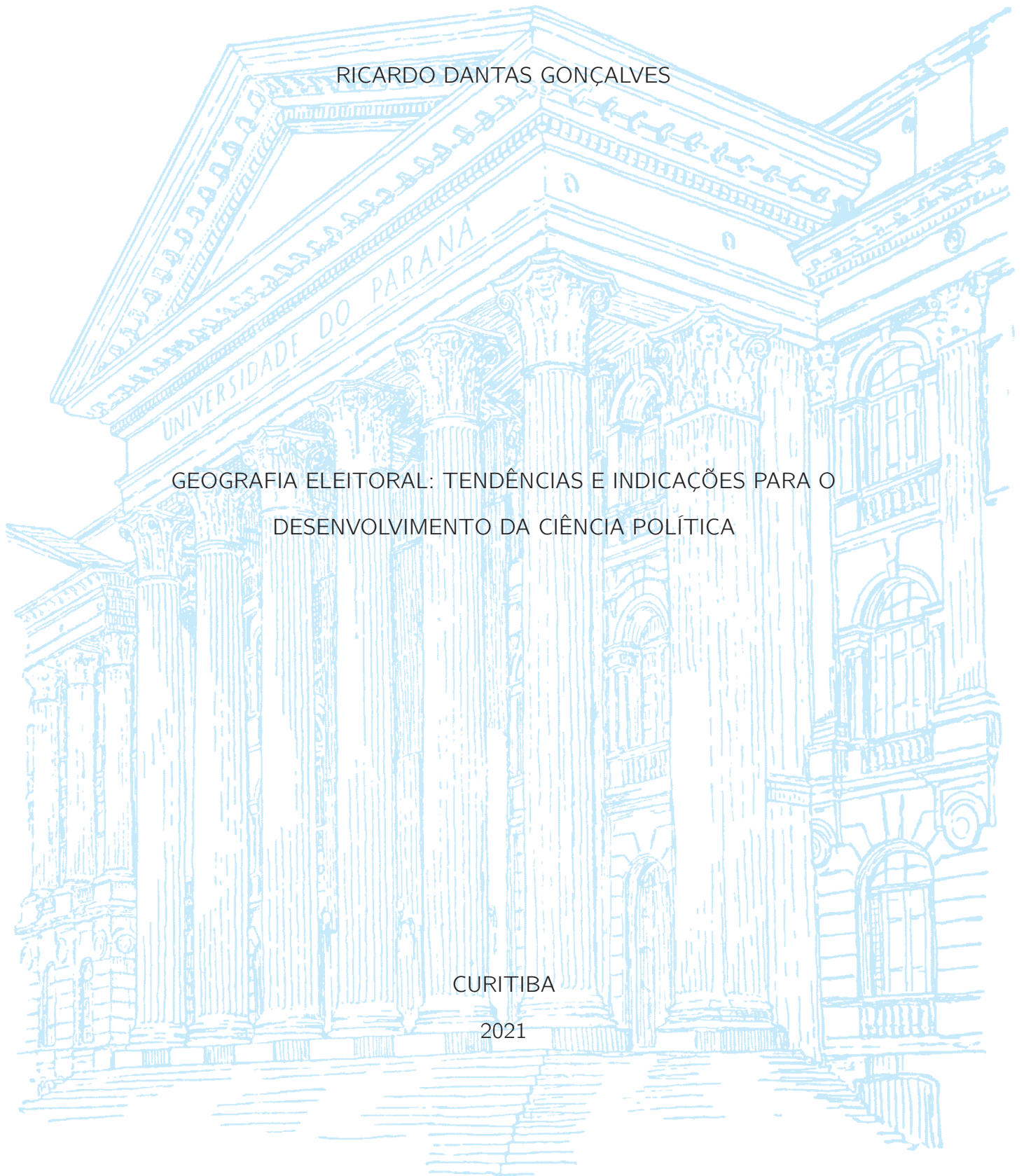
UNIVERSIDADE FEDERAL DO PARANÁ

RICARDO DANTAS GONÇALVES

GEOGRAFIA ELEITORAL: TENDÊNCIAS E INDICAÇÕES PARA O  
DESENVOLVIMENTO DA CIÊNCIA POLÍTICA

CURITIBA

2021



RICARDO DANTAS GONÇALVES

GEOGRAFIA ELEITORAL: TENDÊNCIAS E INDICAÇÕES PARA O  
DESENVOLVIMENTO DA CIÊNCIA POLÍTICA

Tese apresentada como requisito parcial à obtenção do grau de Doutor em Ciência Política no Programa de Pós Graduação em Ciência Política do Setor de Ciências Humanas, Letras e Artes da Universidade Federal do Paraná.

Área de concentração: *Comportamento, Representação e Comunicação Política.*

Orientador: Prof<sup>o</sup> Dr. Rodrigo Rossi Horochovski.

Coorientador: Prof<sup>o</sup> Dr. Ivan Jairo Junckes.

CURITIBA

2021

FICHA CATALOGRÁFICA ELABORADA PELO SISTEMA DE BIBLIOTECAS/UFPR –  
BIBLIOTECA DE CIÊNCIAS HUMANAS COM OS DADOS FORNECIDOS PELO AUTOR

Aparecida Noeli Furquim Geffer – CRB 9/1309

Gonçalves, Ricardo Dantas

Geografia eleitoral: tendências e indicações para o desenvolvimento da ciência política / Ricardo Dantas Gonçalves – Curitiba, 2021.

Tese (Doutorado em Ciências Política) – Setor de Ciências Humanas da  
Universidade Federal do Paraná.

Orientador : Prof. Dr. Rodrigo Rossi Horochovski

Coorientador : Prof. Dr. Ivan Jairo Junckes

1. Geografia . 2. Eleições. 3. Pesquisa eleitoral. 4. Ciência Política. 5. Pesquisa  
– metodologia. I. Horochovski, Rodrigo Rossi. II. Junckes, Ivan Jairo. III. Título.

CDD – 324.981

## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIA POLÍTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **RICARDO DANTAS GONÇALVES** intitulada: **GEOGRAFIA ELEITORAL: TENDÊNCIAS E INDICAÇÕES PARA O DESENVOLVIMENTO DA CIÊNCIA POLÍTICA**, sob orientação do Prof. Me.RODRIGO ROSSI HOROCHOVSKI, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 13 de Agosto de 2021.

Assinatura Eletrônica

19/08/2021 15:19:53.0

RODRIGO ROSSI HOROCHOVSKI  
Presidente da Banca Examinadora

Assinatura Eletrônica

19/08/2021 17:25:26.0

ADRIANO NERVO CODATO

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

19/08/2021 16:06:09.0

JOSÉ ANTONIO CHEIBUB

Avaliador Externo (TEXAS A&M UNIVERSITY)

Assinatura Eletrônica

19/08/2021 17:50:34.0

CLAUDIA ROBBI SLUTER

Avaliador Externo (UNIVERSIDADE DO RIO GRANDE DO SUL)

Assinatura Eletrônica

19/08/2021 19:21:58.0

SONIA LUIZA TERRON

Avaliador Externo (INSTITUTO BRASILEIRO DE GEOGRAFIA E  
ESTADÍSTICA)

*Para todas Ayabás, para todas elas*

## **AGRADECIMENTOS**

Sou grato ao Rafael Sturaro Bernardelli, a Maria Cecília Eduardo, a Juliana Inez Luiz de Souza, ao Bruno Fernando da Silva e aos meus orientadores Rodrigo Rossi Horochovski e Ivan Jairo Junckes. Vocês me ajudaram de forma direta na elaboração deste trabalho, minha sincera gratidão.

D'après une opinion courante, les élections ne sont qu'un domaine d'incohérence et de fantaisie. En les observant à la fois de près et **de haut**, je suis arrivé à la conclusion contraire. Si, selon le mot de Goethe, l'enfer même a ses lois, pourquoi la politique n'aurait-elle pas les siennes?

---

André Siegfried (1913)

## RESUMO

O espaço geográfico é essencial para as análises de diversas linhas de pesquisa da Ciência Política, como Sistemas Eleitorais, Representação Política e Comportamento Eleitoral. Isso motiva o desenvolvimento de operacionalizações para a apreensão dos fenômenos a partir da análise espacial. Entretanto, existe uma tendência de naturalização do espaço no *mainstream* da Ciência Política e é comum que as operacionalizações realizadas apresentem limitações. O objetivo que guia esta tese é indicar caminhos de desenvolvimento metodológico e tendências de pesquisa. Com isso, busco fomentar o desenvolvimento da Geografia Eleitoral na Ciência Política brasileira. A pesquisa tem dois trabalhos empíricos principais: uma análise bibliométrica e cientométrica da produção da Geografia Eleitoral e o desenvolvimento, comparação e teste de um modelo original de análise espacial do voto. Os resultados indicam desalinhamentos entre a produção da Geografia Eleitoral em nível internacional e no Brasil. O maior deles é a ausência de conexão com a Geografia Humana e, mais especificamente, com as pesquisas com abordagens pós-estruturalistas e de identidade. Para lidar com as limitações metodológicas, testei as operacionalizações mais recorrentes na literatura e apresento como modelo final uma proposta original, baseada em distâncias medidas por redes transitáveis e com peso variável pela densidade local.

Palavras-chave: Metodologia Política. Geografia Eleitoral. Análise Espacial. Resultados eleitorais agregados. Dados espaciais.



## **ABSTRACT**

Geographic space is essential for several research lines of Political Science, such as Electoral Systems, Political Representation, and Electoral Behaviour. This motivates the development of operationalizations for the comprehension of political phenomena through spatial analysis. However, there is a tendency to naturalize the space in the mainstream of Political Science, and it is usual that the current operationalizations present limitations. The goal of this thesis is to suggest methodological development paths and research directions. Hence, I intend to stimulate the development of Electoral Geography in Brazilian Political Science. This work has two main empirical trials: a bibliometric and scientometrical analysis targeting the Electoral Geography literature; and the development, comparison, and testing of an original spatial model for turnout analysis. The results indicate mismatches between the international and the Brazilian literature. The most important difference found in the Brazilian research is the absence of a connection with Human Geography and, specifically, with post-structuralist and identity approaches. To tackle the methodological limitations, I have tested the most usual operationalizations found in the literature, although the final model is original, based on density-dependent weighted distances extracted from transit networks.

Keywords: Political Methodology. Electoral Geography. Spatial Analysis. Aggregated election results. Spatial data.

## LISTA DE FIGURAS

1.1	Estrutura do Estudo . . . . .	22
2.1	Representação de fenômenos do mundo real como objetos espaciais . . . . .	27
2.2	Tipos de problema da área modificável (MAUP) . . . . .	30
2.3	Diagrama das relações espaciais por categoria . . . . .	33
2.4	Relações espaciais topológicas de tipo Q1. . . . .	34
2.5	Estrutura da rede de operação política ( <i>brokerage</i> ) . . . . .	44
3.1	Evolução no número de artigos (1959-2020), por país (Brasil e outros) e áreas (Ciência Política e outras) . . . . .	51
3.2	Distribuição de países por número de artigos publicados e citações recebidas	52
3.3	Distribuição de Áreas de pesquisa dos artigos. . . . .	55
3.4	Número de autores por artigo, 1997-2020. . . . .	57
3.5	Mapa de árvore dos 20 autores mais citados no Geral, na Ciência Política e no Brasil . . . . .	59
3.6	Grafo das relações entre países e áreas de pesquisa . . . . .	62
3.7	Grafo unimodal das relações entre países a partir dos autores citados . . . . .	64
3.8	Grafo unimodal das relações entre áreas de pesquisa a partir dos autores citados . . . . .	66
3.9	Grafo das relações entre palavras-chaves . . . . .	68
3.10	Grafo das relações entre países e palavras-chaves. . . . .	70
3.11	Grafo das relações entre áreas de pesquisa e palavras-chaves . . . . .	73
4.1	Esquemática dos métodos. . . . .	80
4.2	Detalhe das Relações espaciais A2-Q1. . . . .	81
4.3	Detalhe das Relações espaciais P-DL . . . . .	81
4.4	Distribuição de eleitores por locais de votação, 2018, Paraná . . . . .	84

4.5	Mapa de densidade dos locais de votação, Paraná, 2018. . . . .	87
4.6	Histograma das menores distâncias entre vizinhos dos locais de votação do Paraná, 2018 . . . . .	88
4.7	Histograma do número de vizinhos por local de votação com <i>threshold</i> de 29,30 km, Paraná, 2018. . . . .	89
4.8	Mapa de relações entre Locais de Votação com <i>threshold</i> de 29,30 km, Paraná, 2018 . . . . .	90
4.9	Histograma do número de vizinhos por local de votação com contiguidade de ordem 1, Paraná, 2018. . . . .	91
4.10	Distorções nas conexões entre unidades espaciais de polígonos de Thiessen e matriz de vizinhança Q1, Paraná, LV, 2018 . . . . .	92
4.11	Exemplo de viés de distância em linha reta e de Decomposição Voronoi, Paraná, LV, 2018 . . . . .	93
4.12	Comparação de fatores de distância para relações espaciais com decaimento exponencial com fator local DL. . . . .	96
4.13	Gráfico de Violino com Boxplot de resultados de I de Moran nos métodos P-DL e A-Q1, amostra de candidaturas aos cargos de Deputado Estadual e Federal, 2018, Paraná . . . . .	100
4.14	Gráfico de Dispersão de resultados de I de Moran nos métodos P-DL e A-Q1, amostra de candidaturas aos cargos de Deputado Estadual e Federal, 2018, Paraná . . . . .	101
4.15	Teste de Bland-Altman para concordância entre métodos P-DL e A-Q1 para o índice I de Moran, amostra de candidaturas aos cargos de Deputado Estadual e Federal, 2018, Paraná. . . . .	103
4.16	Gráfico de Dispersão de médias e diferenças de I de Moran nos métodos P-DL e A-Q1, amostra de candidaturas aos cargos de Deputado Estadual e Federal, 2018, Paraná . . . . .	104
4.17	Gráfico de Dispersão de médias e diferenças de I de Moran nos métodos A2-Q1 e A-Q1, amostra de candidaturas aos cargos de Deputado Estadual e Federal, 2018, Paraná . . . . .	105

4.18	Gráfico de Montanha (Função de Distribuição Acumulada Empírica Dobrada) das diferenças de resultados dos índices I de Moran nos métodos A2-Q1 e P-DL em comparação com A-Q1, amostra de candidaturas aos cargos de Deputado Estadual e Federal, 2018, Paraná. . . . .	106
4.19	Histogramas Comparativos das diferenças de resultados dos índices I de Moran nos métodos A2-Q1 e P-DL em comparação com A-Q1, amostra de candidaturas aos cargos de Deputado Estadual e Federal, 2018, Paraná	107
4.20	Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato Gustavo Bonato Fruet, 2018, Paraná, Brasil. . .	108
4.21	Mapa Exemplo de Divergência do tipo Franjas Municipais. . . . .	109
4.22	Mapa Exemplo de Divergência do tipo Padrão Anulado . . . . .	110
4.23	Mapa Exemplo de Divergência do tipo Simplificação . . . . .	111
4.24	Indicador Local de Associação Espacial de candidatura indígena, Ivan Kaingang, Paraná, 2018 . . . . .	115
4.25	Indicador Local de Associação Espacial de candidatura indígena, Ivan Kaingang, Paraná, 2018 (1) . . . . .	116
4.26	Indicador Local de Associação Espacial de candidatura indígena, Ivan Kaingang, Paraná, 2018 (2) . . . . .	117
4.27	Indicador Local de Associação Espacial de candidatura indígena, Ivan Kaingang, Paraná, 2018 (3) . . . . .	118
4.28	Indicador Local de Associação Espacial de candidatura indígena, Ivan Kaingang, Paraná, 2018 (4) . . . . .	119
4.29	Indicador Local de Associação Espacial de candidatura indígena, Ivan Kaingang, Paraná, 2018 (5) . . . . .	120
4.30	Locais de Votação, 2018, Brasil, Mato Grosso, Aldeia Japuira . . . . .	122
A.1	Diagrama das etapas do método de mapeamento do estado da arte e identificação de lacunas e tendências de pesquisa. . . . .	142
A.2	Diagrama de Venn proporcional da distribuição e interseção dos artigos entre bases de dados (Coleção Principal WoS e SciELO), área (Ciência Política) e país (Brasil) . . . . .	148

B.1	Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato ARTIDOR JOSÉ DELCI SOARES, 2018, Paraná, Brasil. . . . .	153
B.2	Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato ANTONIO WANDSCHEER, 2018, Paraná, Brasil. . . . .	154
B.3	Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato LUIZ ALSIONI MARTINS DA SILVA, 2018, Paraná, Brasil. . . . .	155
B.4	Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato PLAUTO MIRO GUIMARÃES FILHO, 2018, Paraná, Brasil. . . . .	156
B.5	Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato LUIZ CLAUDIO ROMANELLI, 2018, Paraná, Brasil. . . . .	157
B.6	Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato JOSE CARLOS SAHAGOFF RAAD, 2018, Paraná, Brasil. . . . .	158
B.7	Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato FELIPE RAMON DOS PASSOS, 2018, Paraná, Brasil. . . . .	159
B.8	Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato JORGE GOMES DE OLIVEIRA BRAND, 2018, Paraná, Brasil. . . . .	160
B.9	Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato MABEL CORA CANTO, 2018, Paraná, Brasil. . . . .	161
B.10	Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato GUSTAVO BONATO FRUET, 2018, Paraná, Brasil. . . . .	162
C.1	Terras Indígenas com Legenda de Etnias, Paraná. . . . .	163

## LISTA DE TABELAS

2.1	Presença de justificativas para definição de Objetos Espaciais. . . . .	43
4.1	Estatísticas descritivas de candidaturas e de votação para os cargos de Deputado Federal e Estadual, 2018, Paraná. . . . .	83
4.2	Estatísticas descritivas da densidade demográfica do Paraná a partir dos municípios (hab/km <sup>2</sup> ). . . . .	86
A.1	Testes de construção de corpus analítico por palavras-chave e base de dados - Web of Science, 06/02/2020 . . . . .	145
A.2	Testes de construção de corpus analítico por palavras-chave e base de dados - Scopus, 06/02/2020 . . . . .	146
A.3	Testes de construção de corpus analítico por palavras-chave e base de dados - Web of Science, 25/05/2020 . . . . .	147

## LISTA DE ACRÔNIMOS E SIGLAS

CPBR	Ciência Política Brasileira
e.g.	Por exemplo, do latim <i>exempli gratia</i>
GE	Geografia Eleitoral
i.e.	Isto é, do latim <i>id est</i>
WoS	<i>Web of Science</i>
VSPC	<i>Visual spatial pattern comparison</i>
TSE	Tribunal Superior Eleitoral
PROS	Partido Republicano da Ordem Social
PSL	Partido Social Liberal
PSD	Partido Social Democrático
OSM	OpenStreetMap
TI	Terra Indígena
UIF	Limite interno inferior
A-Q1	Modelo de análise espacial com objeto espacial de tipo área e relação espacial de contiguidade de primeira ordem
A2-Q1	Modelo de análise espacial com objeto espacial de tipo área criada a partir de decomposição de pontos e relação espacial de contiguidade de primeira ordem
P-DL	Modelo de análise espacial pontual com relação espacial de distância com fator de peso local.

## LISTA DE SÍMBOLOS

$\sigma$	Desvio-padrão
$\mathcal{B}$	Relação Espacial
$(i, j)$	Par de objetos espaciais
$\mathcal{W}$	Matriz de conexão espacial
$OE$	Objeto espacial



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>18</b>
1.1	CONTEXTO E PROBLEMA	18
1.2	CONTRIBUIÇÕES DA TESE	19
1.3	FONTES E MÉTODOS	20
1.4	ESTRUTURA DO TRABALHO	21
<b>2</b>	<b>GEOGRAFIA ELEITORAL: APLICAÇÕES, MÉTODOS E LIMITAÇÕES DA PRODUÇÃO BRASILEIRA</b>	<b>23</b>
2.1	<i>RATIONALE</i>	23
2.2	MÉTODOS ESPACIAIS E VOTO	25
2.2.1	Análise Espacial	26
2.3	OPERACIONALIZAÇÕES DA GE BRASILEIRA	35
2.3.1	Objetos Espaciais e Escalas nos estudos de voto no Brasil	35
2.3.2	<i>W</i> , Brasil	40
2.3.3	Discussão	42
2.4	CONSIDERAÇÕES FINAIS	46
<b>3</b>	<b>GEOGRAFIA ELEITORAL: O CAMPO, A PRODUÇÃO BRASILEIRA E SUAS LACUNAS</b>	<b>47</b>
3.1	<i>RATIONALE</i> E ESTRATÉGIA METODOLÓGICA	47
3.2	RESULTADOS	49
3.2.1	Análise Bibliométrica	49
3.2.2	Análise Cientométrica	60
3.3	LACUNAS	74
3.3.1	Uma Geografia Eleitoral mais <i>Geográfica</i>	74
3.3.2	Novos Fenômenos Espaciais	75
3.4	CONSIDERAÇÕES FINAIS	76

<b>4</b>	<b>96 MIL OBJETOS ESPACIAIS RELACIONADOS POR ROTAS TRANSITÁVEIS</b>	<b>78</b>
4.1	<i>RATIONALE</i>	78
4.1.1	Uma Operacionalização Original	78
4.1.2	Um Caso Original	82
4.2	PROCEDIMENTOS: O DESENVOLVIMENTO DO MODELO	82
4.2.1	Bases de Dados	82
4.2.2	<i>W</i> — Rotas Transitáveis e Peso Local	86
4.2.3	Preparação e Processamento Estatístico de Dados	97
4.3	CONCORDÂNCIA ENTRE MÉTODOS	98
4.3.1	Comparação de Índices	99
4.3.2	Comparação Visual de Padrões Espaciais	107
4.3.3	Discussão	111
4.4	ANÁLISE: UM ESTUDO ORIGINAL DA REPRESENTAÇÃO INDÍGENA	112
4.4.1	Estratégia Empírica	113
4.4.2	Resultados	114
4.4.3	Discussão	121
4.5	CONSIDERAÇÕES FINAIS	123
<b>5</b>	<b>CONCLUSÃO</b>	<b>125</b>
5.1	SUMÁRIO DE RESULTADOS	125
5.2	POSSÍVEIS IMPLICAÇÕES	126
5.3	LIMITAÇÕES DO ESTUDO	129
5.4	RECOMENDAÇÕES PARA PESQUISAS FUTURAS	131
	<b>REFERÊNCIAS</b>	<b>133</b>
	<b>APÊNDICE A – MÉTODOS E PROCEDIMENTOS DO MAPEAMENTO DA GEOGRAFIA ELEITORAL</b>	<b>141</b>
A.0.1	Definição de base de dados	143
A.0.2	Processamento estatístico de dados	150

	<b>APÊNDICE B – MAPAS DO INDICADOR LOCAL DE ASSOCIAÇÃO ESPACIAL . . . . .</b>	<b>152</b>
B.1	CANDIDATOS COM MAIORES DIFERENÇAS DE <i>SCORE</i> ENTRE MÉTODOS: <i>A – Q1 - P – DL</i> . . . . .	152
B.2	CANDIDATOS COM MAIORES DIFERENÇAS DE <i>SCORE</i> ENTRE MÉTODOS: <i>P – DL - A – Q1</i> . . . . .	158
	<b>APÊNDICE C – ETNIAS POR TERRAS INDÍGENAS DO PARANÁ</b>	<b>163</b>

## 1 INTRODUÇÃO

Esta é uma tese metodológica que tem como objetivo fomentar o desenvolvimento da Geografia Eleitoral na Ciência Política brasileira. Nesse sentido, coloco uma lupa sobre as limitações da produção da área. Convido os cientistas políticos a uma reflexão mais aprofundada sobre as necessidades teóricas por trás dos modelos de análise espacial de voto e as lacunas da literatura nacional.

### 1.1 CONTEXTO E PROBLEMA

O estudo acadêmico do voto apresenta um grande conjunto de evidências que indica efeitos significativos da composição e do contexto geográfico e de regras eleitorais de agregações espaciais. Por isso, em áreas de pesquisa como Representação Política, Comportamento Eleitoral e Sistemas Eleitorais, o *espaço geográfico* é uma categoria fundamental de análise. Ver o conjunto de argumentos teóricos e de evidências empíricas em, e.g., Agnew, 1996; Ames, 2003; Borges et al., 2016; de Carvalho, 2003; Faganello e Fernandes, 2018; Gorecki e Marsh, 2012; Marques, 2018; Marzagão, 2013; Meneguello e Neto, 2012; Pattie e Johnston, 2000, 1999; Silotto, 2019; Soares e Terron, 2008; Sonnleitner, 2017; Terron e Soares, 2010; Zolnerkevic, 2019.

Na contramão dessas evidências, existe uma tendência de naturalização do espaço na Ciência Política. Na disciplina, o espaço é, essencialmente, ignorado, interpretado como um meio, uniforme e independente, sobre o qual ocorrem as ações políticas (Jones e Clark, 2020). A crítica de sub-exploração da localização geográfica correspondente aos dados da disciplina é lugar comum (King, 1996; Elazar, 1999; Terron, 2012), sobretudo na América Latina (Rodrigues-Silveira et al., 2017; Sonnleitner, 2013; Marzagão, 2013; Gonçalves, 2016).

Esse é um problema pertinente porque os dados político-eleitorais dificilmente são manipuláveis. Diferentemente de experimentos clínicos, com toda a sorte de controles de efeito das variáveis explicativas, os cientistas políticos raramente têm controle sobre seus dados. As nossas pesquisas são, em sua grande maioria, de caráter observacional e, por isso, os dados são estruturados espacialmente, mesmo que de forma implícita (Salvatore e Ruggeri, 2021). Mas, a estatística aplicada e ensinada na Ciência Política, na maior parte dos casos, não lida com problemas advindos de dependência e heterogeneidade

espacial (Salvatore e Ruggeri, 2021). Os dados são interpretados seguindo premissas de aleatoriedade e independência como se fossem experimentos da psicologia (Goodchild e Janelle, 2010). Os modelos de análise mais comuns da área tem como base regressão linear tradicional e métodos econométricos onde a premissa de independência entre os objetos é um pressuposto estatístico. Ou seja, de uma perspectiva metodológica, negligenciar a localização geográfica dos dados pode invalidar os resultados da área<sup>1</sup>.

Outro ponto relevante é que ao marginalizar a Geografia Eleitoral a Ciência Política deixa de utilizar valiosas ferramentas analíticas. A geografia é uma das maneiras mais claras de entender a política. Por um lado, o da cartografia eleitoral, os dados organizados em mapa apresentam padrões instantaneamente reconhecíveis e interpretáveis. Conhecemos muito da geografia e, por isso, ao conectar uma variável que desejamos explicar ao espaço, estamos conectando o que não sabemos a informações que já conhecemos (King, 1996, p. 161). Por outro lado, técnicas de análise da estatística espacial possibilitam estudar os fenômenos político-eleitorais em diferentes dimensões espaciais que revelam as relações entre práticas sociais e o espaço (Zavala, 2012). Esse intercambio entre áreas tem capacidade de aumentar a profundidade das análises da Ciência Política (Sonnleitner, 2013).

A motivação desta tese surge desse contexto: sub-exploração de uma área com alto potencial para a disciplina. O objetivo que guia este trabalho é indicar caminhos de desenvolvimento metodológico e tendências de pesquisa. Com isso, busco fomentar o desenvolvimento da Geografia Eleitoral na Ciência Política brasileira.

## 1.2 CONTRIBUIÇÕES DA TESE

Minhas principais contribuições são:

1. Entrego um **guia de desenvolvimento** para a Geografia Eleitoral brasileira, como um guia de correção de rota para a internacionalização da área:
  - Analisei atributos bibliométricos e características relacionais da produção da área. Os resultados indicam desalinhamentos entre a produção da Geografia Eleitoral em nível internacional e no Brasil. Os dados são utilizados para mostrar os espaços de oportunidade de desenvolvimento de pesquisas futuras.
2. Indico **caminhos metodológicos** para aumentar a qualidade dos estudos de voto agregado no Brasil:

---

<sup>1</sup>Ver mais sobre o problema da dependência espacial em Anselin e Bera (1998).

- Realizei uma síntese das operacionalizações de análise espacial na Geografia Eleitoral brasileira. A reflexão indica que podemos aprimorar a definição de matriz de vizinhança com maior atenção aos pressupostos teóricos e a superfície modelada. Além disso, explorações intramunicipais podem revelar padrões desconhecidos, principalmente a partir de locais de votação, sem limites de área pré-estabelecidos.
3. Desenvolvo uma **operacionalização original** da análise espacial de voto no Brasil:
- Desenvolvi um método avançado de representação espacial dos resultados eleitorais brasileiros. O modelo é original e avança, principalmente, na definição de matriz de relações espaciais.

### 1.3 FONTES E MÉTODOS

A pesquisa tem dois trabalhos empíricos principais: uma análise bibliométrica e cientométrica da produção da Geografia Eleitoral e o desenvolvimento, comparação e demonstração do modelo espacial de análise proposto.

Para mapear o estado da arte e identificar lacunas e tendências de pesquisa da Geografia Eleitoral, combino métodos bibliométricos e cientométricos. Os métodos principais são *Co-citation*, análise da estrutura do campo a partir do compartilhamento de referências entre artigos (Small, 1973), e *Co-word*, análise da estrutura relacional dos termos identificadores da produção científica (Callon et al., 1983). A amostra soma 226 artigos, tendo como fonte a plataforma *Web of Science* (WoS) e o termo de busca “*electoral geography*”. O pré-processamento, para tratar as inconsistências e garantir a confiabilidade das observações, foi realizado com auxílio do *software* OpenRefine (v. 3.3). O processamento estatístico de redes bibliométricas, foi realizado com auxílio dos *softwares* Gephi (v. 0.9.2) e Sci2 (v. 1.3).

Para o desenvolvimento do modelo que proponho, o P-DL, utilizo três bases de dados: resultados eleitorais agregados em locais de votação, do Repositório de Dados Eleitorais do Tribunal Superior Eleitoral (TSE); coordenadas geográficas dessas unidades, também com o TSE como fonte principal e Hidalgo (2021) e Pindograma (2020) como fontes auxiliares; e, por fim, os pesos das relações espaciais entre os objetos espaciais em rede transitável, construído a partir do conjunto de rede de ruas do Brasil do *OpenStreetMap* (OSM). Esses dados foram pré-processados nas linguagens R e Python e no *software* Qgis (v. 3.16). Para a análise de concordância entre métodos utilizei o

teste de Bland-Altman com índices de autocorrelação global (I de Moran) e *visual spatial pattern comparison* (VSPC) de mapas de autocorrelação local. Na análise de caso das bases eleitorais indígenas utilizei o mesmo método de VSPC e autocorrelação local. As estatísticas espaciais foram processadas com auxílio dos *softwares* ArcGIS Pro (v. 2.8) e GeoDa (v. 1.18). As estatísticas clássicas foram processadas nos *softwares* NCSS (v. 12.0) e MedCalc (v. 19.0).

#### 1.4 ESTRUTURA DO TRABALHO

A tese está estruturada em quatro partes, além desta introdução:

- Primeiro, no Capítulo 2, apresento os componentes básicos da análise espacial e exploro como a produção brasileira operacionaliza as suas pesquisas;
- No Capítulo 3, desenvolvo um conjunto de indicações de lacunas e oportunidades para o desenvolvimento da agenda da pesquisa da Geografia Eleitoral na Ciência Política Brasileira. Essas recomendações são sustentadas por resultados bibliométricos e cientométricos da produção da área;
- No Capítulo 4, mostro uma operacionalização original de análise espacial de voto no Brasil. Esse modelo é comparado com a operacionalização mais comum da área. Também ilustro o potencial da proposta com uma análise exploratória das bases eleitorais de uma candidatura indígena;
- Por fim, no Capítulo 5, faço um sumário dos resultados, contribuições e possíveis implicações da pesquisa. Além disso, reconheço as principais limitações do estudo e dou recomendações para pesquisas futuras;

A Figura 1.1 ilustra como se dá a relação entre os capítulos. Os resultados de oportunidades metodológicas (Capítulo 2) e temáticas (Capítulo 3) dão base ao modelo e a análise desenvolvidas (Capítulo 4), que culminam nas conclusões da pesquisa (Capítulo 5).

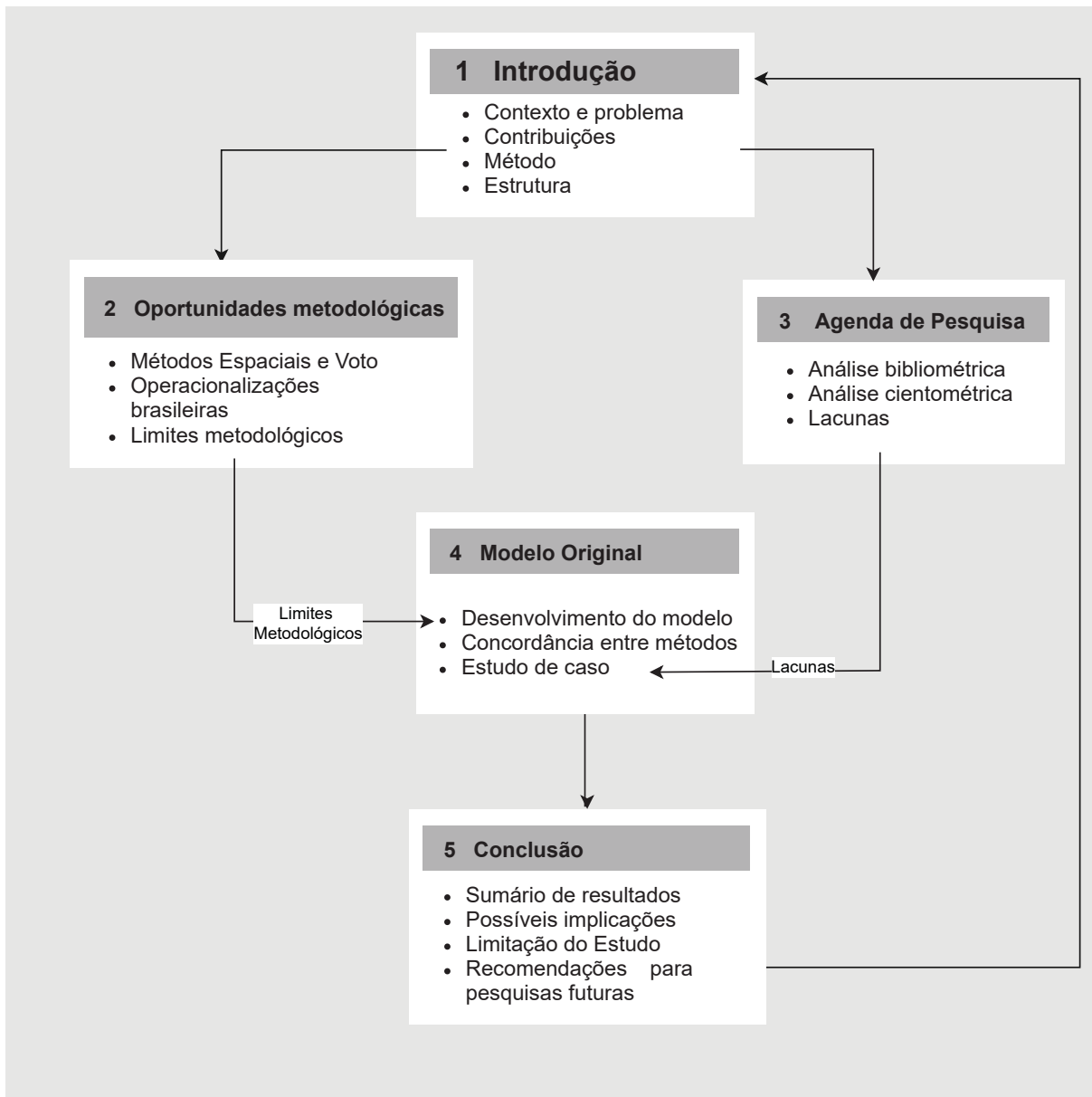


Figura 1.1: Estrutura do Estudo

Fonte: O autor (2021)



## 2 GEOGRAFIA ELEITORAL: APLICAÇÕES, MÉTODOS E LIMITAÇÕES DA PRODUÇÃO BRASILEIRA

Na Ciência Política Brasileira (CPBR, a partir daqui), a análise de dados eleitorais espacialmente agregados é usada para tratar de temas diretamente relacionados a problemas da democracia do país. Por um lado, destacam-se as consequências espaciais do Sistema Eleitoral brasileiro, como a distritalização informal da competição e representação política, incentivos a comportamentos parlamentares concentracionistas e disputas intrapartidárias dada a combinação de lista aberta, distritos de grande magnitude e sistema proporcional<sup>1</sup>. Por outro lado, verifica-se a importância das estruturas geográficas sobre o comportamento eleitoral, como a relação entre direcionamento de políticas públicas e o deslocamento de bases geoeleitorais<sup>2</sup>.

Sem deixar de reconhecer a contribuição dada pela CPBR a partir dessas análises, é preciso observar que, muitas vezes, as operacionalizações feitas por esses trabalhos apresentam limitações. Neste capítulo, trato das pesquisas da área em busca dessas limitações. O objetivo é indicar soluções para o desenvolvimento de uma análise espacial de voto robusta para os resultados eleitorais brasileiros. É possível resumir o problema ao qual este capítulo se dedica na seguinte pergunta:

*Pergunta de pesquisa 1:* Como podemos aprimorar a operacionalização dos componentes principais da análise espacial dos votos no Brasil?

O texto está estruturado em três partes: a primeira parte apresenta as motivações para a realização da pesquisa; a segunda traz os componentes básicos da análise espacial; e, por fim, apresento as operacionalizações da literatura da Geografia Eleitoral brasileira e indico caminhos de desenvolvimento.

### 2.1 RATIONALE

Embora o estudo territorial do voto tenha desempenhado um papel de destaque na fundação da Ciência Política moderna (Rodrigues-Silveira et al., 2017), críticas metodológicas levaram a Geografia Eleitoral para a margem da disciplina. O principal

---

<sup>1</sup>Ver mais em Gonçalves (2016, p. 52-59).

<sup>2</sup>Ver mais em Terron (2009) e Simoni (2017).

exemplo disso é o viés da área modificável, usado como argumento para rejeitar as análises espaciais nas Ciências Sociais desde Robinson (1950). Por isso, aprimorar os nossos modelos de análise é relevante não só para obter conclusões mais confiáveis, mas para que as teorias da Geografia Eleitoral se desenvolvam na disciplina. Enquanto a análise espacial não for aplicada de maneira mais precisa nós estaremos defendendo conclusões com pouca força em uma disciplina quantitativamente orientada (O'Loughlin, 2008, p. 34).

Neste capítulo exploro as definições metodológicas dos trabalhos da Geografia Eleitoral (GE, a partir daqui) no Brasil. Reúno aspectos metodológicos e as justificativas utilizadas pela produção da GE no Brasil<sup>3</sup>. O objetivo não é fazer uma revisão exaustiva da teoria e método de cada trabalho<sup>4</sup>. A ideia é identificar as operacionalizações das pesquisas que estudam voto no Brasil a partir da perspectiva geográfica. Mais especificamente, trato dos componentes básicos da análise espacial: as unidades espaciais de análise e as relações espaciais. Discutir a operacionalização da pesquisa espacial é essencial para evitar erros de especificação dos modelos, diminuir as críticas direcionadas às pesquisas espaciais e fazer melhores representações da realidade empírica. A motivação é explorar limitações e oportunidades metodológicas a partir dos dados eleitorais disponíveis.

O Brasil soma um território de 8.510 milhões de km<sup>2</sup>, distribuídos entre 5.570 municípios, mais de 15 mil bairros e 800 mil logradouros. Uma divisão territorial que representa um quadro complexo, de grande diversidade socioeconômica e cultural (Soares e Terron, 2008). É um grande território de notáveis desigualdades econômicas, políticas e sócio-demográficas (Terron e Soares, 2010). As decisões metodológicas para representar essa realidade não são simples, mas, ao mesmo tempo, um universo com tamanha diversidade territorial pode revelar muitos fenômenos políticos.

Como a Geografia é uma disciplina baseada em conceitos espaciais — i.e., área, caminho, centro, contexto, direção, distância, escala, fronteiras, limite, linhas, movimento, nós, periferia, polígonos, pontos, rede, região — existem questões teórico-metodológicas específicas que devem ser levadas em conta por esse tipo de análise. O problema é que normalmente essas questões são desconhecidas pelos cientistas políticos (Ethington e

---

<sup>3</sup>Representada pelos artigos da seguinte busca sistemática: foram coletados todos os artigos publicados em revistas indexadas pela plataforma *Web of Science* incluindo *SciELO Citation Index* e *Core Collection* com a identificação "*Electoral Geography*" (ver Apêndice A para mais maiores detalhes metodológicos). Os resultados obtidos em (Gonçalves, 2016) também são utilizados como fonte. Em pesquisa anterior (Gonçalves, 2016) encontrei 35 pesquisas que tratavam de dimensões espaciais das eleições brasileiras de alguma forma. Essas pesquisas incluíam artigos em periódicos não indexados, livros e *paper* de congresso ver Gonçalves (2016, p.47)

<sup>4</sup>Para uma revisão mais detalhada das explorações da área ver Gonçalves (2016) páginas 51-68

McDaniel, 2007). A seção a seguir se dedica a apresentar alguns conceitos básicos dos métodos espaciais e, mais especificamente, do processo de formalização empírica da análise espacial (as unidades espaciais de análise, ou objetos espaciais, e a matriz de conexão espacial).

## 2.2 MÉTODOS ESPACIAIS E VOTO

A Ciência Política é uma disciplina fundamentalmente interdisciplinar. É impossível ignorar a natureza multiobjetiva dos processos político-sociais. A interface entre áreas é parte da condição de explicar os fenômenos políticos, por isso, trocamos ideias, teorias, dados e métodos com disciplinas que vão da "*anthropology to zoology — and everything in between*" (Druckman et al., 2009, p.486). Uma das áreas com as quais compartilhamos métodos e conceitos é a Geografia — e disciplinas afins como a Cartografia, a Ciência da Informação Geográfica e as Ciências Geodésicas.

Os métodos importados dessas disciplinas vão de análises visuais simples a modelos computacionais complexos<sup>5</sup>. O modo como a Ciência Política dialoga com os métodos dessas disciplinas pode ser dividido em duas categorias: a cartografia eleitoral e a análise espacial de fenômenos político-eleitorais.

A visualização de dados espaciais auxilia a identificar padrões, tendências e ilustrar teorias. A representação dos dados eleitorais em mapa permite descrever a distribuição espacial das variáveis, identificar ocorrências *outliers* (atípicas) e padrões de relação dos dados. Quando analisados ao longo do tempo, e.g., ajudam na compreensão dos componentes explicativos mais duradouros do comportamento eleitoral (Zolnerkevic, 2019). Os mapas são ótimos instrumentos para gerar hipóteses (King, 1996). Esses são alguns dos motivos pelos quais a abordagem cartográfica, ou o mapeamento das eleições e observação da distribuição dos votos, segue sendo uma ferramenta relevante de investigação. Mas a versão moderna da área utiliza técnicas de Análise Espacial (Terron, 2012).

---

<sup>5</sup>É preciso lembrar que os conceitos e métodos importados pela Ciência Política possuem caráter auxiliar para as análises. Diferentemente de disciplinas como a Engenharia Cartográfica, a Ciência da Informação Geográfica e Ciências Geodésicas, utilizamos a cartografia e análise espacial como ferramentas de análise, como a estatística. Alerto quem lê que a exposição a seguir é superficial, os conceitos foram filtrados apenas para o essencial e dialogam com pessoas pesquisadoras da Ciência Política. Para um guia mais aprofundado ver Smith et al. (2015). Sobre esse ponto agradeço os comentários da avaliadora externa professora doutora Claudia Robbi Sluter.

### 2.2.1 Análise Espacial

A análise espacial é associada aos métodos estatísticos tradicionais e computacionais, mas as características particularidades dos dados e a dependência natural entre as observações impossibilitam a extensão dos requisitos da estatística clássica (Smith et al., 2015). A principal característica da disciplina é a inclusão explícita do espaço geográfico nos modelos de análise (Druck et al., 2004). De outra forma, o que torna uma pesquisa *espacial* é a interpretação de fenômenos a partir da incorporação do seu espaço geográfico de maneira direta no modelo analítico.

A análise espacial é uma forma de interpretar os fenômenos olhando para onde as coisas estão e como elas se conectam. Para a operacionalização de conceitos na análise espacial fazemos abstrações empíricas para representar os lugares (também chamados de objetos espaciais) e as conexões entre eles (relações espaciais). Esses conceitos passam por definições operacionais para se tornarem concretamente mensuráveis pela análise espacial. O que é essencial nessa definição é entender que "*talking about space implies talking about (spatial) objects and relationships*" (Matsakis et al., 2010, p. 49). O objeto é um lugar que pode ser posicionado no espaço euclidiano, normalmente representado por um ponto, uma área ou uma linha. As relações espaciais são as diversas formas como os objetos estão conectados, e.g., "o objeto A está a 5 km do objeto B", "o objeto C é adjacente ao objeto D". Essa definição pode parecer longe de uma pesquisa social, mas é o quadro branco do desenho das pesquisas espaciais.

Na Ciência Política esse quadro é preenchido de diversas formas. Os objetos espaciais são representações de, e.g., indivíduos (eleitores), ações de atores políticos (campanhas eleitorais, alianças locais), unidades administrativas (distritos eleitorais, municípios, países), agrupamentos de indivíduos (étnicos, culturais, socioeconômicos, de classe). Já as relações espaciais representam três mecanismos causais principais (Neumayer e Plümper, 2016): i) aprendizado ou imitação (e.g. eleitores próximos interagem e se comunicam mais frequentemente e podem aprender mais uns com os outros); ii) exterioridades (e.g. cidades próximas fazem câmbios de mercadorias, profissionais, serviços e por isso se influenciam mutuamente por, e.g., competição); e iii) coerção (e.g. milicianos exercem força sobre eleitores do território que dominam). As seções a seguir se dedicam a apresentar essas duas principais definições no processo de formalização empírica.

### 2.2.1.1 Objetos Espaciais e Escalas Geográficas

Os objetos espaciais (*OE*) são representações abstratas da realidade, tentativas de retratar atributos do mundo empírico através de pontos, linhas, polígonos e estruturas contínuas em mapas (Long e Robertson, 2018). Cada representação espacial determina sua visualização, modelagem e análise (Yao e Murray, 2013; Druck et al., 2004).

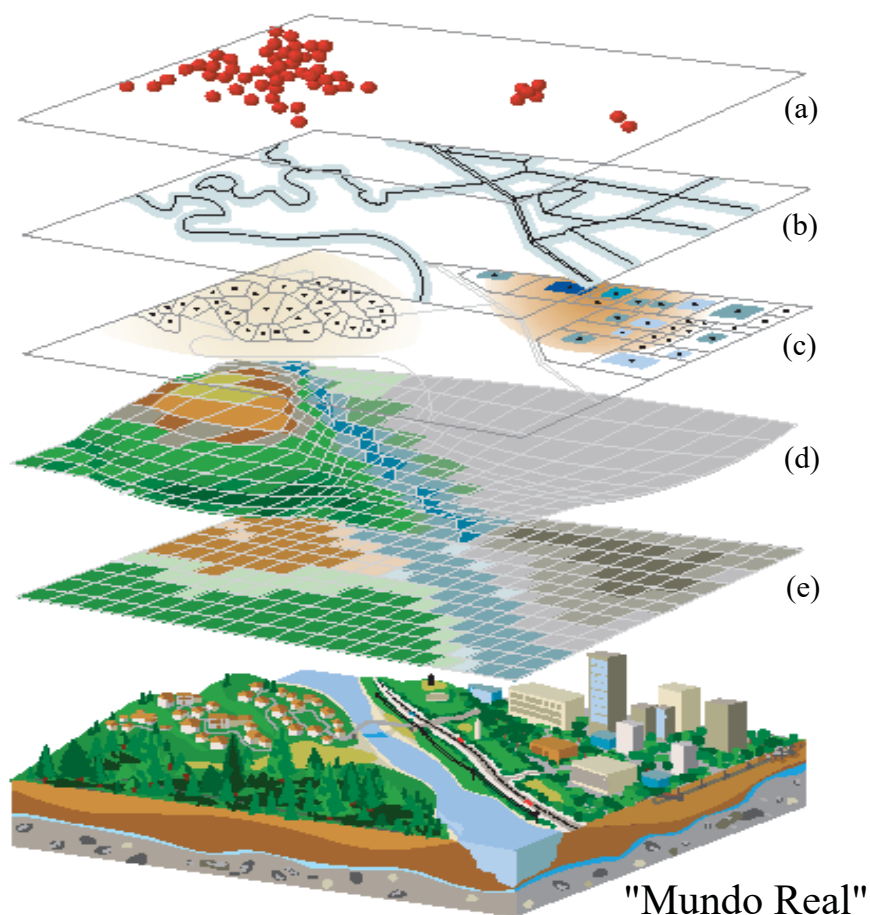


Figura 2.1: Representação de fenômenos do mundo real como objetos espaciais  
 Fonte: Adaptado pelo autor (2021) de Sánchez-Lozano et al. (2013)

A Figura 2.1 demonstra como o “mundo real” é representado a partir de primitivas geométricas da estrutura de dados: pontos (a) , linhas (b) , áreas ou polígonos fechados (c) e estrutura matricial (d) e (e).

- (a) Ponto — É a representação definida por um par de coordenadas. Representa processos pontuais ou a ocorrência de eventos no espaço (e.g., local de um crime, local de um protesto), uma pessoa (e.g., participante de um *survey*, usuário de um aplicativo), um imóvel (e.g., local de votação, uma casa).

- (b) Linha — É um objeto que possui mais de um ponto associado a sua geometria, mas sem formar uma área fechada. A linha pode ser reta ou curva, com direção ou sem, e.g, uma rua, um rio, um caminho percorrido para determinado lugar ou um trilho de trem.
- (c) Polígono — A área, ou polígono fechado, é uma série de pontos onde o primeiro é o mesmo que o último. É o tipo de dado espacial básico mais utilizado na Ciência Política como, e.g, uma cidade, um país, uma zona eleitoral, um setor censitário.
- (d) e (e) Estrutura matricial — Diferentemente dos três primeiros tipos de representação, que são de formato vetor, o formato matricial é uma representação dada por um conjunto de "pixels" localizados em coordenadas contíguas. Ou seja, os dados são divididos em células de tamanho igual, como um campo escalar. A estrutura matricial representa fenômenos mais complexos, distribuídos de forma contínua no espaço, é mais presente nas ciências naturais (e.g., chuvas, qualidade do ar, temperatura, elevação), mas também representa fenômenos sociais (e.g. densidade populacional, emissão de luz como *proxy* de atividade da economia local) (Salvatore e Ruggeri, 2021).

Cada *OE* contém informações sobre sua localização em relação à superfície terrestre, usualmente na forma de latitudes e longitudes. Para a exploração de fenômenos são associadas outras informações ao *OE*. No caso da GE a associação básica é feita com resultados eleitorais. Os votos podem ser associados a *OE* de diferentes formatos (e.g., ponto, área), tamanhos (e.g., bairros, cidades), níveis (e.g., local, regional, nacional). Por isso, a definição do *OE* que recebe os dados de resultados eleitorais é um problema corrente para a análise dos fenômenos de interesse. Esse problema pode ser visto de duas perspectivas: a metodológica e a teórica.

Usualmente, cientistas políticos trabalham com *OE* do tipo áreas com contagens e taxas agregadas. Em alguns casos os objetos são agregados em áreas através de junções espaciais (e.g., locais de votação, que são originalmente feições pontuais, são agregados por cidade, que são do tipo área), com a justificativa de, e.g., melhorar a visualização do fenômeno estudado (Faganello, 2017). Esse tipo de transformação gera um viés estatístico conhecido como *modifiable areal unit problem* (MAUP). Ao menos desde Robinson (1950), sabemos que a agregação escolhida afeta o resultado obtido e, por consequência, as conclusões da análise.

Do ponto de vista metodológico, é preciso definir qual objeto melhor se adequa à análise proposta visando uma estruturação espacial que não gere inferências enviesadas (Gonçalves, 2016). A área definida para a análise espacial tem impacto direto nos coeficientes – ver, e.g., evidências de dados eleitorais na França (Russo e Beauguitte, 2014) e no Brasil (Gonçalves, 2016). O MAUP é efeito da heterogeneidade espacial e opera de duas maneiras (de Ávila e Monasterio, 2009; Tobler, 1989): i) efeito do tamanho, quando unidades maiores agregam unidades menores, uma mudança de *agregação* da área (e.g. bairros agregados em município); ii) efeito do zoneamento, unidades dispostas em áreas de *arranjos* diferentes, uma mudança de localização — ilustrados na Figura 2.2

% votos no candidato A

10%	0%	5%
20%	30%	10%
5%	20%	0%

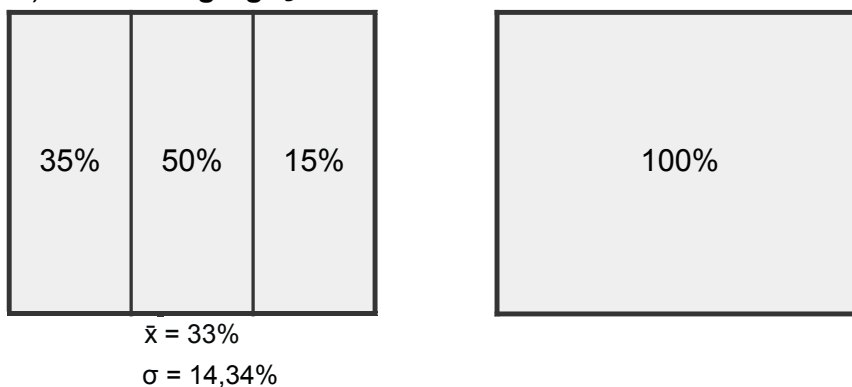
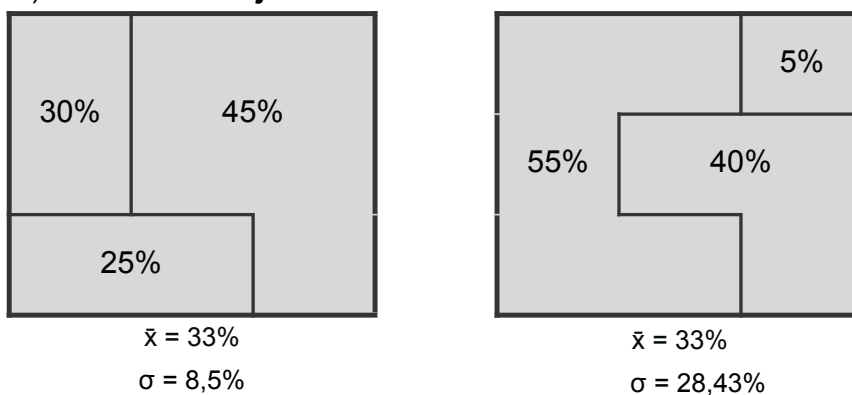
a) Efeito de **agregação** sobre o % de votos no candidato Ab) Efeito de **arranjo** sobre o % de votos no candidato A

Figura 2.2: Tipos de problema da área modificável (MAUP)

Fonte: O autor (2021)

Essa é uma discussão relevante e de impacto para o desenvolvimento da GE na Ciência Política. As proposições de Robinson (1950) impactaram a produção da disciplina nas décadas seguintes à sua publicação. Estudos com resultados eleitorais agregados espacialmente foram relegados, tratados automaticamente como “falácia



ecológica” (Simoni, 2017). Causa preocupação aos analistas identificar variações de grande magnitude e até inversão de sinais nos resultados estatísticos dos dados agregados. Mas o MAUP é inerente às análises de dados espacialmente agregados, por isso, essa é uma questão mais teórica do que empírica (King, 1996).

A escala geográfica é um dos conceitos fundamentais da Geografia. Essa não é a escala matemática simples, como a razão da dimensão real e a representada, como empregada na Cartografia, mas sim um conjunto de definições teórico-conceituais que permitem observar os fenômenos que lhes dão sentido (Terron, 2009), uma "escala de análise". Como é comum com conceitos-chave, não há consenso sobre o que o termo significa ou como ele deve ser operacionalizado.

Existem duas conceituações principais de escala na Geografia Humana<sup>6</sup> (Marston et al., 2005), nas quais se encontram os estudos de dados eleitorais. A primeira é a escala geográfica entendida pela metáfora de pirâmide, em uma escala de complexidade ou de nível hierárquico. A segunda entende que o termo deve representar a horizontalidade das relações. A escala hierárquica é a conceituação *padrão* para a análise de atores e ações políticas (Delaney e Leitner, 1997).

Na perspectiva de diferentes níveis “verticais” de análise o fenômeno responde a uma estrutura conectada. O material genético e o espaço sideral podem ser considerados como os limites inferiores e superiores dessa escala de conexão do comportamento humano (Flint, 2006, p. 11). Mas uma hierarquia mais recorrente é a que vai do indivíduo ao global, passando por local, regional e nacional. Para ilustrar o ponto com um tema político-eleitoral, podemos pensar em protesto:

Protest can be enacted at the individual scale, by breaking laws seen by the individual as unjust or by wearing clothes or tattoos that make a political statement. [...] But protest can also involve vigils outside of, say, abortion clinics or protests at animal hunts or laboratories conducting tests on animals. These “localized” acts require individual commitment and are also often motivated by national campaigns aimed at influencing the national legislative process. Increasingly, protest politics do not stop at the national scale; abortion politics, for example, are a component of discussions over the form of US foreign aid as well as a component of the missionary activity of many churches. (Flint, 2006, p. 11-12).

---

<sup>6</sup>Uma terceira conceituação busca combinar as duas escalas e também há abordagem ontológica plana, ver mais em Marston et al. (2005).

Para Rodrigues-Silveira (2013, p. 168) “a política é sempre um fenômeno multi-escalar”. Além da escala ser dependente ela é dinâmica, em constante transformação por ações políticas, econômicas e sociais de diferentes atores (Delaney e Leitner, 1997). Mas, é comum que na análise de instituições e ações políticas a escala de análise seja única e fixa. Uma noção “*of geographic scale as an unproblematic, pre-given and fixed hierarchy of bounded spaces*” (Delaney e Leitner, 1997, p. 93).

Considerar a característica hierárquica da escala para a definição de *OE* possibilita novas interpretações dos fenômenos. O detalhamento excessivo pode esconder uma percepção geral, ao mesmo tempo que uma grande abrangência pode ocultar interações fundamentais para entender o fenômeno eleitoral (Terron, 2009). A articulação de escalas tem grande potencial analítico, podendo evidenciar novos elementos sobre fenômenos políticos, “permite relevar dimensões tradicionalmente não consideradas e, em particular, conectar distintas escalas de análise na construção do conhecimento sobre um determinado processo político ou social” (Rodrigues-Silveira, 2013, 168).

### 2.2.1.2 *Relações Espaciais*

As relações espaciais entre os objetos são operacionalizadas a partir de uma matriz de conexão  $W$ <sup>7</sup>. Por definição,  $W$  determina quais observações são espacialmente dependentes umas das outras e o grau dessas dependências (Neumayer e Plümer, 2016, p. 175).  $W$  é uma matriz  $n \times n$  que representa a conexão par a par de cada *OE*.

A escolha de  $W$  tem papel fundamental no modelo de análise espacial. Medir o efeito do espaço no fenômeno de interesse com uma matriz que subestima ou superestima as conexões entre objetos gera estimativas enviesadas (Darmofal, 2006). Existem diversos métodos para definir (1) se existe relação entre os objetos e (2) qual o peso da relação. Tratarei de três categorias de métodos de definição mais comumente encontrados na Ciência Política: duas com base na posição espacial dos objetos a (1a) topológica e a de (1b) distância, e as baseadas no mecanismo causal de forma direta, as de tipo (1c) não-espaciais. E duas formas de peso das relações: (2a) medidas binarizadas, *1 root, simple minded* e (2b) contínuas, probabilísticas, de difusão.

---

<sup>7</sup>Nos trabalhos brasileiros  $W$  normalmente é encontrada como “matriz de vizinhança”. Ver Neumayer e Plümer (2016) sobre a escolha do termo “conexão”.

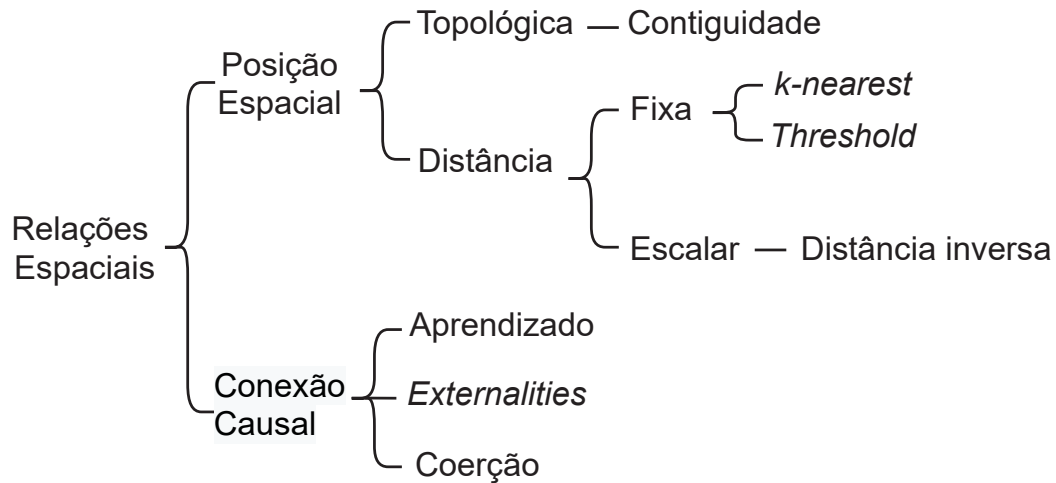


Figura 2.3: Diagrama das relações espaciais por categoria

Fonte: O autor (2021)

As análises espaciais feitas por cientistas políticos usam majoritariamente  $W$  baseadas na posição espacial dos objetos (Neumayer e Plümper, 2016, p. 179). A definição mais simples de conexão entre os  $OE$  é a que considera contiguidade de primeira ordem, binária e  $1\ root$  (Darmofal, 2006). Essa matriz é a mais presente nos  $OE$  de tipo área e define conexão entre unidades que dividem fronteiras imediatas. Usualmente é encontrada na literatura como *queen1* ( $Q1$ , aqui), nome baseado nos movimentos da rainha no tabuleiro de xadrez, e o número é baseado em quantas casas do tabuleiro a rainha caminha<sup>8</sup>.  $W$  com esse tipo de definição é binária no sentido de que ou os  $OE$  estão conectados ou não estão — uma relação espacial de 1 e 0 entre as unidades (ver Equação 2.1). A relação é  $1\ root$  porque o grau de influência que um objeto sofre é usualmente igual a um, então se  $i$  divide fronteira com quatro unidades cada uma delas tem 0,25 do peso total da dependência espacial.

$$W_{ij} = \begin{cases} 1, & \text{se } i \text{ e } j \text{ estão conectados} \\ 0, & \text{se não} \end{cases} \quad (2.1)$$

<sup>8</sup>A contiguidade também pode ser definida pelo movimento de outras peças no xadrez como o bispo, considerando apenas as conexões diagonais, e a torre, apenas conexões verticais e horizontais. Mas essas definições tem pouca utilidade para a Ciência Política (Darmofal, 2006).

A  $W$  de tipo Q1 assume a conexão entre  $OE$  com base na propriedade topológica básica *toca*. Na prática, são consideradas conectadas áreas (e.g. cidades) que dividem fronteiras. Outra categoria de conexão dependente da localização são as baseadas em distância. Diferentemente de Q1, ao considerar as distâncias, as conexões são menos dependentes da geometria dos objetos.

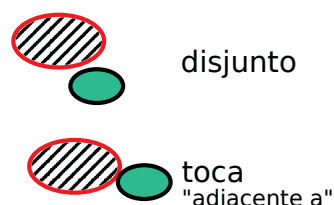


Figura 2.4: Relações espaciais topológicas de tipo Q1

Fonte: Adaptado pelo autor (2021) de Kainz (2010)

As relações por distância podem ser de tipo binário como *threshold* e *k-nearest neighbors* (k-NN) fixos ou contínuos como decaimento por distância ou o inverso da distância. O mais simples desses exemplos é a definição por *threshold*, um limite de distância (e.g. centímetros, metros, quilômetros), além do qual os  $OE$  não são considerados conectados. Outra definição comum é a  $W$  de tipo k-NN, um algoritmo que assume que  $i$  tem conexão com os  $k$   $OE$  mais próximos. Esse é um algoritmo classificador simples que considera a proximidade entre os objetos, mas não enquanto distância euclidiana. Outra alternativa com distância, mas com conexão escalar, é a  $W$  de distâncias inversas. Nessa configuração, assume-se que todos os  $OE$  estão conectados, mas a força da relação é proporcional à distância entre eles (Darmofal, 2006; Marzagão, 2013).

Assim como os  $OE$ , a definição de  $W$  é sobretudo teórico-conceitual. Mas para esse elemento não há tanta discussão. As definições de existência e força de conexão entre objetos são frequentemente atribuídas por conveniência e convenções estatísticas e não decorrentes de teoria (Neumayer e Plümper, 2016, p. 179). Para Neumayer e Plümper (2016) e Salvatore e Ruggeri (2021) as definições de  $W$  mais apropriadas para estudar fenômenos políticos são as que capturam o mecanismo causal pelo qual a dependência espacial opera. O argumento é que a distância *em si* não é a causa, "*spatial effects are caused by transactions, contact, or interactions between the sources and recipients of spatial stimulus*" (Neumayer e Plümper, 2016, p. 176). O trabalho de Kim et al. (2003) é um exemplo de  $W$  definida por relações econômicas entre os  $OE$ .

A próxima seção apresenta quais os  $OE$ , ou escalas, e as definições de  $W$  presentes nas explorações da GE brasileira.

## 2.3 OPERACIONALIZAÇÕES DA GE BRASILEIRA

Nesta seção, exploro como as pesquisas lidam com a disponibilidade de dados eleitorais e com a realidade territorial brasileira em suas operacionalizações. Apresento como as pesquisas da GE brasileira exploram os fenômenos de interesse e em quais pontos elas podem melhorar. Mais especificamente, trato das limitações nas operacionalizações dos dois componentes principais da análise espacial, o *OE* e *W*. O fim é indicar caminhos de desenvolvimento a partir dos dados eleitorais disponíveis.

### 2.3.1 Objetos Espaciais e Escalas nos estudos de voto no Brasil

Para entender as escalas de voto no Brasil é preciso primeiro entender a forma de organização territorial das eleições e da divulgação dos dados pelo TSE. O Código Eleitoral (Lei nº 4.737 de 1965) institui as regras de organização do sistema eleitoral brasileiro. A partir dessas regras, o TSE organiza as eleições com três unidades territoriais: a seção eleitoral, o local de votação (ou lugar da votação) e a zona eleitoral.

As *seções eleitorais* são as unidades mais desagregadas dos resultados eleitorais brasileiros. É a unidade de votação que possui uma mesa receptora de votos, com a urna e os mesários. O artigo 117 da lei, determina que o número de eleitores vinculados a cada seção eleitoral não seja menor que 50 e maior que 400, nas capitais, e 300, nos demais municípios.

O artigo 135 determina que cada seção é designada pelo juiz eleitoral para um *local de votação*, e no parágrafo 2º, a regra define que o local deverá ser, preferencialmente, um edifício público. Usualmente, o local de votação é uma escola e as diferentes seções eleitorais vinculadas ali são as diferentes salas de aula. Mas, os locais também podem ser igrejas, outros prédios públicos, como hospitais e prefeituras e até propriedades particulares.

A seção também é vinculada a uma *zona eleitoral*. Essa é uma unidade administrativa dos Tribunais Regionais Eleitorais (TREs) e a qual responde um juiz eleitoral. As zonas eleitorais seguem uma série de normas (Resolução nº 23.422) que envolvem número de eleitores por município e densidade demográfica com variação por região do país (e.g. na região norte uma zona eleitoral em municípios com densidade demográfica de até 2 hab/km<sup>2</sup> agrega um mínimo de 12 mil eleitores, já em capitais o mínimo é de 100 mil eleitores). Por isso, algumas zonas são dadas por agregação simples de municípios pequenos e outras por arranjos de pedaços de municípios populosos.

O eleitor é vinculado a uma seção eleitoral e, por consequência, um local de votação e zona eleitoral, pelo seu *domicílio eleitoral*. O artigo 42 do código eleitoral define que o domicílio eleitoral é, primeiro, definido pela proximidade ao lugar de residência ou moradia do eleitor, mas também permite vinculação por vínculos políticos, econômicos, sociais ou familiares.

Na prática as seções eleitorais, quando vinculadas aos locais de votação, se sobrepõem geograficamente. As seções eleitorais não possuem endereços próprios. Ou seja, os locais de votação são os *OE* pontuais *originais* dos dados de resultados eleitorais brasileiros. Todos os outros *OE* são agregações e arranjos a partir do ponto que agrega as seções. De outra forma, é a partir dos locais de votação que é possível vincular cada seção a uma outra escala.

Abaixo, apresento as escalas identificadas na produção brasileira.

- Em explorações de efeitos composicionais e contextuais sobre o voto:
  - Locais de Votação (do Nascimento, 2017; Faganello, 2017);
  - Bairros (Nicolau e Terron, 2012);
  - Distritos Administrativos da Cidade de São Paulo (Zolnerkevic, 2019);
  - Municípios (Avelar e Walter, 2008; Soares e Terron, 2008; Terron e Soares, 2010; Faganello e Fernandes, 2018);
  - Microrregiões (da Silva, 2014);
  - Regiões Administrativas do Estado de São Paulo (Meneguello e Neto, 2012);
  - Estados (Faganello e Fernandes, 2018);
- Em explorações relacionadas a sistemas eleitorais:
  - Locais de Votação (Fedozzi e Correa, 2015);
  - Zonas Eleitorais (Fleischer, 1976);
  - Municípios (Borges et al., 2016, 2011);
  - Mesorregiões (Cervi, 2009; Vieira, 2012; Borges et al., 2011)
  - Regiões de Influência das Cidades de São Paulo (Silotto, 2019);

Ao total, os trabalhos analisados aqui e em Gonçalves (2016, 2020) utilizam 10 *OE* nas análises de voto no Brasil, todos de caráter político-administrativo. Não é

comum encontrar justificativas quanto às escolhas de escala, sobretudo quando o *OE* é o município. Na maior parte das pesquisas a escala só é informada, mas sem nenhum desenvolvimento de significado substantivo. Abaixo organizo as justificativas encontradas para o uso de algumas delas.

### 2.3.1.1 Bairros

Os bairros são subdivisões intramunicipais, em alguns municípios são espaços de relações reconhecidos apenas por cidadãos, em outros são unidades estabelecidas por meio de leis municipais e, via de regra, não possuem autonomia (Cazzolato, 2005). Por serem definidos por meio do legislativo municipal, não há consenso sobre a função e fatores que tornam um território um bairro. Há casos em que os bairros são estabelecidos por lei e, inclusive, sub-divididos em sub-bairros, como o caso da Lei complementar nº 18, de 08 de Julho de 2004 de Feira de Santana (2004) na Bahia. como Feira de Santana. E há casos onde os bairros não são delimitados, reconhecidos, e não possuem funções de unidades operativas do poder público como foi o caso de São Paulo (Cazzolato, 2005).

Nicolau e Terron (2012) e Rodrigues-Silveira (2013) oferecem um conjunto de argumentos em favor dos bairros como escala de estudos de voto. Em resumo, os argumentos são de três ordens:

- Prática/analítica — garantem um número suficientemente relevante de casos para análises estatísticas (Nicolau e Terron, 2012, p. 4);
- Teórica/simbólica — são unidades tradicionais e com forte significado cultural e histórico (Nicolau e Terron, 2012, p. 3). Possuem significado reconhecido pelos atores políticos, servem de referência socioeconômica, cultural, étnica e de sociabilidades. É o local "por excelência onde se constroem identidades e se delimitam interesses" (Rodrigues-Silveira, 2013, p. 179);
- Administrativa/legal — os autores indicam que os bairros constituem unidades estáveis, estabelecidas e delimitadas administrativamente pelo poder municipal, e reconhecidas por outros órgãos, como o Instituto Brasileiro de Geografia e Estatística (IBGE) que considera essa unidade para os Censos (Nicolau e Terron, 2012, p. 3). Isso faz com que os bairros sejam objetos de intervenção

estatal e, por isso, são ambientes articuladores de interesses e conflitos políticos (Rodrigues-Silveira, 2013, p. 179)<sup>9</sup>.

### 2.3.1.2 *Municípios*

Os municípios são unidades autônomas, mas com baixa hierarquia. A sua criação, incorporação ou desmembramento depende de leis estaduais. No Brasil há hoje 5.570 unidades (IBGE, 2021). Essa é a escala mais presente nos trabalhos da GE do Brasil. Também é a escala para a qual existe o maior número de argumentos, são eles:

- Prática/analítica — alta disponibilidade de dados (Terron, 2009);
- Teórica/simbólica — para alguns autores, o município é o palco de disputas regionais entre elites (Avelar, 2008, p. 98). É “a unidade eleitoral **de fato** no Brasil”(de Carvalho, 2003, p. 68, negrito do autor), pois, em grosso modo, o município é a base dos deputados, para onde são direcionadas as emendas orçamentárias. Também é o território das alianças e campanhas eleitorais (Terron, 2009, p. 55). E, por fim, é a escala do *domicílio eleitoral* de todo político (Nicolau e Terron, 2012, p. 2) e existe uma tradição de associação dos candidatos a uma trajetória prévia na política municipal, "como é o caso de parlamentares ligados a uma família política com tradição no município, ex-prefeitos e ex-veredores, etc"(Borges et al., 2016, p. 34);
- Administrativa/legal — é um objeto constitucionalmente definido na estrutura administrativa do país (Nicolau e Terron, 2012, p. 2), um território institucionalizado (Terron, 2009, p. 55), é o distrito eleitoral nas eleições para prefeito e vereador (Nicolau e Terron, 2012, p. 2).

### 2.3.1.3 *Microrregiões*

As microrregiões fazem parte da Divisão Regional do Brasil proposta pelo IBGE, em 1987, para fins estatísticos. São partes das mesorregiões, espaços que agrupam municípios com base em semelhanças de quadro natural, rede de comunicação, identidade

---

<sup>9</sup>Embora essa possa ser a realidade no Rio de Janeiro, Fortaleza e Porto Alegre, os casos analisados pelos trabalhos citados, não há referência à obrigatoriedade da delimitação político-administrativa de bairros na Constituição Federal (1988) e nem nos regulamentos do Estatuto da Cidade (Lei nº 10.257, de 10 de julho de 2001). Ainda que exista a obrigatoriedade de um plano diretor, de desenvolvimento urbano, em municípios com mais de 20 mil habitantes, não há normas sobre a delimitação de bairros, ou de uso enquanto unidade administrativa.



regional e processos sociais. As mesorregiões são agregados espaciais que somam 558 unidades no país e são definidas por especificidades da estrutura de produção agropecuária, industrial, extrativista e de pesca, que resultam em relações sociais, econômicas, quadro natural, particulares dentro das mesorregiões. Os argumentos para o uso dessa escala são apenas de ordem prática/analítica:

- Prática/analítica — facilitam a análise gráfica, sendo um meio termo entre os municípios, uma escala muito grande, e as unidades federadas, uma escala muito pequena (Jacob et al., 2011, p. 191). Possuem número fixo de unidades, o que facilita a exploração longitudinal (do Socorro Braga e Zolnerkevic, 2020, p. 11).

#### *2.3.1.4 Regiões de Influência das Cidades de São Paulo*

Regiões de Influência das Cidades (REGIC) são agregações espaciais, realizadas por pesquisa do IBGE, que delimitam regiões a partir de relacionamentos entre as cidades brasileiras com base na análise dos fluxos de bens, serviços e gestão. No Estado de São Paulo o estudo define 15 regiões com essas características. O trabalho de Silotto (2019) apresenta os seguintes argumentos para justificar o uso dessa escala:

- Teórica/simbólica — essas regiões são ambientes de circulação de informações que propiciam vieses contextuais sobre os cidadãos (Silotto, 2019, p. 7-8).

#### *2.3.1.5 Regiões Administrativas do Estado de São Paulo*

As Regiões Administrativas de São Paulo são regionalizações definidas pela Secretaria de Planejamento e Gestão. Essa escala agrega os 645 municípios do Estado em 15 grandes regiões. O agrupamento de municípios é feito para fins de planejamento, prestação de serviços públicos e gestão da administração pública e são definidos por decreto do governador do Estado (e.g. decreto nº32.141, de 14 de Agosto de 1990). As agregações são justificadas pela dinâmica social, econômica e demográfica das regiões (SEADE, 2016). E esse é o argumento para a escolha de Meneguello e Neto (2012):

- Teórica/simbólica — a caracterização dessas regiões exercem papel importante na formação de preferências, dados por contexto socioeconômico e regional e relações entre centro e periferia (Meneguello e Neto, 2012, p. 135).

### 2.3.1.6 *Distritos Administrativos da Cidade de São Paulo*

Os Distritos Administrativos da Cidade de São Paulo são áreas administrativas, referências para a administração pública municipal (Lei nº 10.932, de 15 de Janeiro de 1991). Zolnerkevic (2019) utiliza 57 dos 91 distritos e os agrega em quatro classes com base em renda média da região. O argumento utilizado é de ordem prática/analítica:

- Prática/analítica — os dados disponíveis para a pesquisa realizada possuem disparidade no número de casos por unidade de observação. A agregação foi feita como forma de uniformizar o número de casos por unidade (Zolnerkevic, 2019, p. 200).

### 2.3.1.7 *Locais de Votação*

(LV)

Os LV são os menores *OE* da divulgação dos resultados eleitorais brasileiros. No pleito de 2018, foram 94.246 locais para os quais o eleitorado se deslocou para votar. As pesquisas de Faganello (2017); do Nascimento (2017); Fedozzi e Correa (2015) apresentam os seguintes argumentos para a defesa dessa unidade de análise:

- Prática/analítica — para Faganello (2017, p. 85) essa é a maneira mais otimizada de análise de dados eleitorais no espaço. do Nascimento (2017, p. 101) justifica o uso de LV pelo alto grau de granularidade obtido, que permite observar fenômenos de maneira muito localizada e precisa. Na linha de pesquisa em conexão eleitoral, Fedozzi e Correa (2015, p. 20) indicam que permite ampliar as possibilidades analíticas, possibilitando a identificação de uma diversidade de padrões espaciais que de outra forma são classificados de forma genérica;
- Teórica/simbólica — corresponde melhor à distribuição efetiva dos fenômenos eleitorais (Faganello, 2017, p. 36).

Na próxima seção faço o mesmo procedimento de organização de justificativas da operacionalização das pesquisas, mas para as definições de *W*.

### 2.3.2 *W, Brasil*

O trabalho de Ames (2003) é o primeiro, até onde consta, a utilizar estatística espacial para explorar os dados eleitorais brasileiros. O autor utiliza o índice *I* de Moran,

mas a definição de  $W$  não é descrita de forma clara: "as matrizes de vizinho mais próximo usadas para calcular a estatística de Moran I foram derivadas das coordenadas dos mapas"(Ames, 2003, p. 382).

Além de Ames (2003), ao menos outros sete trabalhos utilizaram o mesmo índice, ou a versão local dele, que é dependente de definição de  $W$ : Soares e Terron (2008), Terron (2009), Terron e Soares (2010), Marzagão (2013), Rodrigues-Silveira (2013), Gonçalves (2016), do Nascimento (2017) e Faganello (2017). Nenhum deles utiliza matriz do tipo não-espacial, mas estão presentes as topológicas e de distâncias binárias e escalares.

### 2.3.2.1 *Relações Topológicas*

A  $W$  de tipo topológico assume a conexão entre  $OE$  com base na propriedade básica *toca*. Os trabalhos com dados eleitorais brasileiros utilizaram três configurações para definir as conexões desse tipo:

- Q1 — a configuração mais presente é a que considera a relação entre as unidades imediatamente vizinhas, ou contíguas. Presente nos modelos de Soares e Terron (2008), Terron (2009), Terron e Soares (2010), Marzagão (2013), Gonçalves (2016), do Nascimento (2017) e Faganello (2017). Marzagão (2013, p. 276) e Terron (2009, p. 63-64), são os únicos que justificam a decisão por essa operacionalização, apontam para a dificuldade em definir  $W$  pela enorme heterogeneidade territorial dos municípios. Marzagão (2013, p. 276) indica que os padrões históricos de colonização brasileira resultaram em pequenos municípios no Sul, Sudeste e Nordeste e grandes no Norte e em parte do Centro-Oeste, por isso não é possível encontrar uma distância única para representar todas as regiões do país.
- Contiguidade Torre — a pesquisa da autora do Nascimento (2017, p. 112) utiliza três definições de  $W$ , entre elas as conexões verticais e horizontais (do movimento da torre no xadrez). A autora não oferece argumentos que justifiquem essa decisão.

### 2.3.2.2 *Distância*

Duas pesquisas utilizaram  $W$  estabelecidas por distância entre os  $OE$ :

- Distância fixa definida por pico de *Z-score* — Faganello (2017) utilizou o método de *Incremental Spatial Autocorrelation* para o  $W$  do indicador local de autocorrelação. Essa é uma ferramenta que vai aumentando gradativamente o *threshold* de distância entre os *OE* até chegar ao pico de significância estatística da autocorrelação do modelo. O método indica em qual distância fixa o teste de autocorrelação resulta no maior resultado de significância no *Z-score*.
- *8-nearest* — o artigo do Rodrigues-Silveira (2013) utiliza uma  $W$  que relaciona os oito *OE* mais próximos, mas não explica essa decisão.
- Inverso da distância — o trabalho do Marzagão (2013) utilizou uma matriz de distâncias inversas medidas a partir do centroide dos municípios. Na operacionalização do autor, cada objeto é relacionado a todos os demais objetos, mas os pesos são inversamente proporcionais às respectivas distâncias.

### 2.3.3 Discussão

O objetivo de uma operacionalização empírica é buscar a melhor forma de representar a realidade com os dados disponíveis. No caso da operacionalização da análise espacial de voto no Brasil, buscamos representar, em um mesmo modelo, um território de grandes municípios com baixa densidade de eleitorado e regiões metropolitanas de alta densidade. Ao mesmo tempo que a superfície é composta por relevos e obstáculos territoriais dos mais diversos (e.g. rios, montanhas, áreas de preservação ambiental). E para isso temos os dados eleitorais divulgados por seções eleitorais, que se agrupam em LV e podem ser agregadas e arranjadas em diferentes escalas — ao menos 11, como os dados apresentados aqui mostram. Mas como modelar os atores e as relações entre eles nesse universo?

A operacionalização é sobretudo teórica. Tanto o *OE* quanto a  $W$  são entidades vazias que dependem do sentido que o pesquisador desenvolve. Para sete dos *OE* encontrei argumentos que os justificam — Tabela 2.1.

Tabela 2.1: Presença de justificativas para definição de Objetos Espaciais

Objeto Espacial	Prática/analítica	Teórica/simbólica	Administrativa/legal
Bairros	✓	✓	✓
Municípios	✓	✓	✓
Microrregiões	✓	n/a	n/a
REGIC	n/a	✓	n/a
Regiões Administrativas do Estado de São Paulo	n/a	✓	n/a
Distritos Administrativos da Cidade de São Paulo	✓	n/a	n/a
Locais de Votação	✓	✓	n/a

Fonte: O autor (2021).

As justificativas teórica/simbólica são de dois tipos: as de explorações contextuais e as mais características de sistemas eleitorais.

As justificativas de base em teorias contextuais são utilizadas para os bairros (forte significado cultural e histórico; referências de sociabilidades; local de construção de identidades e interesses), REGIC (ambiente de circulação de informações), Regiões Administrativas do Estado de São Paulo (lugar de formação de preferências), LV (maior correspondência com o lugar do fenômeno eleitoral). Em escalas grandes é difícil transpor as teorias de efeitos contextuais. Qualquer escala maior do que o espaço de vizinhança onde ocorre a interação social pode estar desalinhada da teoria (Johnston et al., 2001, p. 88). Por essa crítica, os bairros e LV seriam os *OE* mais adequados para esse tipo de exploração. O primeiro, pela força de construção de identidades e sociabilidade. O segundo, por ser a menor unidade e com uma média de aproximadamente 300 eleitores, o que representa melhor as diferenças locais (Gonçalves, 2016).

Os LV servem de *proxy* para uma vizinhança. Os TREs exigem comprovante de residência atualizado para determinar o LV de cada eleitor/a. Resultados de uma pesquisa realizada pelo IBOPE Inteligência, em 2014, mostraram que 88% dos entrevistados afirmaram votar próximo do domicílio, 4% responderam que é longe de casa, mas ainda assim é possível ir a pé e 8% responderam que é longe de casa, e é preciso pegar ao menos uma condução para chegar ao local (do Nascimento, 2017). Hoje, devido ao movimento de recadastramento biométrico do eleitorado, os LV, em tese, são uma representação ainda melhor de onde o/a eleitor/a mora. E.g, em 2018, aproximadamente 90% do eleitorado do Paraná passou por recadastramento.

Os trabalhos mais relacionados a explorações de sistemas eleitorais, sobretudo a conexão eleitoral, apresentam argumentos para a utilização da escala de município. Em resumo, defendem que são territórios das alianças e disputas entre elites, tradicionalmente associados a políticos e famílias políticas específicas, além de serem a unidade que recebe as emendas orçamentárias. Essa é uma linha que pode se beneficiar de *OE* menores. A literatura que considera o município como "unidade eleitoral de fato" acaba por ignorar a rede de operação política que se ramifica em diversas escalas intramunicipais até chegar em *microbrokers* que também se organizam espacialmente. É certo que manter vínculos políticos territorialmente ancorados, sobretudo, com destinação de emendas orçamentárias está entre as principais atividades dos parlamentares (Meireles, 2019, p. 137). Mas essa não é uma relação que se esgota entre deputado e prefeito, de maneira isolada. *Brokers* atuam em uma rede de relações de diferentes níveis entre os representantes formais e informais — Figura 2.5. São atores que "possuem conexões substantivas com territórios da cidade, agrupamentos em geral espacialmente localizáveis em que se configuram dinâmicas sociais e de poder (seja em ruas, bairros, comunidades, associações, movimentos, conselhos participativos etc.)" (Hoyler, 2020, p. 11). Para os quais as explorações dão pouca atenção.

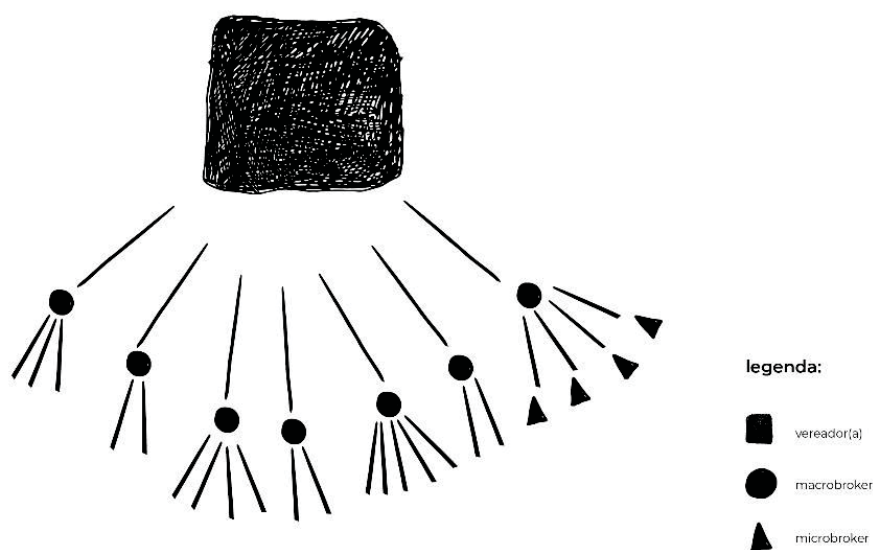


Figura 2.5: Estrutura da rede de operação política (*brokerage*)

Fonte: Hoyler (2020).

Os bairros e municípios são os *OE* para os quais os autores mais desenvolveram argumentos. São os únicos que apresentam definições do tipo "administrativa". O fato da unidade ser delimitada administrativamente traz os benefícios de disponibilidade de dados, por serem uma referência para intervenções de políticas públicas. Mas, a estabilidade da delimitação do *OE* é um limitador da territorialidade dos fenômenos político-eleitorais: "territórios podem ser construídos e desconstruídos continuamente pela ação de indivíduos ou grupos que exercem algum domínio ou poder em determinada área" (Terron, 2009, p. 319). A concepção contemporânea de territórios entende que o espaço está em constante mudança pelas ações de indivíduos e grupos. O domínio do território não se resume ao estatal, nessa linha mais recente também são discutidas a territorialidade do tráfico, da prostituição, dos trabalhadores sem terra (Terron, 2009, p. 54).

Se a concepção é de que os territórios são construídos e desconstruídos continuamente, como usar uma unidade tão estável e estabelecida por atores oficiais? Ou melhor: como usar apenas as unidades administrativamente estabelecidas? Por esses argumentos, as explorações a partir de LV parecem uma oportunidade de pesquisa inovadora. Explorações sem delimitações pré-estabelecidas podem revelar padrões desconhecidos dos fenômenos políticos.

A maior limitação encontrada é quanto à definição de *W*. Um componente com alto impacto nos resultados e pouco desenvolvimento na produção analisada. A literatura só elaborou justificativas para a matriz de tipo Q1. Os argumentos para justificar a operacionalização de relações espaciais com contiguidade imediata entre os objetos é baseada na disparidade do tamanho dos municípios. Nenhuma das referências ofereceu justificativas teóricas.

Distância é uma informação espacial essencial para os modelos de análise. Entretanto, a literatura avançou pouco nas formas de exploração dessa informação. A especificação de *W* é mais importante do que os trabalhos da área sugerem com a falta de especificação. É documentado, ver Salvatore e Ruggeri (2021) e Neumayer e Plümpner (2016), que as especificações de conexões topológicas e binárias não são representações satisfatórias da realidade e dos mecanismos causais previstos em teorias da Ciência Política. Dos modelos apresentados pela literatura, a *W* utilizada por Marzagão (2013), de distâncias inversas, é a mais robusta, por definir os pesos de conexão de forma escalar. E a mais distante da realidade das conexões previstas em teorias da GE é a utilizada por do Nascimento (2017). Não há justificativa para a Ciência Política utilizar uma matriz que só considere as conexões verticais e horizontais de uma área (Darmofal, 2006).

Em suma, embora não exista um modelo "certo", uma operacionalização promissora é a que considera os LV para definição de  $OE$  e explora mais profundamente as definições de distâncias escalares para formar a  $W$ .

## 2.4 CONSIDERAÇÕES FINAIS

Este capítulo foi motivado pelo objetivo de analisar as operacionalizações da GE no Brasil, com o fim de indicar caminhos de desenvolvimento metodológico com os nossos dados eleitorais. Resumi o problema na seguinte pergunta:

*Pergunta de pesquisa 1:* Como podemos aprimorar a operacionalização dos componentes principais da análise espacial dos votos no Brasil?

Para respondê-la, realizei uma síntese qualitativa da produção da área. Organizei os argumentos e justificativas usados para embasar as escolhas de  $OE$  e  $W$ .

*Resposta:* A principal limitação das operacionalizações dos modelos de análise espacial da GE brasileira é em relação as  $W$ . As justificativas são escassas e pouco trabalhadas. O que mostra, que podemos aprimorar o modelo com maior atenção aos pressupostos teóricos e à superfície modelada. Quanto aos  $OE$ , os padrões intramunicipais podem revelar padrões desconhecidos, principalmente a partir de LV, sem limites de área pré-estabelecidos.



### 3 GEOGRAFIA ELEITORAL: O CAMPO, A PRODUÇÃO BRASILEIRA E SUAS LACUNAS

A pesquisa deste capítulo combina métodos bibliométricos e cientométricos para visualizar, analisar e avaliar a atividade científica e a produção da GE. Um ponto central desta tese é o de que a GE tem grande potencial de desenvolvimento na Ciência Política Brasileira (CPBR). Por isso, o objetivo que guia a investigação em tela é mapear o estado da arte e identificar lacunas e tendências de pesquisa para estabelecer prioridades de investimento na área. Esse tipo de estudo, que explora o campo científico, ajuda a subsidiar tomadas de decisões (Grácio, 2020). Em suma, meu ponto é que a GE tem espaço para crescer na CPBR e aqui aponto em que direção podemos crescer. Aqui os resultados nos ajudam a responder a seguinte pergunta:

*Pergunta de pesquisa 2:* Quais os principais desalinhamentos entre a produção da Geografia Eleitoral brasileira e as tendências internacionais?

O texto está estruturado em três partes. A primeira parte apresenta as motivações para a realização da pesquisa e resume as estratégias metodológicas adotadas. A segunda parte traz os resultados da pesquisa e está subdividida pelo método aplicado. Apresento primeiro uma análise bibliométrica da literatura e na segunda uma análise cientométrica do campo de estudos. Com isso, busco mapear a organização e o desenvolvimento da GE. Por fim, os resultados das duas análises são usados para sustentar indicações de lacunas e oportunidades para o desenvolvimento da agenda de pesquisa da GE na CPBR.

#### 3.1 RATIONALE E ESTRATÉGIA METODOLÓGICA

Descrever a *geografia eleitoral* não é uma tarefa simples. Os dados, que serão apresentados neste capítulo, indicam que o termo foi usado em artigos publicados em 107 periódicos, em ao menos 28 áreas do conhecimento, escritos por 268 autores/as, vinculados a 170 instituições de 35 países diferentes. Trata-se de um campo grande e complexo, naturalmente multi-temático e multidisciplinar e com mais de 100 anos de produção. A heterogeneidade das publicações é um indicativo do potencial da área.

Diversas teorias, estruturas conceituais, aplicações e desenvolvimentos empíricos da GE estão em andamento desde o início do século 20 (Johnston e Pattie, 2006).

Embora existam diversas revisões narrativas sobre diferentes aspectos da GE — e.g., sobre sua gênese (Zavala, 2012), trajetória histórica (Terron, 2012), reflexões metodológicas (Sonnleitner, 2013), conceitos principais (Jones et al., 2004), principais objetivos e problemas (Agnew, 1990), técnicas de análise espacial (O’Loughlin, 2008) e resumos gerais (Johnston e Pattie, 2006; Rodrigues-Silveira et al., 2017) — só foi possível encontrar uma análise da área a partir de métodos quantitativos e com busca sistemática (Leib e Quinton, 2011). Diferentemente da análise realizada neste capítulo, Leib e Quinton (2011) consideram a GE apenas enquanto subárea da Geografia. A fonte do material utilizado pelos autores é formada exclusivamente por periódicos da Geografia. Os critérios de inclusão do material também limitam a análise para artigos publicados entre 1990 e 2007 e escritos em língua inglesa. Outra diferença significativa é que os autores exploram o material a partir de estatísticas descritivas simples e análise qualitativa de tópicos.

A fim de fornecer um mapa que oriente o desenvolvimento da GE brasileira, combino métodos bibliométricos e análise de rede. Com a análise bibliométrica é possível quantificar as características das pesquisas do campo e com isso descrever as tendências e panoramas intelectuais (Kim et al., 2016). Com a análise cientométrica, a partir de redes, é possível apreender como os atores do campo se articulam e quais as estruturas e hierarquias da produção da área (Horochovski et al., 2019). A identificação da estrutura de conexão serve para orientar os pesquisadores do campo sobre a agenda de investigação (Feretti et al., 2018).

Para tanto, foram coletados todos os artigos publicados em revistas indexadas pela plataforma *Web of Science* incluindo *SciELO Citation Index* e *Core Collection* com a identificação “*Electoral Geography*”<sup>1</sup>. Mais especificamente, utilizo as análises de *co-citation* e *co-word*:

- *Co-citation* ou Acoplamento Bibliográfico — Método para modelar a estrutura do campo científico assumindo que o compartilhamento de referências indica a existência de uma relação intelectual significativa. Os resultados indicam relações entre ideias, conceitos-chave e métodos entre os atores dentro do campo (Small, 1973).

---

<sup>1</sup>No Apêndice A apresento detalhes dos métodos e procedimentos empregados na análise.

- *Co-word*, Co-ocorrência de Palavras ou Comunidades Temáticas — Método que mostra a estrutura relacional dos termos identificadores da produção científica (Callon et al., 1983). Assume-se que as palavras-chave fazem referências ao objeto da pesquisa, aos métodos, temas e perguntas, funcionando como *proxy* do conteúdo do artigo (Strozzi et al., 2017).

As próximas seções apresentam os resultados bibliométricos e cientométricos que formam o *mapa* da GE, começando pela análise descritiva dos atributos estáticos da produção e, em seguida, tratando das características relacionais dessa produção.

## 3.2 RESULTADOS

### 3.2.1 Análise Bibliométrica

A análise feita nesta seção é uma abordagem descritiva de atributos bibliométricos. Exploro as distribuições de seis variáveis bibliométricas: autoria, ano de publicação, país, número total de citações, periódico e área de pesquisa do periódico. Também foram incluídas duas variáveis *dummy* como forma de isolar a Ciência Política e o Brasil para a análise dos dados.

O corpus analisado é formado por um total de 226 artigos, distribuídos entre 1959 e 2020, que perfazem uma média de 3,7 artigos por ano. Entretanto, é preciso lembrar que a história do campo, usualmente, é traçada a partir de André Siegfried. O autor é amplamente indicado como fundador da GE (Zavala, 2012; Terron, 2012; Sonnleitner, 2013; Rodrigues-Silveira et al., 2017; Gonçalves, 2016). Siegfried discutia *l'aspect géographique* das eleições francesas ao menos desde 1913. A primeira publicação indexada pode indicar o início da versão contemporânea da produção da área.

A Figura 3.1 mostra que a média por ano não tem uma distribuição proporcional entre o período ( $\sigma=4,878$ ). Nas primeiras duas décadas a produção total é de sete artigos e a área permanece com um número de publicações anuais menor que três até a metade dos anos 1990. Os primeiros 36 anos, do total de 61 anos da produção mapeada, representam menos de 11,9% do total. Diferentemente da realidade entre 1996 e 2000, quatro anos somados representam 13,7% da amostra — com destaque para a produção em 1996 e 1997, com nove trabalhos publicados em cada ano, marca que só foi igualada em 2008. Leib e Quinton (2011, p.10) indicam dois possíveis motivos dos picos de produção da GE para o período 1995-2002. O primeiro fator é o incentivo dado pelos periódicos nessa época, como edições especiais sobre o tema. O segundo diz respeito à eventos políticos

de importância para a GE, como as decisões da Suprema Corte Norte-Americana sobre redistribuição das eleições presidenciais e as eleições pós-guerra fria em ex-estados comunistas (Leib e Quinton, 2011).

Uma diferença entre a produção do Brasil frente a dos outros países é em relação à presença de trabalhos da área de Ciência Política. No Brasil, outras áreas aparecem concentradas em 2012 e representam 20% do total das produções do país. Considerando todos os demais países, os trabalhos de outras áreas aparecem desde o começo da amostra e representam 64,3% da produção total. Em alguns casos, como o de 2006, todos os trabalhos publicados são de outras disciplinas. Esses dados apontam para um baixo diálogo entre disciplinas na GE no Brasil.

No geral, considerando toda a produção, a Figura 3.1 apresenta uma distribuição assimétrica negativa, com alta concentração entre 2015 e 2019. Essa é uma distribuição esperada dada a democratização do acesso a ferramentas cartográficas e geoestatísticas, e bancos de dados geoespacializados abertos (Terron, 2012). Entretanto, esse não é o padrão encontrado ao considerar a produção feita no Brasil. Dos 226 artigos publicados, nove foram elaborados por pesquisadores de instituições brasileiras. Os primeiros trabalhos brasileiros encontrados na pesquisa datam de 2008 e não é possível verificar um aumento consistente na década seguinte. Isso evidencia um descompasso entre o desenvolvimento internacional e o brasileiro da GE.

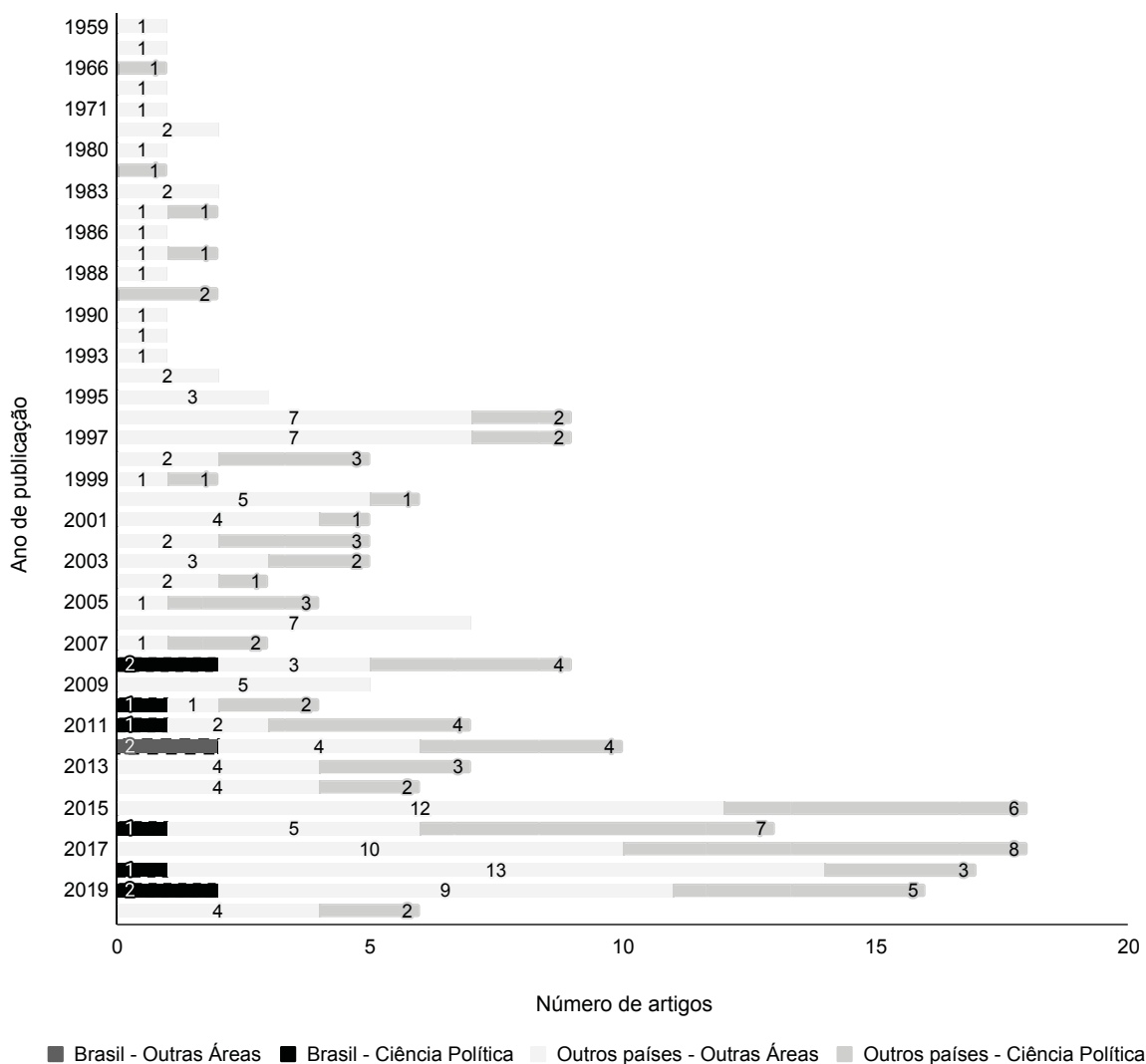


Figura 3.1: Evolução no número de artigos (1959-2020), por país (Brasil e outros) e áreas (Ciência Política e outras)

Fonte: O autor (2021)

Os artigos estão vinculados a instituições de 35 países, que variam de um a 68 artigos publicados e de zero a 996 citações recebidas totais, como mostra a distribuição na Figura 3.2.

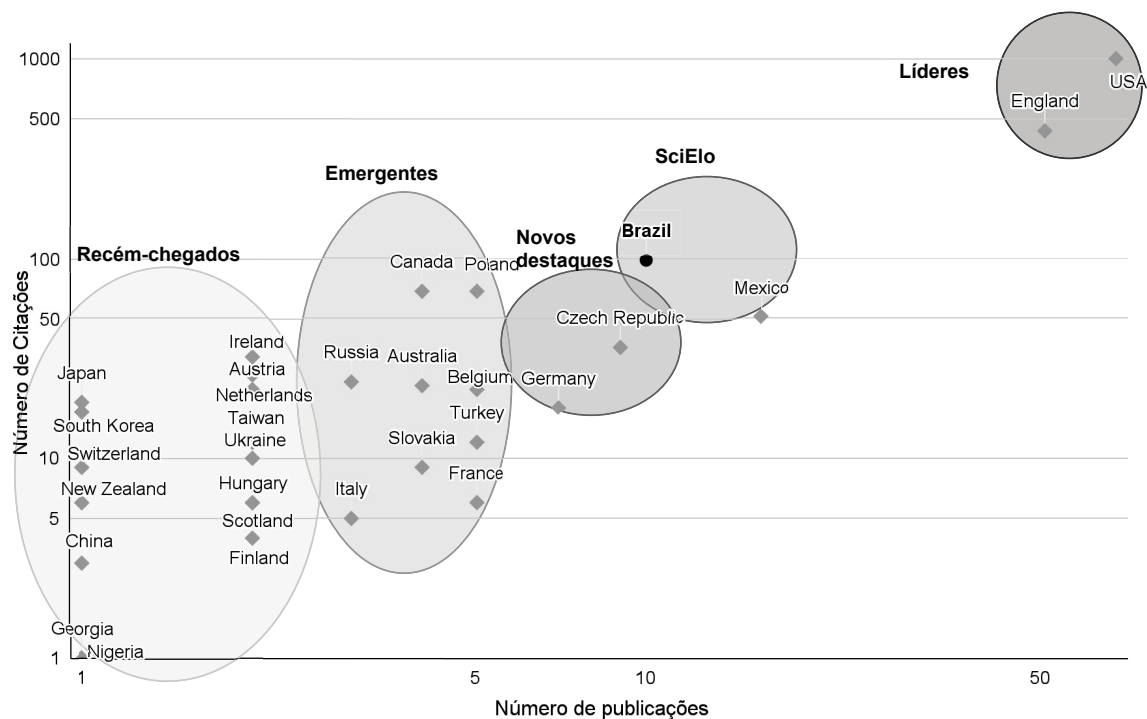


Figura 3.2: Distribuição de países por número de artigos publicados e citações recebidas

Fonte: O autor (2021)

Nota: Eixos apresentados em escala logarítmica de base 10. O mesmo artigo pode ser escrito por autores/as vinculados a instituições de mais de um país, a soma total aqui é de 227, sendo que 28 artigos não contam com nenhuma identificação de instituição e em outros 10 casos uma autoria não possui essas informações.

Os países podem ser divididos em cinco grupos, considerando as relações entre número de citações e de publicações:

- **Recém-chegados** — O resultado mostra que 20 países (57,1% da amostra) podem ser considerados *latecomers*. Os trabalhos desenvolvidos por esse grupo começam em 1998 (39 anos depois do primeiro caso), na Hungria, e em 4 casos (Nigéria, Espanha, Colômbia e Ucrânia) a primeira publicação foi realizada em 2019. Em número de publicações, o grupo conta com 12 países com uma publicação e oito com duas, já em citações o grupo tem um mínimo de zero (Argentina, Irã, Colômbia e Espanha) e máximo de 32 (Irlanda). Não parece existir outro padrão que una esses novos produtores da área, já que fazem parte desse grupo a Suíça e Nigéria, de notável diferença de desenvolvimento econômico e democrático, e China e Nova Zelândia, com sistemas de governo incomparáveis.

- Emergentes — Os dados mostram uma classe intermediária com 10 países que somam 38 artigos (16,7% das publicações). O número de artigos varia entre três (Itália e Rússia) e cinco (Bélgica, França, Polônia e Turquia). Já as citações desse conjunto de países varia entre 5 (Itália) e 68 (Canadá e Polônia), com total de 255. Essa é uma classe heterogênea quanto ao ano de entrada nas produções. De um lado, a França conta com sua primeira publicação em 1966, do outro, a Eslováquia com a primeira publicação em 2016.
- Novos destaques — Alemanha e República Tcheca também são recém-chegados, mas contam com sete e nove publicações respectivamente. As produções em instituições alemãs vão de 2014 a 2018 e 2012 a 2020 no caso tcheco. A primeira soma 18 citações e a segunda 36 citações totais.
- SciElo — México e Brasil são os próximos nessa lista com 16 e nove publicações respectivamente, mas isso pode ser explicado pela inclusão da base de dados SciElo. Considerando apenas a principal base da Web of Science, o Brasil conta com uma publicação e o México com três. No total, os artigos brasileiros somam 98 citações, o que coloca o país na terceira posição, e México conta com 51 citações.
- Líderes — Os resultados apontam alta concentração da produção nos Estados Unidos da América (EUA) e Inglaterra, 52,4% dos artigos foram produzidos em instituições desses países. Os EUA são o principal produtor da GE, com 68 publicações, apresentando trabalhos desde 1969, mas as pesquisas se intensificam a partir de 1995, chegando ao topo em 2015, com sete publicações no ano. A Inglaterra é o segundo país em produção total com 51 trabalhos e pesquisas publicadas desde 1971, tendo seu auge de produção em 2019, com seis artigos. A concentração é ainda maior na soma de citações, 1.433 ou 70% das citações estão nesse grupo, sendo 437 para Inglaterra, uma média de 8,56 citações por artigo, e 996 para os EUA, média de 14,64 citações por artigo.

A pesquisa de Leib e Quinton (2011) mostrou o mesmo padrão de concentração das publicações nos EUA e Inglaterra: considerando a produção em periódicos da Geografia, esses países somam 59% dos artigos. Mas, uma diferença entre os resultados da regionalização da produção do campo é o caso da Rússia. Aqui a Rússia aparece na base da classe intermediária. Nos resultados dos autores/as a Rússia aparece com destaque, em

terceiro lugar na produção total, principalmente como reflexo do interesse nos processos eleitorais pós-dissolução da União Soviética.

Os trabalhos analisados foram produzidos por pesquisadores vinculados a 170 instituições diferentes, mas, assim como a distribuição dos artigos por país, a produção apresenta concentração em instituições da Inglaterra e dos EUA. São quatro universidades (Universidade de Bristol, Universidade de Sheffield, Universidade Internacional da Flórida e Universidade do Colorado) que agregam mais de 20% da produção total. Na outra ponta, 68,8% (n=117) das instituições constam em uma publicação e 16,5% (n=28) em duas publicações. Esses dados mostram que a área é formada por uma grande diversidade de instituições, mas reforçam a posição dos EUA e da Inglaterra enquanto centros da produção da GE.

No Brasil, a Universidade Estadual de Campinas (UNICAMP) e a Universidade de Brasília (UNB) são as principais instituições, com seis e cinco presenças, respectivamente. Entretanto, o Instituto Universitário de Pesquisas do Rio de Janeiro (IUPERJ) e a Universidade do Estado do Rio de Janeiro (UERJ) são destaques por serem as instituições da autora e do autor (Soares e Terron, 2008; Terron e Soares, 2010) que concentram mais de 75% das citações.



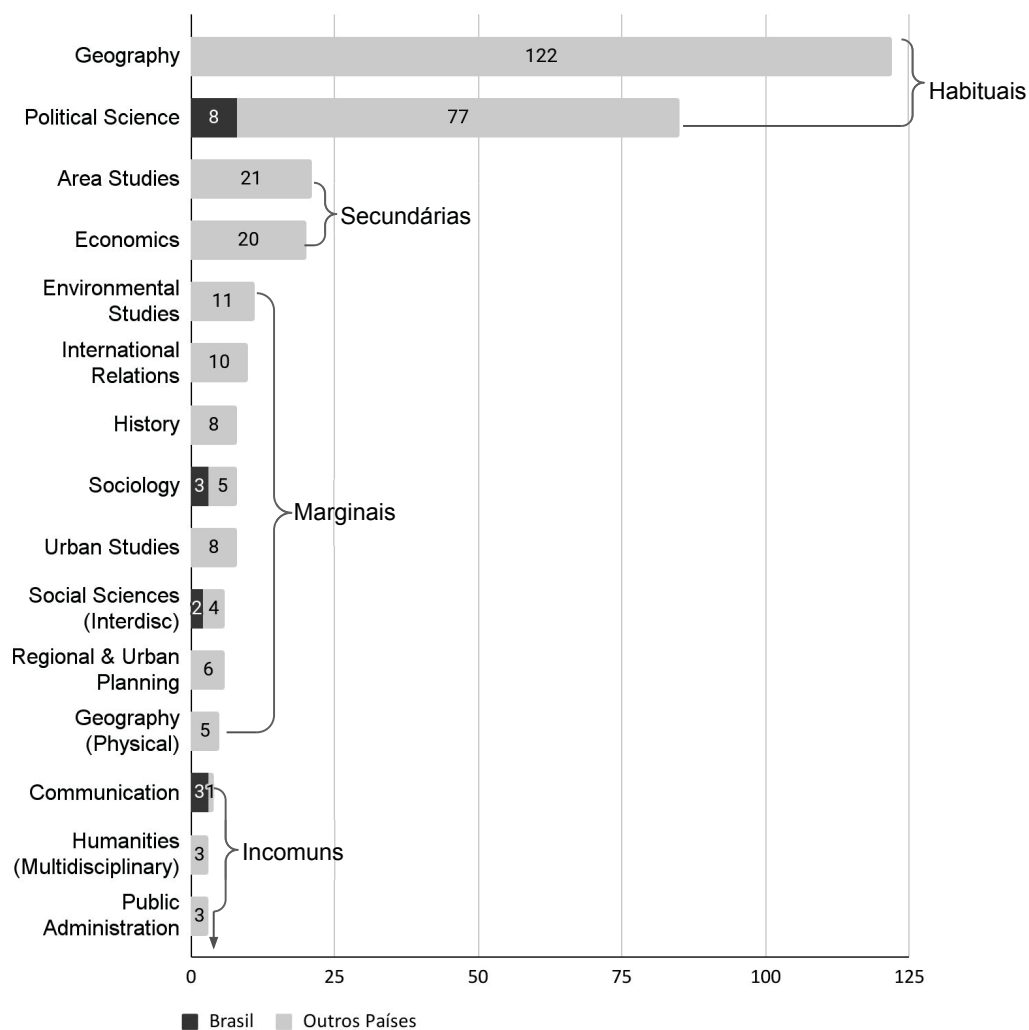


Figura 3.3: Distribuição de Áreas de pesquisa dos artigos

Fonte: O autor (2021)

Nota: Áreas incomuns com menos de três casos foram ocultadas da visualização, são elas: Direito, Demografia e Estudos de Desenvolvimento com duas presenças e; Linguagem e Linguística, Literatura, Estudos Culturais, Problemas Sociais, Relações Industriais e de Trabalho, Ciências Multidisciplinares, Matemática, Abuso de Substâncias e Saúde Pública, Ambiental e Ocupacional que contam com uma presença. O mesmo artigo pode ser indexado em mais de uma área de pesquisa, o total aqui é de 336 áreas para os 226 casos.

A GE é interdisciplinar por natureza e os dados confirmam isso para o cenário geral. Os trabalhos estão vinculados a 63 departamentos e a 28 áreas de conhecimento. Existe uma heterogeneidade de disciplinas que vai desde explorações em áreas como Abuso de Substâncias (n=1), Direito (n=2), Literatura (n=1), Comunicação (n=4) e

Matemática ( $n=1$ ). A distribuição entre as áreas tem alta variação ( $\sigma=26,16$ ). É possível dividi-las em quatro grupos:

- Habituais — 207 casos, ou 61,61% da amostra, estão concentrados nas áreas que podemos considerar comuns ou esperadas, a Geografia e Ciência Política;
- Secundárias — A área de Estudos de Área e a de Economia formam a “segunda fila” das produções, somando 41 presenças, que representam 12,2% dos casos;
- Marginais — Oito áreas podem ser consideradas como periféricas, variam entre cinco e 11 presenças, somando 18,4% da amostra; e
- Incomuns — 16 áreas apresentam menos de quatro presenças, somando 7,7% dos casos.

Os trabalhos brasileiros se concentram majoritariamente na Ciência Política, mas se destacam nas áreas marginais de Sociologia e Ciências Sociais (interdisciplinar) e é destaque na área incomum de comunicação (esse resultado é explicado pelo periódico de publicação dessas pesquisas). O Brasil não apresenta nenhum caso na Geografia, a principal área de exploração, nem nas áreas secundárias e nas principais áreas periféricas. Essa é uma lacuna importante, que impacta os temas explorados e métodos aplicados na GE no país.

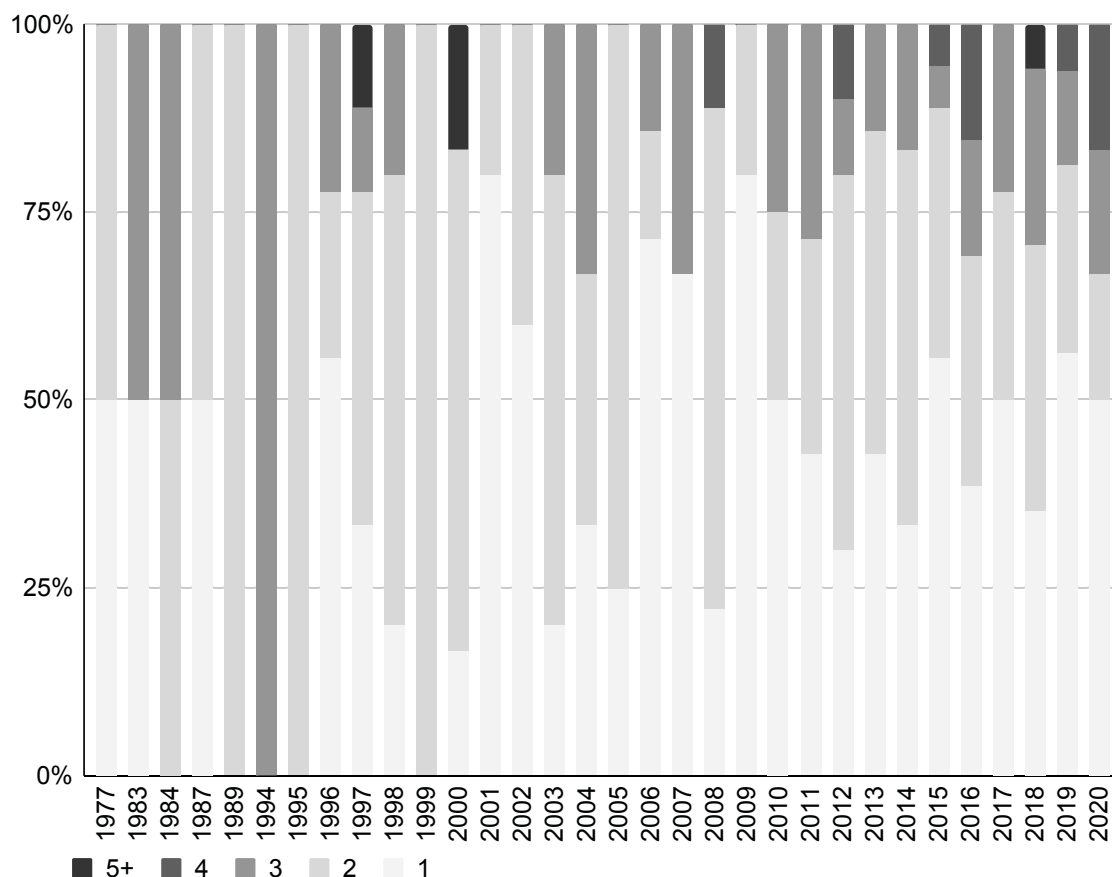


Figura 3.4: Número de autores por artigo, 1997-2020

Fonte: O autor (2021)

Nota: 1959, 1965, 1966, 1969, 1971, 1980, 1982, 1986, 1988, 1990, 1992 e 1993 estão ausentes pois em cada um desses anos foi publicado apenas um artigo.

Os artigos foram escritos por 268 autores/as únicos, com média de 1,827 autor por artigo. Os números indicam que a área apresenta baixo movimento de trabalhos em equipe. Considerando toda a amostra, são 99 trabalhos escritos por apenas uma autoria, que representam 43,8% da produção total, 83 (36,7%) por duas autorias, 34 (15%) por três, sete por quatro (3%) e três por seis (1,3%).

Não há uma tendência clara na dinâmica da produção ao longo dos anos, como o Gráfico 2 evidencia. Há, e.g., em 2008 um pico de 78% de artigos em grupo, seguido de 80% de produções de autoria única em 2009. Considerando todo o período, os resultados não mostram tendência de aumento de coautorias ao longo dos anos ( $R^2 = 0,026$ ). Mas é possível verificar que, a partir de 2011, a produção em equipe se intensifica e a variação

anual se torna menor. O resultado acompanha uma tendência observada em outros campos, já que “a crescente complexidade do ofício acadêmico impõe o trabalho em equipe” (Horochovski et al., 2019, p. 600).

O Brasil apresenta um cenário de maior produção em coautoria, pois 72% dos artigos foram escritos por duas ou três pessoas, quase 16 pontos percentuais acima do geral. Mas, considerando apenas os trabalhos da Ciência Política, a variação não apresenta diferença significativa: 44% das publicações foram assinadas por uma única pessoa.

Os artigos apresentam um total de 9.352 referências<sup>2</sup>, com uma média de 41,4 referências citadas por artigo. São 4.302 citações únicas<sup>3</sup>, sendo que 72,3% (n=3.112) delas possui uma presença. Esse resultado é um indicativo da diversidade de discussões movimentadas pela GE, mas também pode indicar baixa coesão da área.

Assim como outros resultados já citados, os dados indicam que a concentração de autores/as ingleses e americanos é alta: entre os 20 autores/as mais citados da área, 11 são ingleses e nove americanos; considerando os trabalhos da Ciência Política, 13 são ingleses e sete americanos. Existe pouca diferença entre os mais citados totais e os mais citados da Ciência Política: há mudança de posições, mas a diferença geral é de dois autores/as.

---

<sup>2</sup>Principalmente artigos de revistas, livros e capítulos de livros, mas também são encontradas fontes de dados (como a *Pew Research Center*) e instituições oficiais (como a *North Carolina State Board Of Elections*) entre as referências citadas.

<sup>3</sup>Chegamos nas citações únicas desconsiderando ano, título, periódico, doi e demais informações das referências bibliográficas, consideramos apenas a informação de autoria, normalizadas conforme descrito no Apêndice A

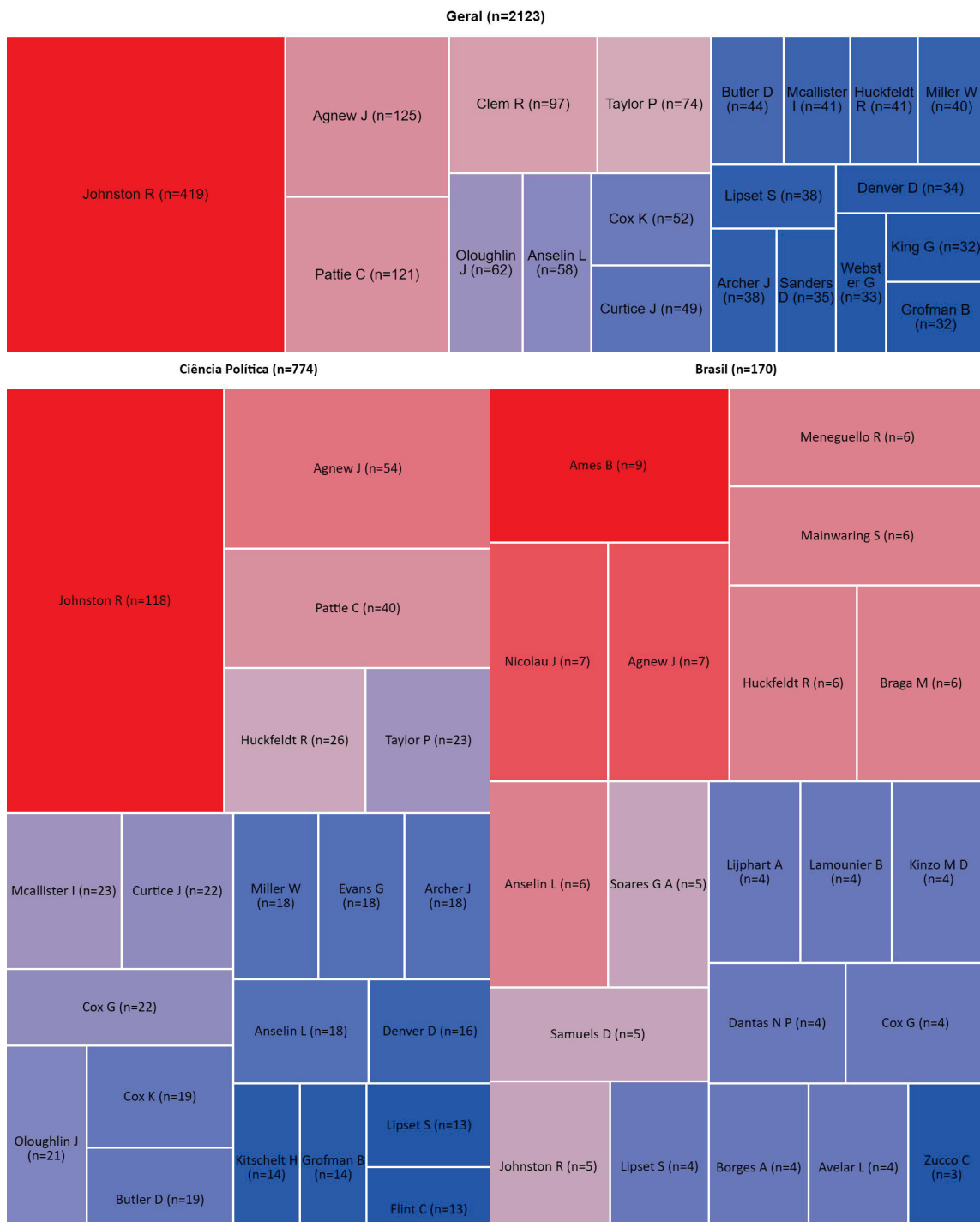


Figura 3.5: Mapa de árvore dos 20 autores mais citados no Geral, na Ciência Política e no Brasil

Fonte: O autor (2021)

A produção do Brasil apresenta um padrão diferente em dois sentidos. Primeiro, os autores/as americanos são 2,3 vezes mais presentes que os ingleses. Outra diferença é

a concentração de citações de autores/as brasileiros: 50% das citações são de autores/as brasileiros, 35% de autores/as americanos e 15% de ingleses. Esse resultado é um indicativo de pouco diálogo da GE realizada no Brasil com a internacional, sobretudo, com países não líderes da área, como os grupos de países emergentes, novos destaques e recém-chegados.

Outro resultado que mostra a dispersão da GE está relacionada aos periódicos das publicações. Os 226 artigos estão distribuídos em 107 periódicos diferentes, uma média de 2,1 artigo por revista, com desvio padrão de 2,69. Os dados mostram 65,4% dos periódicos com um caso e 22,4% contendo entre dois e três artigos. O principal periódico da GE é o *Political Geography*, localizado na Inglaterra, que tem um índice  $h^4$  de 92, pertencendo ao quartil de revistas mais bem avaliadas pelo *Scientific Journal Ranking* (SJR, 2021) em Geografia, Planejamento e Desenvolvimento, História, Sociologia e Ciência Política desde 1999. A revista *Eurasian Geography and Economics*, também de base na Inglaterra, possui um índice  $h$  de 41 e tem foco na região da Eurásia (China, Rússia, Índia, Sudeste Asiático, antiga União Soviética e Europa). Esse periódico pertence ao primeiro quartil em Geografia, Planejamento e Desenvolvimento, Economia e Econometria (idem) e é o segundo em concentração dos trabalhos da GE. *Post-Soviet Geography and Economics* é o terceiro lugar em número de publicações, esse periódico foi substituído pelo *Eurasian Geography and Economics* em 2002. A *Electoral Studies* concentra nove artigos, tem um índice  $h$  de 66, pertence ao primeiro quartil de revistas de Ciência Política e Relações Internacionais desde 2001 e é "*widely recognised as a major journal in the field [...] of elections and voting*" (ibidem). As próximas seções apresentam os resultados relacionais da produção da GE.

### 3.2.2 Análise Cientométrica

Na exploração cientométrica com redes bibliométricas aplicada aqui, os países, áreas de pesquisa, autores/as citados e palavras-chave, convertem-se em vértices conectados entre si por arestas. Isso permite a exploração de como a área se articula, ou de outra forma, como esses atributos bibliométricos estão conectados. O foco da exploração é identificar lacunas e tendências de pesquisa da área mostrando a posição do Brasil e da Ciência Política, por isso as unidades de observação agrupam os países e áreas de pesquisa. As redes representam a forma como os atores da área interagem e se agrupam.

---

<sup>4</sup>Medida de produtividade e impacto científico, mais sobre o índice em Thomaz et al. (2011).

Mais especificamente, elas formam estruturas que indicam as dinâmicas de proximidade entre diferentes áreas de pesquisa e países a partir de *co-citation* e *co-word*.

Os resultados estão divididos por método. Na primeira parte dos resultados, *co-citation*, os autores/as citados foram utilizados como *proxies* de uma relação temática e/ou teórica entre países e áreas. A ideia é que o fato de as pesquisas terem referências em comum é indicativo de uma relação intelectual significativa dentro de um campo (Small, 1973). O grau da inter-relação entre as áreas e países é determinada pelo número de citações idênticas entre *papers*. Na segunda, *co-word*, as conexões são estabelecidas a partir da associação de palavras-chave agregadas por país e área de pesquisa. Esses resultados podem evidenciar tendências e lacunas da área e servem para orientar os pesquisadores do campo (Feretti et al., 2018).

### 3.2.2.1 *Co-citation*

O primeiro grafo, Figura 3.6, mostra a relação entre os países e as áreas de pesquisa. O objetivo é identificar tendências e entender o lugar que o Brasil ocupa na rede, verificando quais relações são construídas pela GE no geral e no caso brasileiro. No grafo, o Brasil se encontra no agrupamento de cor roxa, à direita na extremidade superior.

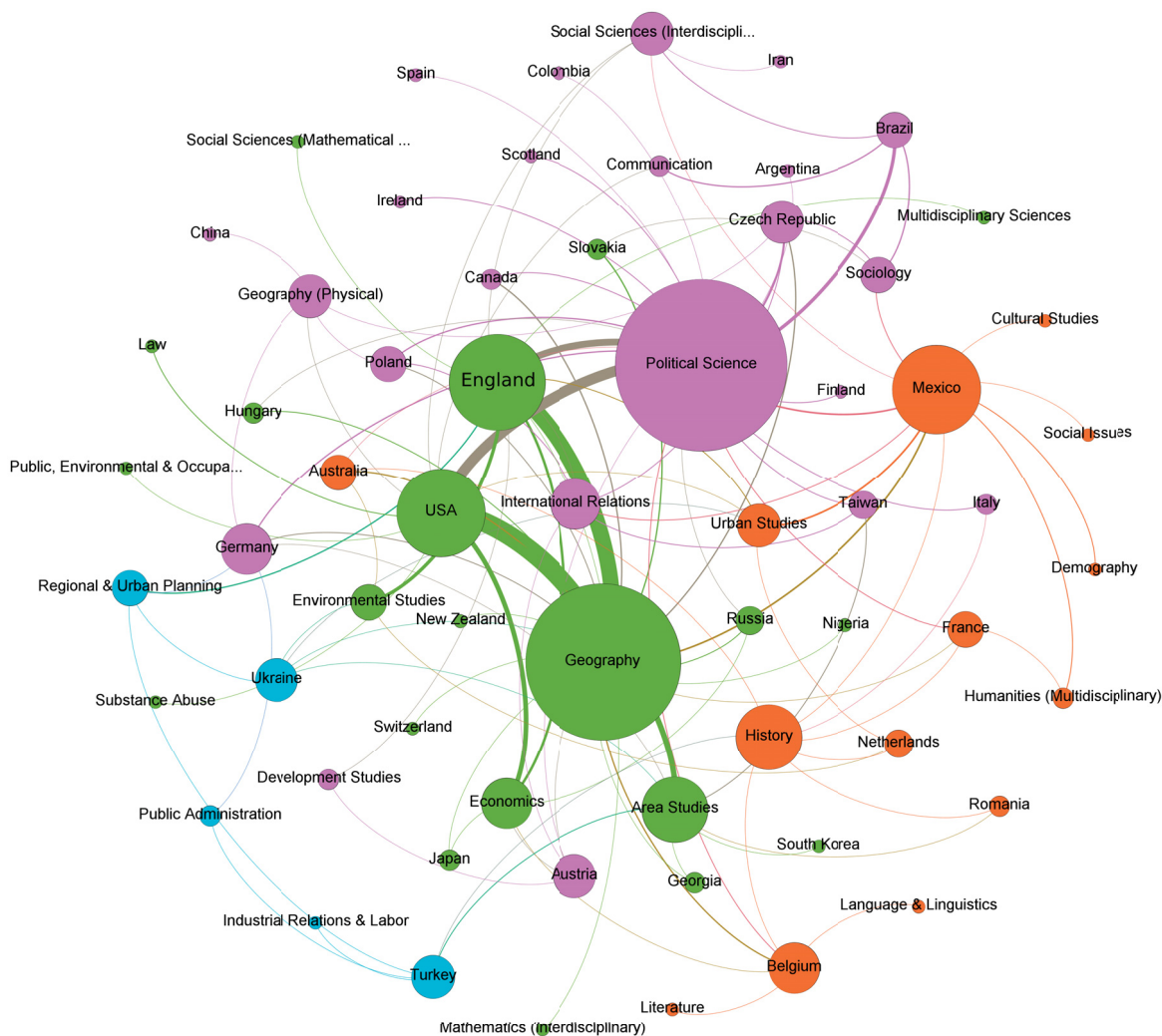


Figura 3.6: Grafo das relações entre países e áreas de pesquisa

Fonte: O autor (2021)

Nota: grafo bimodal; modularidade = 0,254; nós = 63; arestas = 117; distribuição da rede por modelo *Fruchterman Reingold*.

As conexões entre países e áreas de pesquisa formam sete agrupamentos. A Ciência Política é o ator central dessa rede com conexões com 22 países e 87 laços totais. Essa área se conecta mais fortemente com EUA, Inglaterra e Brasil (25, 16 e oito ligações respectivamente) e faz parte do agrupamento roxo, que agrega 36,5% da rede. Uma característica desse grupo é a presença de países com poucos laços: Irlanda, Espanha,



Escócia, Finlândia, Itália, Argentina e Colômbia apresentam produções exclusivamente na Ciência Política. Ao total, esse grupo conta com 16 países e sete áreas do conhecimento.

A Geografia é a segunda área em centralidade, apresentando laços com 20 países que totalizam 118 ligações totais muito concentradas nos EUA (46) e na Inglaterra (37). É a área central do grupo de cor verde que possui 11 países e 10 áreas de pesquisa. O destaque aqui é a maior pluralidade de disciplinas ao seu redor, principalmente das Ciências Exatas, e a presença de países do leste europeu e Ásia.

O México é o país com a maior interdisciplinaridade na produção de GE, e é o centro do grupo laranja. Esse agrupamento conta com 14 nós, seis países e oito áreas de pesquisa. Destaque para a maior presença de “humanidades”, como Literatura, Linguística, Problemas Sociais e História.

Por fim, Ucrânia e Turquia formam um agrupamento próprio, o azul, com a característica “prática” da GE, publicando nas áreas de Planejamento Regional e Urbano, Administração Pública e Relações Industriais e Trabalhistas.

O próximo grafo, Figura 3.7, é uma representação unimodal<sup>5</sup> das relações dos países construídas a partir dos autores/as citados. A figura indica agrupamentos de países que citam um conjunto de literatura em comum. As relações dos países, a partir da literatura compartilhada entre eles, formam seis agrupamentos principais e deixam cinco países isolados. No grafo, o Brasil se encontra isolado, com cor cinza, à esquerda abaixo do grupo rosa.

---

<sup>5</sup>Mais informações sobre a transformação bimodal para unimodal realizada pode ser encontrada em: <<https://github.com/jaroslav-kuchar/Multimode-Networks>>.

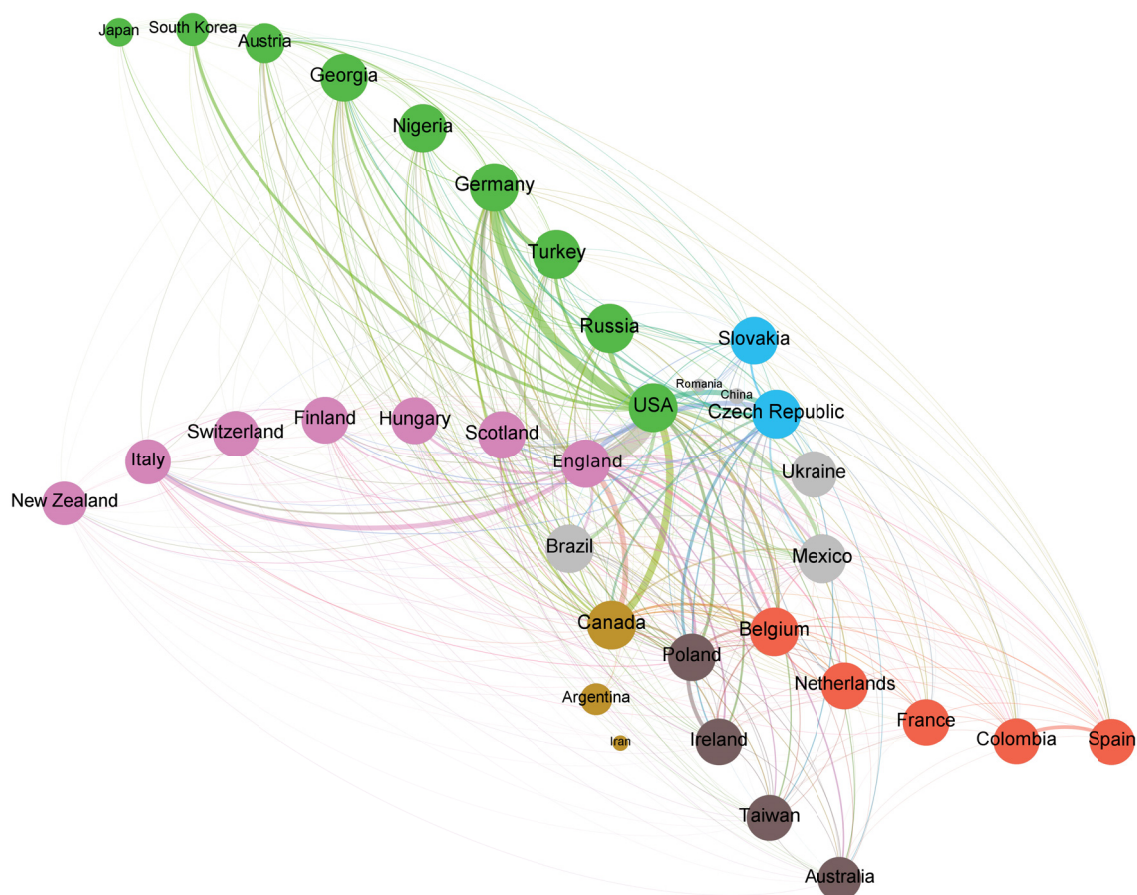


Figura 3.7: Grafo unimodal das relações entre países a partir dos autores citados

Fonte: O autor (2021)

Nota: grafo unimodal de pesos estabelecidos pelas relações dos países a partir de autores/as citados; nós = 35; arestas = 442; modularidade = 0,368.

O Brasil é um dos países isolados, o que indica pouco diálogo da nossa produção da GE com a do cenário internacional. Entretanto, diferentemente da Romênia e da China, que também se encontram isolados na rede, mas apresentam um número ínfimo de relações com a comunidade internacional (nenhuma relação para o primeiro e duas para o segundo), o Brasil apresenta 290 citações compartilhadas. EUA, Inglaterra, Alemanha, Canadá, República Tcheca e México são os países que mais compartilham literatura com os artigos brasileiros, com 44, 33, 21, 18, 18 e 17 relações, respectivamente. México e

Ucrânia também aparecem isolados: o primeiro apresenta 317 relações (51 com EUA, 35 com Inglaterra, 25 com República Tcheca, 21 com Canadá e 17 com o Brasil); o segundo soma 119 conexões (23 com EUA, 15 com Alemanha e 13 com Inglaterra).

O principal agrupamento, o verde, conta com nove países, sendo os EUA o principal deles. No total, esse grupo apresenta 524 conexões internas, com destaque para as relações entre EUA e Alemanha (n=109), Turquia e Alemanha (n=66) e EUA e Rússia (n=64). O segundo *cluster*, de cor rosa, mantém 204 conexões internas entre sete países, sendo a Inglaterra o principal deles. As principais relações nesse grupo são entre a Inglaterra e a Itália (n=61), Hungria (n=33) e Escócia (n=31). Na sequência, o grupo laranja, liderado pela Bélgica, totaliza cinco países e conta com 113 relações internas, sendo a principal entre Colômbia e Espanha (n=45). O próximo *cluster* é o marrom, com a Polônia como nó central de uma rede de quatro países. A principal relação do grupo é entre Irlanda e Polônia (n=48) e o total de relações internas é de 83. Canadá, Argentina e Irã formam um grupo fraco, de quatro relações internas totais. Por fim, República Tcheca e Eslováquia dividem 26 citações conjuntas e formam o grupo azul.

O próximo grafo é uma representação unimodal das relações das áreas de pesquisa construídas a partir dos autores/as citados. As relações a partir da literatura compartilhada entre elas formam quatro agrupamentos. O principal *cluster* é o da Ciência Política e o da Geografia, seguido da interdisciplinaridade não usual, com a Matemática e Saúde, em terceiro está o grupo de GE de humanidades e o menor, em número de áreas, é o grupo da GE *prática*.

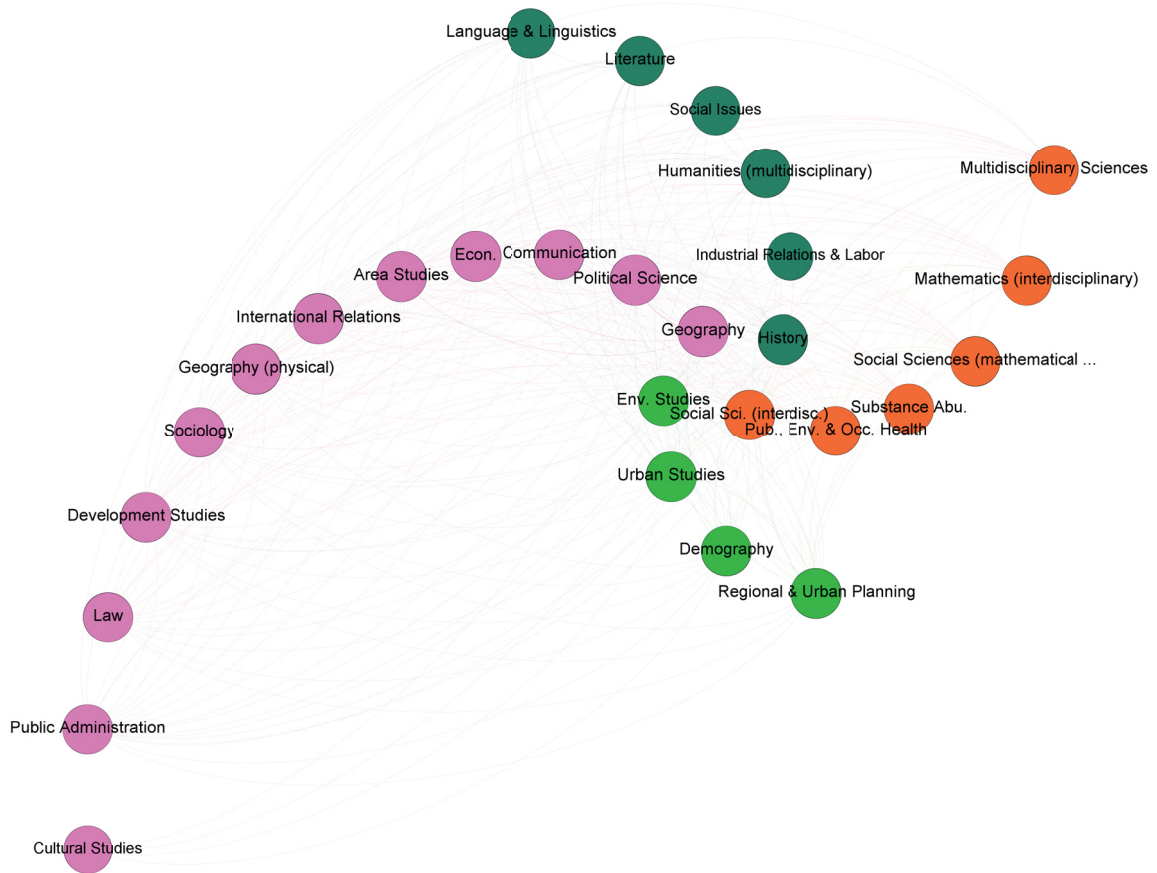


Figura 3.8: Grafo unimodal das relações entre áreas de pesquisa a partir dos autores citados

Fonte: O autor (2021)

Nota: grafo unimodal de pesos estabelecidos pelas relações dos países a partir de autores/as citados; nós = 35; arestas = 442; modularidade = 0,368.

O grupo rosa representa a corrente principal da GE, com o total de 3.816 conexões internas e tendo a Geografia como o principal ator do grupo. Os principais laços são da Geografia com a Ciência Política ( $n=863$ ), com os Estudos de Área ( $n=391$ ) e com a Economia ( $n=367$ ). Na sequência, vem a Ciência Política, que divide citações principalmente com Relações Internacionais ( $n=196$ ), Estudos de Área ( $n=195$ ), Economia ( $n=184$ ) e Sociologia ( $n=184$ ).

O segundo *cluster* é o laranja, da GE de interdisciplinaridade não usual, somando 160 conexões e tendo as Ciências Sociais (interdisciplinar) como principal nó. As relações de maior peso nesse grupo são entre Ciências Sociais (interdisciplinar) e Ciências Multidisciplinares (n=74), Saúde Pública, Ambiental e Ocupacional e Abuso de Substâncias (n=31) e Ciências Sociais (métodos matemáticos) e Matemática (interdisciplinar).

O terceiro agrupamento é o da GE das “humanidades”, esse grupo soma 154 conexões internas. A História aparece como principal área, fazendo 25 conexões com a Literatura e Linguagem e Linguística e 22 conexões com Problemas Sociais e Humanidades (multidisciplinar).

Por fim, os dados mostram um grupo de quatro áreas de características de aplicações práticas: Estudos Urbanos, Estudos Ambientais, Planejamento Urbano e Regional e Demografia. No total, o grupo soma 354 relações internas a partir da literatura compartilhada.

### 3.2.2.2 *Co-word*

Na análise de *co-word* realizada aqui, são exploradas as conexões estabelecidas pelo compartilhamento de palavras-chave na produção da GE. Começando pelas relações apenas entre as palavras-chave, os dados formam 37 agrupamentos totais, em que oito deles são os principais, com ao menos 5% dos termos.

O maior *cluster* é o de cor rosa e soma 14,72% das palavras-chave. O principal nó do grupo é “*Elections*” e conecta termos de característica de exploração do comportamento eleitoral como “*Voting Pattern*”, “*Voting Behavior*”, “*Voter Choice*”, “*Friends and Neighbours*”, “*Behavioral Pattern*”. É um resultado esperado que o centro da rede seja “eleições”. Os estudos de comportamento eleitoral representam o maior número de trabalhos Leib e Quinton (2011), a geografia associada ao voto, seja por fatores composicionais ou contextuais, é a linha mais importante da área (Agnew, 1990). O segundo maior nó do grupo é “*Political Parties*” que se conecta com pequenos atores como “*Latin America*” e “*Populism*”. Nesse grupo aparecem técnicas como “*Statistical Modeling*”, “*Discriminant Analysis*”, “*Gravity Models*” e “*Entropy-maximizing*”. Analisar a geografia da performance, organização e mobilização de partidos políticos é outro dos grandes objetivos da GE (Agnew, 1990).

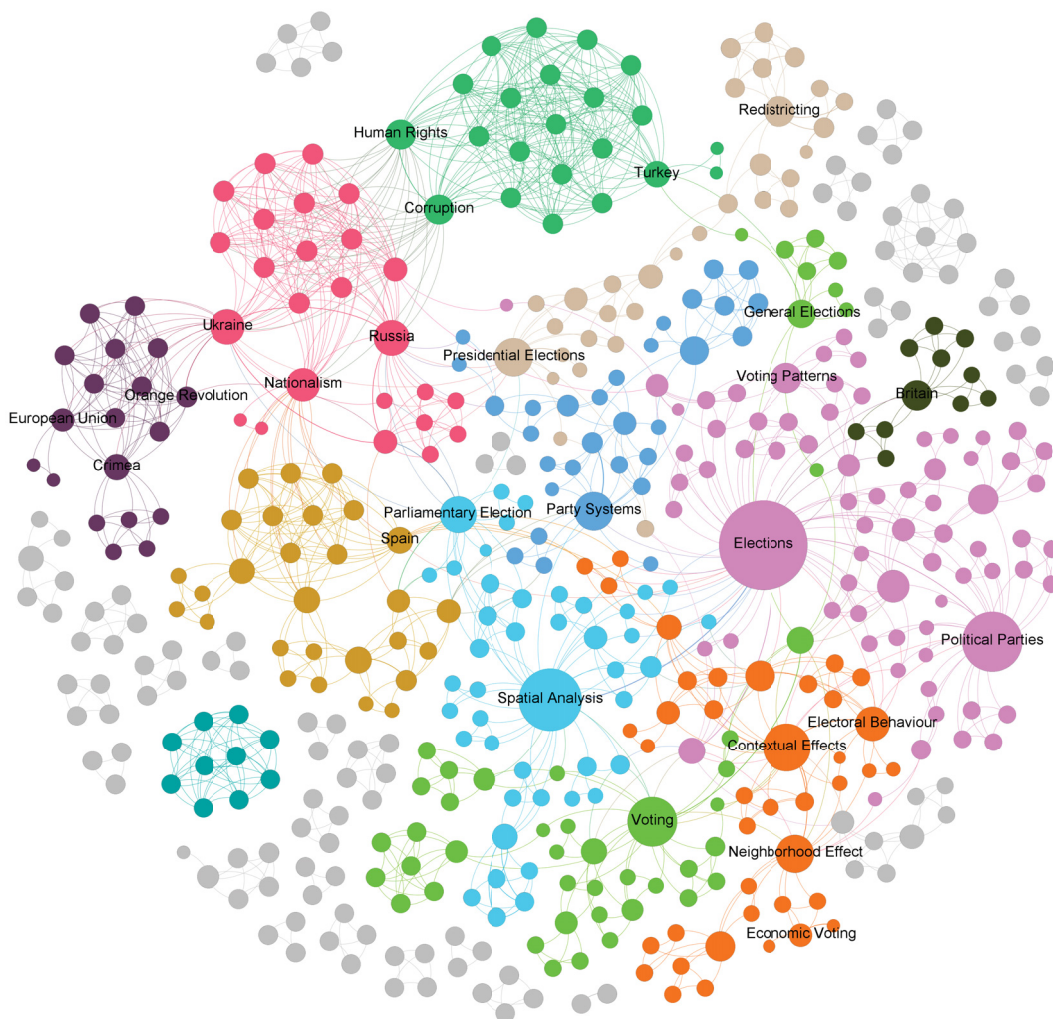


Figura 3.9: Grafo das relações entre palavras-chaves

Fonte: O autor (2021)

Nota: Grafo bimodal total; modularidade = 0,822; nós = 489; arestas = 1503. Em destaque palavras com maior grau de intermediação

O *cluster* verde claro soma 9,41% dos termos, o que o coloca na segunda posição. O principal termo aqui é “*Voting*”, mas, diferentemente do primeiro grupo, aparece conectado com questões sociais (“*Gay Rights*”, “*Social Issues*”, “*Class*”, “*Anti-immigration*”) e com movimentos políticos de direita (“*Religious Right*”, “*American Conservatism*”, “*Right-wing*” e “*Donald Trump*”). As técnicas de exploração presentes

nos termos deste grupo são “*Spatial Clustering*” e “*Cartography*”. Esse grupo mostra uma vertente de trabalhos sobre voto que incluem discussões de identidade e minorias.

O próximo grupo é o azul claro, com 9% dos termos, tendo como principal nó o método “*Spatial Analysis*”. Além desse, estão presentes termos como “*Multinomial Logit*”, “*Geospatial Technologies*” e “*Geographically Weighted Regression*”, o que caracteriza o grupo por palavras relacionadas a métodos e as técnicas. Outros termos presentes neste conjunto de artigos são “*Turkish Election*” e “*Czech Republic*”.

Com 8,18% dos termos, o grupo laranja também apresenta palavras ligadas a métodos e técnicas: “*Spatial Statistics*”, “*Spatial Models*”, “*Spatial-Structural Models*” e “*Regression Analysis*”. O principal nó desse grupo é “*Contextual Effects*” e estão presentes “*Germany*”, “*Belgium*” e “*Canada*” entre as principais palavras-chave. O grupo marrom claro tem como principais nós os termos “*Presidential Elections*” e “*Redistricting*”, conectando termos relacionado a explorações de “salamandragem” dos distritos eleitorais americanos: “*Red and Blue America*”, “*Partisan Bias*”, “*Partisan Asymmetry*”, “*Gerrymandering*” e “*Electoral Bias*”. Com 6,54% dos casos, esse grupo também apresenta termos de explorações brasileiras como “Lula” e “Bolsa Família”, conectados por Eleições Presidenciais. As técnicas presentes são “*Spatial Econometrics*” e “*Qualitative Analysis*”. Uma das principais críticas ao desenvolvimento da GE é a de *obsessão por metodologia quantitativa* (Agnew, 1990). Os resultados das comunidades temáticas parecem evidenciar essa *obsessão*. Identificadores de técnicas e métodos estão presentes em grande variedade e destaque nos principais agrupamentos.

O grupo azul escuro soma 6,34% dos casos e tem como característica explorações de sistemas representadas nos principais nós — “*Party System*” e “*Electoral System*”. Estão presentes os países “*Japan*” e “*Brazil*” entre os termos, o primeiro conectado com “*Rural Areas*” e “*Agricultural Policy*”, e o segundo conectado com “*Electoral connection*” e “*Federalism*”. Em amarelo escuro estão agrupados os países “*United Kingdom*”, “*France*”, “*Spain*” e “*Hungary*”, e os temas “*Terrorism*”, “*Organized Crime*” e “*Far Right*”. No total o grupo agrega 5,73% dos termos e conta com “*Gaussian Models*” e “*Classification*” como técnicas de pesquisa. O último grupo, com 5,11% dos casos, é o vermelho. Nele estão presentes os países “*Russia*”, “*Ukraine*” e “*Poland*” com os temas “*Nationalism*” e “*Censorship*”.

O próximo grafo relaciona os países produtores da GE e as palavras-chaves dos artigos. Essas conexões formam 19 *clusters*, nove países aparecem isolados, sete em dupla, dois em trio e um quarteto. Os artigos dos EUA formam o maior *cluster*, o laranja, representando 23,85% dos casos. O resultado mostra uma grande diversidade dos termos

usados para identificar os trabalhos de GE do país. As palavras-chave desse grupo são de diversos tipos: as de explorações internas (*Oregon, North Carolina, Louisiana, U.S. Politics*); externas (*Ukraine, Middle East, Sri Lanka, Yemen, India, Israel*); de explorações de minorias e problemas sociais (*Same-sex Marriage, Ethnicity, Gay Right, Sexual Politics, Education, Civil Liberties, Unemployment*).

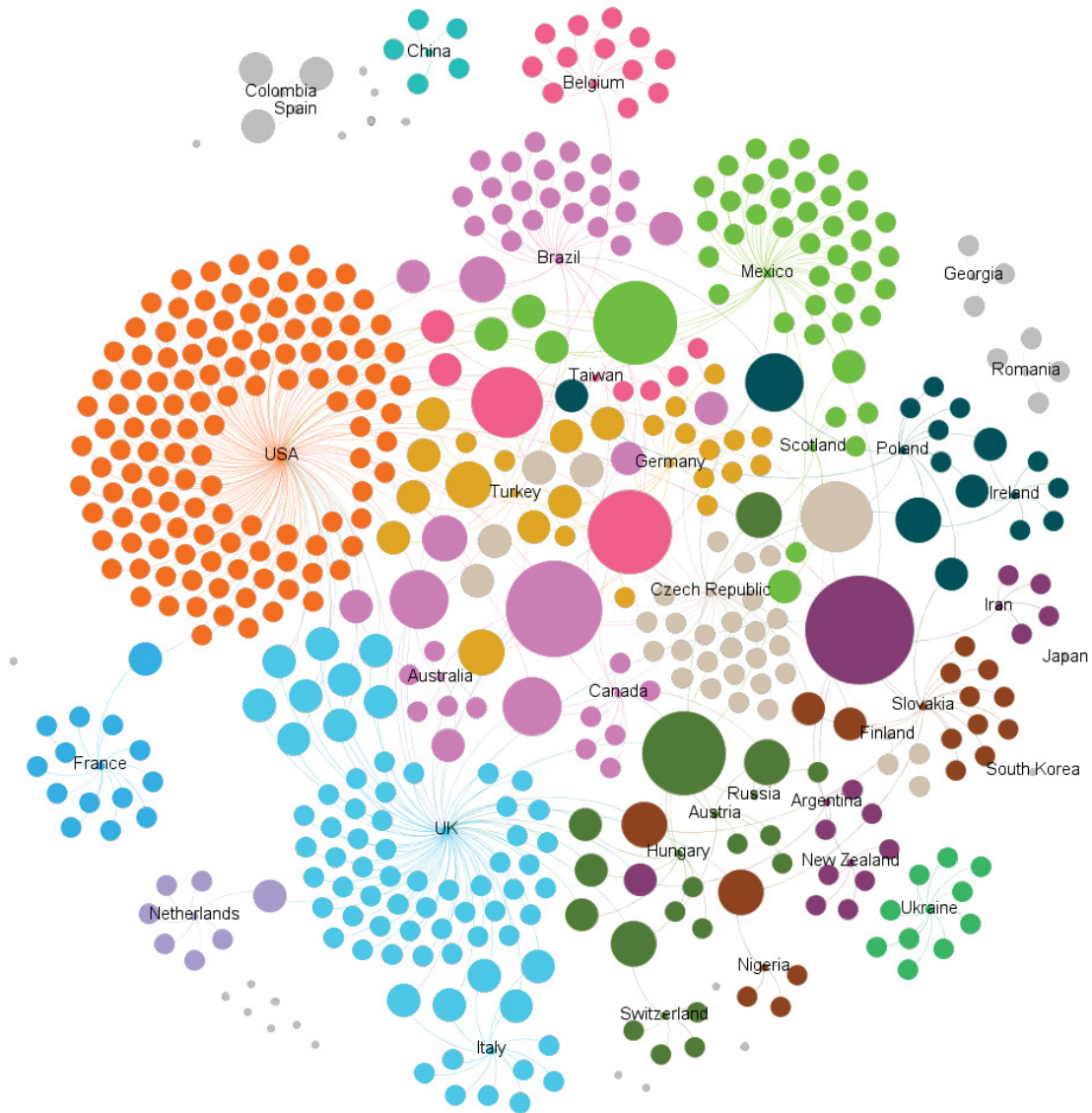


Figura 3.10: Grafo das relações entre países e palavras-chaves

Fonte: O autor (2021)

Nota: Grafo bimodal total; modularidade = 0,734; nós = 524; arestas = 588. Errata: Onde está "UK" ler "England".

Errata: Onde está "UK" ler "England".



Diferente do agrupamento dos EUA, os outros países, que formam grupos isolados, não apresentam essa diversidade de palavras-chave. Coreia do Sul e Japão não mostram palavras-chaves e os outros países isolados exibem como principais conexões: Georgia com “*Societal Cleavages*”; Romênia com “*Parliamentarianism*”, “*Austria-hungary National Movement*” e “*Political Orientations*”; China com “*Geo-strategy*”, “*World Geography*”, “*Geopolitics*” e “*Research Progress*”; Países Baixos com “*Gentrification*”, “*Anti-immigration*”, “*Suburbanisation*” e “*Class*”; Ucrânia com “*Crimea*”, “*Euromaidan Revolution*”; e França com “*Decolonization*”, “*Indigenous People*”, “*Socio-spatial Inequalities*”.

O principal agrupamento, com dois países, é o da Inglaterra e Itália, formando o *cluster* azul claro com 13,74% dos casos. Conectam-se a partir dos termos “*Civil War*”, “*European Politics*”, “*Political Legacies*” e “*Italy*”. O destaque das conexões exclusivas da Inglaterra são termos de técnicas estatísticas (“*Gaussian Models*”, “*Logistic Regression*”, “*Gravity Models*” e “*Entropy-maximizing*”) e da Itália “*Political Subcultures*”.

O próximo grupo de dois países é o verde claro, do qual fazem parte México e Escócia, que somam 8,97% dos termos e estão conectados principalmente por “*Spatial Analysis*” e “*Turnout*”. O grupo se destaca pela presença de técnicas e estatísticas entre as palavras-chave. O México apresenta uma grande variedade de palavras-chave únicas, o que mostra a diversidade da produção.

Com 6,49% dos casos, o grupo marrom claro agrega os termos mais próximos dos artigos da República Tcheca e da Finlândia que são conectados, principalmente, por “*Parliamentary Election*”. A República Tcheca se destaca pela diversidade de palavras-chave únicas, tratando de “*Nationalism*”, “*Extremism*” e “*Borders*”.

Do grupo amarelo escuro fazem parte Turquia e Alemanha, que estão conectados por termos de exploração da política turca (“*Turkish Party System*”, “*Akp*” e “*Turkish Politics*”) e “*General Elections*”.

O grupo vermelho tem a Bélgica e o Taiwan conectados por “*Neighborhood Effect*”. No grupo marrom escuro estão Eslováquia e Nigéria, conectados por “*Voting Patterns*”. No grupo petróleo estão Polônia e Irlanda, que dividem “*Voter Choice*”, “*Friends And Neighbours*”, “*Local Campaigning*” e “*Electoral Behaviour*”. Colômbia e Espanha dividem um trabalho, por isso estão conectados por “*Peace Agreement*”, “*Colombian Armed Conflict*” e “*Farc-ep*”.

Dois grupos são formados por três países: do rosa fazem parte Brasil, Canadá e Austrália, que se conectam principalmente por “*Political Parties*”; do vinho fazem parte Irã, Argentina e Nova Zelândia, conectados principalmente por “*Elections*” e “*Peronism*”.

Um grupo, o verde escuro, é formado por quatro países: Rússia, Áustria, Hungria e Suíça, conectados por “*Party System*”, “*Populism*” e “*Nationalization*”.

Os termos mais compartilhados com os artigos do Brasil são “*Spatial Econometrics*”, “*Legislative Elections*” e “*Sub-national Politics*”. E os específicos fazem referências às explorações do Programa Bolsa-Família como “Lula”.

O próximo grafo apresenta as relações entre áreas de pesquisa e palavras-chave, as quais formam nove agrupamentos principais: dois com uma área de pesquisa, três com duas, um com três, um com cinco e um com seis.

O maior agrupamento é o formado pela Geografia e pelos Estudos de Área, que somam 39,07% dos nós da rede. O destaque desse grupo é a presença de uma grande diversidade de temas, sobretudo temas de minorias e desigualdade. *Cluster* da Geografia mostra mais palavras relacionadas a debates ditos de “minorias” (“*Black*”, “*Gay*”, “*African*”, “*Social*”).

O segundo agrupamento é o rosa, que soma 25,53% dos nós e agrega Ciência Política, Estudos Culturais e Relações Internacionais. Esse grupo se conecta com objetos mais tradicionais da Política relacionados a questões de democracia, como sistemas partidários e competição eleitoral.

O grupo azul claro soma 10,64% da rede e é o terceiro maior, agrupando seis áreas: Demografia, Estudos Urbanos, Economia, Estudos de Desenvolvimento e Estudos Ambientais. Esse grupo conta com termos de técnicas e estatísticas de tratamento de dados espaciais, sendo as principais “*Spatial Analysis*” e “*Spatial Statistics*”. Outro conjunto de termos estão relacionados a comportamento eleitoral.

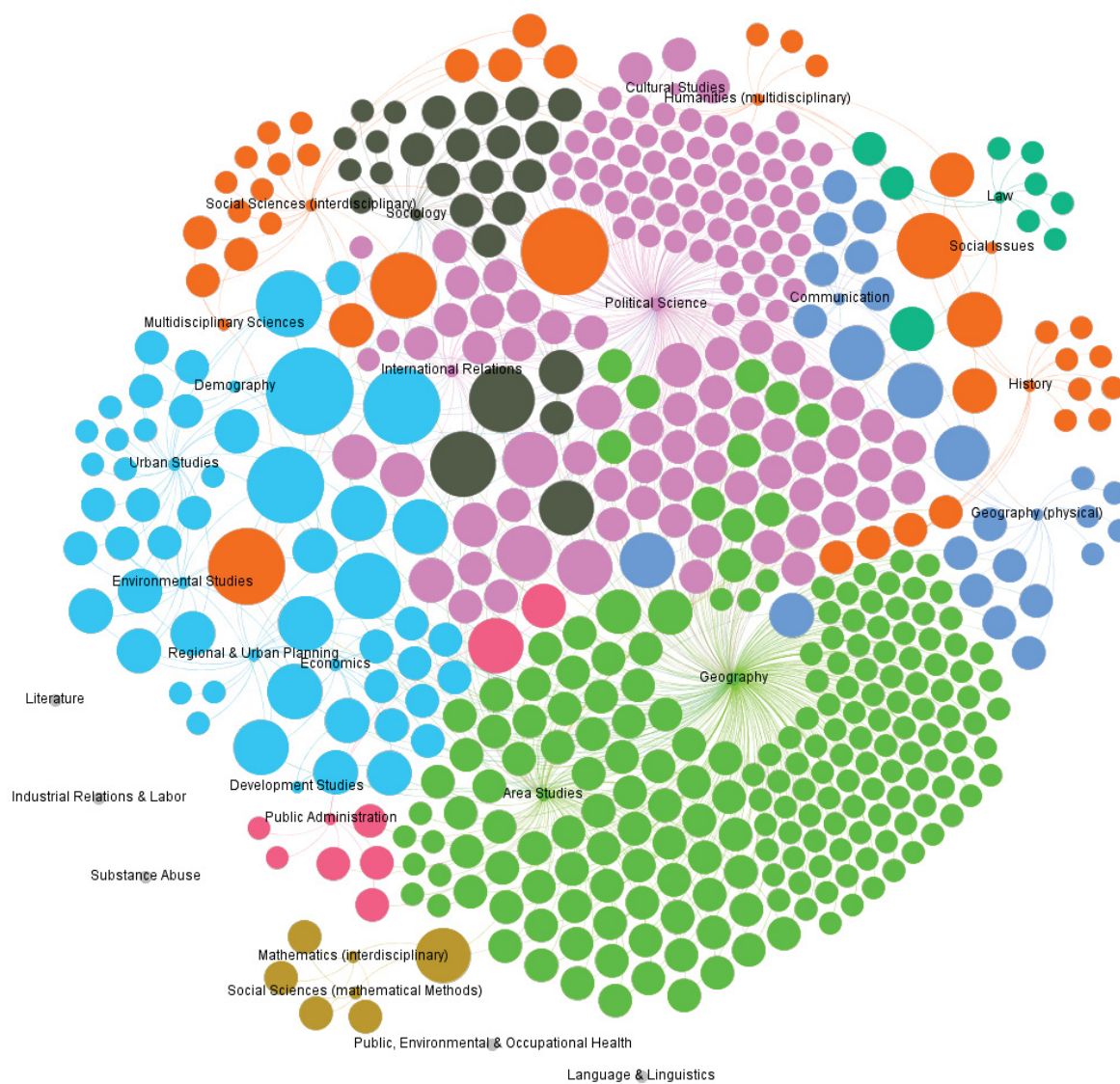


Figura 3.11: Grafo das relações entre áreas de pesquisa e palavras-chaves

Fonte: O autor (2021)

Nota: Grafo bimodal total; modularidade = 0,513; nós = 517; arestas = 844. tamanho dado por grau de entrada.

O grupo laranja soma 8,51% dos nós e conta com cinco áreas: História, Problemas Sociais, Humanidades (multidisciplinar), Ciências Sociais (Interdisciplinar) e Ciências Multidisciplinares. É um agrupamento difuso, contando com termos de explorações eleitorais (“*Elections*”, “*Voting Pattern*”, “*Polarization*”, “*Political parties*” e “*Representation*”) e de técnica (“*Entropy-maximizing*”).

A Sociologia forma um agrupamento isolado com 5,22% dos nós. Destaque para o uso dos termos “*Friends and Neighbours*”, “*Neighborhood Effect*” e “*Contextual Effects*”.

Comunicação e Geografia (física) formam um agrupamento com 5,03% dos nós. Estão conectados por “*Political Behavior*” e “*Voting*”. Comunicação também agrega os termos característicos de explorações do Brasil (“Lula” e “*Electoral Territory*”).

Entre as áreas incomuns, o Direito e a Administração Pública formam *clusters* isolados de menos de 2% dos nós cada. O primeiro se conecta com questões de viés eleitoral (“*Gerrymandering*”). Já o segundo se concentra ao redor de palavras-chave da política turca. Matemática (interdisciplinar) e Ciências Sociais (métodos matemáticos) formam o último agrupamento com ao menos 1% dos nós. As palavras-chave desse grupo são técnicas (como “*Classification*” e “*Gaussian Models*”) e temáticas (como “*Parliamentary Constituencies*” e “*Electronic Petitions*”).

### 3.3 LACUNAS

Para essa seção utilizo todos os dados analisados para sustentar indicações de lacunas e oportunidades para o desenvolvimento da agenda de pesquisa da GE na CPBR. Minha proposta é usarmos essas informações como guia de correção de rota, ou mapa de construção de pontes para o desenvolvimento da área. Os resultados indicam alguns desalinhamentos entre a produção brasileira e a internacional. A principal implicação desses desalinhamentos é a produção da GE do Brasil aparecer isolada no acoplamento bibliográfico, sem formar *ambiente intelectual* com outros países. Abaixo discorro sobre dois caminhos, ou agendas de pesquisa, que podem fechar essa lacuna e nos aproximar dos diálogos atuais da área.

#### 3.3.1 Uma Geografia Eleitoral mais *Geográfica*

A Geografia é a área com maior número de publicações da GE, tanto nos resultados obtidos aqui quanto por Leib e Quinton (2011). É a área mais presente na produção dos EUA, país líder da GE. É o principal ator em número de conexões por acoplamento bibliográfico e em *co-word*, além de ser altamente conectado com *keywords* de método. É destaque pela pluralidade de disciplinas ao seu redor, principalmente das ciências exatas. Mas, na amostra analisada aqui, o Brasil não apresenta nenhum artigo nessa área.

Essa não é uma lacuna nova e não é um problema exclusivo da produção da GE brasileira, Agnew (1990, p.20) já indicava que "*the challenge facing the field today [1990] is to become more geographical*". Essa crítica surge de um movimento de virada para perspectivas pós-estruturalistas, feministas, pós-colonialistas, teorias *queer* e de questões de identidade na Geografia nos anos 80 e 90 que não estava sendo acompanhado pela GE (Leib e Quinton, 2011).

Ao considerar as publicações em periódicos da Geografia entre os anos 2000 a 2010 a lacuna do uso de ideias pós-estruturalistas e de identidade, ou de aproximação com a Geografia Humana, foi superada pela GE no cenário internacional:

*Race, gender, and sexuality were all brought into focus as contestations of spatialized identities as played out within electoral processes. The diversity in these cleavages of power and their usage to produce spaces and places leaves little doubt that post-structuralisms have influenced the current shape of electoral geography.* (Leib e Quinton, 2011, p.19).

Contudo, talvez porque haja pouco diálogo da CPBR com abordagens pós-estruturalistas, essas temáticas estão ausentes na amostra da nossa produção da GE. Esse resultado contrasta com a produção geral da GE no nível internacional e.g.: o maior agrupamento entre áreas de pesquisa e palavras-chave é o formado pela Geografia e termos relacionados a minorias e desigualdade como "*Black*" (Grofman e Handley, 1989), "*Gay*", "*African*" (Clem, 2011), "*Social*". Os EUA, país líder da área, se destaca pela presença de termos como *Same-sex Marriage*, *Gay Right*, *Sexual Politics* (McGing, 2014); A Rússia apresenta pesquisas ao redor da temática *Ethnicity* (White, 2015), e ainda temos os exemplos de "*Indigenous People/(de)colonization* (Blay, 1988)" explorados na França.

Se pensarmos em termos de fases de desenvolvimento da área, o estágio atual é mais próximo da *geografia* (Begu, 2007). Ou de uma visão onde o estudo da dimensão espacial da política não está limitado à questões das instituições formais do Estado, englobando fenômenos dados por origem de gênero, classe, religiosa e de identidade (Ethington e McDaniel, 2007, p. 127). E a nossa produção não se encontra nessa fase de desenvolvimento.

### 3.3.2 Novos Fenômenos Espaciais

Na amostra de artigos brasileiros também estão ausentes explorações de novos fenômenos espaciais. Esse pode ser o segundo caminho de desenvolvimento para a GE

brasileira. Identifiquei duas temáticas para os quais faltam destaque na nossa produção: gentrificação e movimentos políticos de direita.

A Holanda é o destaque entre os novos atores da área com um artigo com alto número de citações com tema “*Gentrification*”. Os estudos urbanos jogam luz sobre mudanças na estrutura demográfica, social, no estilo de vida, valores e padrões de consumo associados aos espaços. O Estado tem papel em uma gama de políticas que impactam a vida urbana nesses novos arranjos espaciais, desde de a construção e distribuição de habitações a manutenção de infraestruturas, como escolas, transporte, segurança, lazer e cultura (Wacquant, 2010, p. 56). Mas, assim como as temáticas de *minorias*, esses fenômenos de gentrificação não constam na amostra.

*Right-wing, Religious Right, Conservatism, Far Right* são termos que aparecem no segundo maior agrupamento de *co-words*, mas não apresentam nenhuma ligação com a produção brasileira. O estudo dos fenômenos relacionados ao crescimento da direita e extrema-direita é uma das tendências mais atuais da área. Representam alguns dos artigos mais recentes da amostra (Mancosu e Ladini, 2020; Bertus, 2017). Na produção brasileira o trabalho de Faganello (2017), que explora bases eleitorais de representantes ligados aos aparatos militares e policiais, é um exemplo da tendência. Mas essa é uma pesquisa de dissertação de mestrado, logo, não é indexada por ferramentas como a WoS e tem baixa visibilidade.

### 3.4 CONSIDERAÇÕES FINAIS

Este capítulo foi motivado pelo objetivo de orientar os pesquisadores da área em direção ao desenvolvimento da GE brasileira. Resumi o problema na seguinte pergunta:

*Pergunta de pesquisa 2:* Quais os principais desalinhamentos entre a produção da Geografia Eleitoral brasileira e as tendências internacionais?

Para embasar minha resposta analisei atributos bibliométricos e características relacionais da produção da área. Os dados indicam a seguinte resposta:

*Resposta:* O maior desalinhamento é a ausência de conexão com a Geografia Humana e, mais especificamente, com as pesquisas com abordagens pós-estruturalistas e de identidade. Um desalinhamento secundário é o de explorações de novos fenômenos

espaciais como a gentrificação e as novas direitas e extremas-direitas.

## 4 96 MIL OBJETOS ESPACIAIS RELACIONADOS POR ROTAS TRANSITÁVEIS

No Capítulo 2 aponte argumentos para a exploração de padrões espaciais de voto no Brasil a partir da escala pontual de locais e votação e com relações de distância entre os objetos. No Capítulo 3 os resultados indicam mais *geografia* para o desenvolvimento da GE no Brasil. Neste capítulo combino esses dois resultados para desenvolver uma operacionalização da análise espacial dos resultados eleitorais brasileiros que lide com as limitações metodológicas e permita a exploração de um tópico da lacuna.

O capítulo está organizado em quatro partes: apresento o *rationale* do modelo proposto; os procedimentos realizados para a operacionalização do modelo P-DL; Testes de concordância entre A-Q1 e P-DL; e por fim, ilustro a potencialidade da proposta com a análise de um caso de território eleitoral indígena.

### 4.1 RATIONALE

#### 4.1.1 Uma Operacionalização Original

O objetivo que guiou o trabalho foi encontrar um método avançado para representar e analisar a realidade espaço-eleitoral brasileira. Para isso, testei as opções mais recorrentes na literatura, e desenvolvi um modelo original baseado em distâncias medidas por rotas terrestres e com peso variável pela densidade local. Um modelo que considera evidências da superfície modelada, suavização, variações e irregularidades ao longo território e a finalidade da modelagem — como recomendado por Smith et al. (2015).

A minha proposta têm três avanços principais na conceituação de relações espaciais: i) explora uma unidade da federação inteira usando a unidade espacial de locais de votação; ii) modela o peso do espaço na relação entre as unidades considerando a distância “real” entre elas, a partir de rotas terrestres de todo o estado; e iii) aplica uma função local de decaimento para o peso das relações entre unidades espaciais proporcional à distância e a concentração de LV.

Comparo a *operacionalização padrão*, o modelo de área com contiguidade de primeira ordem (A-Q1, a partir daqui) com um modelo pontual com métrica de distância local, robusta à variação de densidade/heterogeneidade espacial da distribuição dos eleitores (P-DL, a partir daqui). Para a primeira parte do teste do modelo, a concordância



entre métodos, utilizo três operacionalizações dos dados eleitorais brasileiros. A-Q1 e A2-Q1 são as operacionalizações praticadas pela literatura e P-DL a proposta desenvolvida aqui. Resumo dos modelos:

- A-Q1: Voto agregado em município e relações de contiguidade de primeira ordem.
- A2-Q1: Voto agregado em local de votação decomposto por método de Decomposição Voronoi e relações de contiguidade de primeira ordem.
- P-DL: Voto agregado em local de votação e relações de distância por rotas transitáveis com métrica de ajuste local de variação de densidade.

As Figuras 4.1, 4.2 e 4.3, ilustram as diferenças entre os modelos (em vermelho a seleção de um objeto espacial e em azul a forma como a relação é estabelecida).

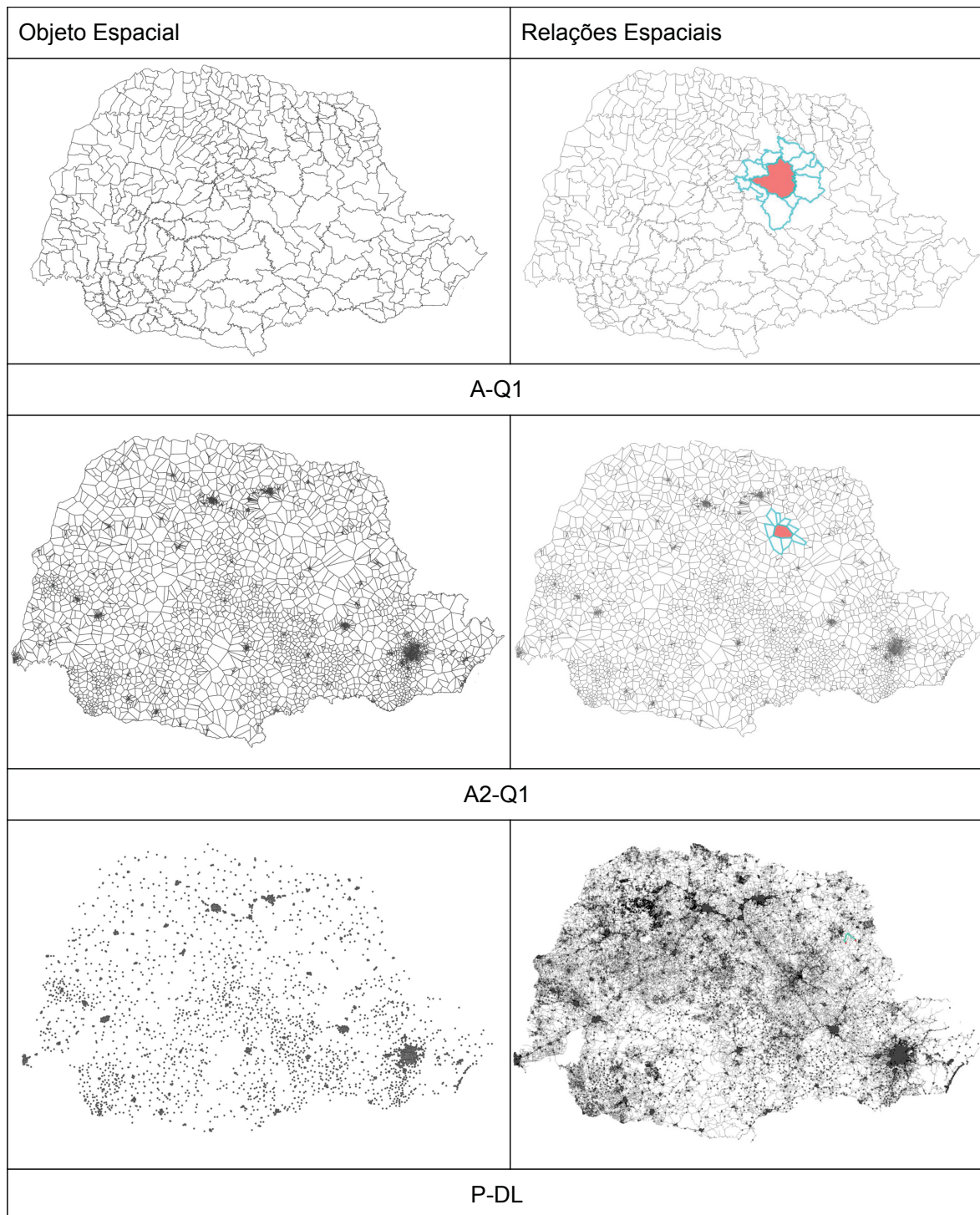


Figura 4.1: Esquemática dos métodos

Fonte: O autor (2021)

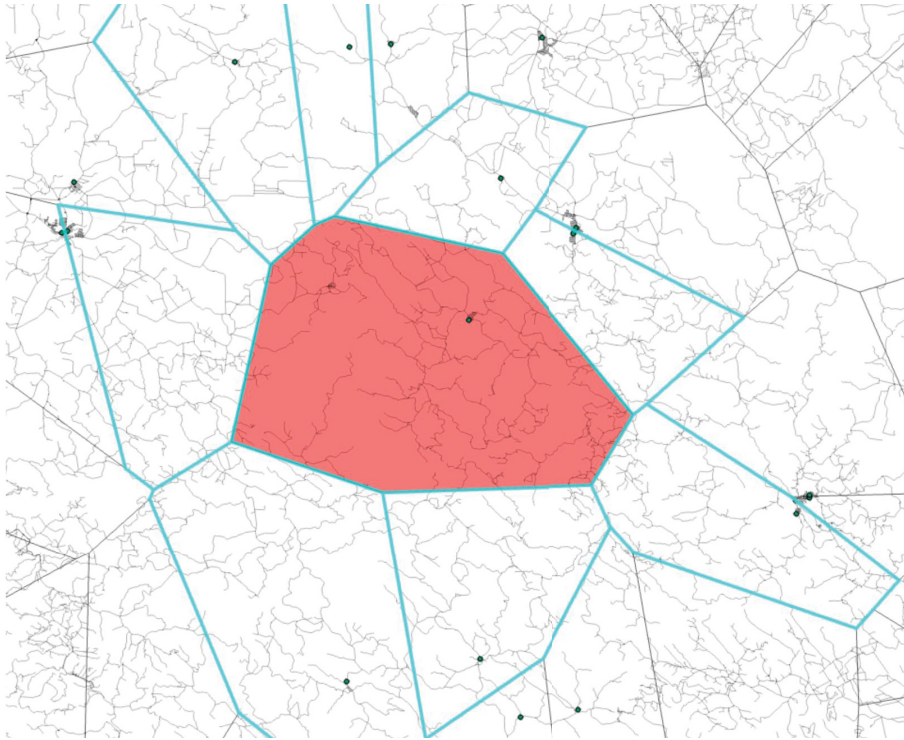


Figura 4.2: Detalhe das Relações espaciais A2-Q1  
Fonte: O autor (2021)

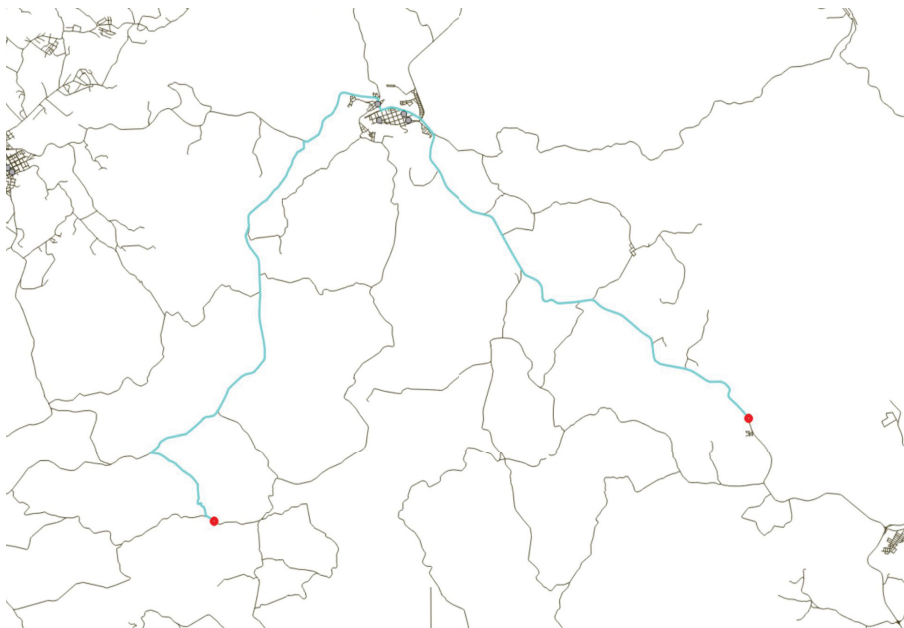


Figura 4.3: Detalhe das Relações espaciais P-DL  
O autor (2021)

#### 4.1.2 Um Caso Original

O principal resultado da exploração de lacunas e limitações da GE no Brasil indica a aproximação com a Geografia Humana e, mais especificamente, com as pesquisas com abordagens pós-estruturalistas e de identidade. Aqui faço isso a partir da exploração da representação eleitoral indígena.

Na CPBR questões relativas aos povos indígenas praticamente não são estudadas. Apenas dois artigos tratam do tema entre 2.621 publicações de sete principais revistas da área no Brasil (Soares, 2018, p. 1). Os principais entraves relatados são de cunho metodológico, em função da “insignificância numérica” do grupo para a Ciência Política (Codato et al., 2017) e de falta de sistematização dos dados (Soares, 2018). Para contornar esses problemas e, com isso, ilustrar o potencial da proposta, aplico o modelo P-DL para explorar a configuração espacial de votos de uma campanha indígena.

A próxima seção apresenta os métodos e procedimentos do desenvolvimento do modelo.

### 4.2 PROCEDIMENTOS: O DESENVOLVIMENTO DO MODELO

#### 4.2.1 Bases de Dados

Este trabalho faz uso de três bases de dados principais. A primeira delas, com resultados eleitorais agregados por Locais de Votação (LV), obtida por meio do Repositório de Dados Eleitorais do Tribunal Superior Eleitoral (TSE). A segunda, com as coordenadas geográficas dos LV, também com o TSE como fonte principal e Hidalgo (2021) e Pindograma (2020) como fontes auxiliares. Por fim, uma base com os pesos das relações espaciais entre as unidades de análise, criada a partir do conjunto de rede de ruas do Brasil do OpenStreetMap (OSM).

##### 4.2.1.1 Definição de Amostra

O estudo foi conduzido com dados do Estado do Paraná. O caso do Paraná interessa por questões práticas<sup>1</sup>. As decisões metodológicas de análise espacial dependem da experiência do analista com o fenômeno em estudo (Terron, 2009). Explorei o mesmo Estado em pesquisas anteriores (Gonçalves, 2016), o que garante maior controle sobre as

---

<sup>1</sup>Isso não significa que o modelo final não seja generalizável, embora seja difícil considerar um caso típico com as tamanhas diferenças regionais do Brasil, a operacionalização lida com problemas compartilhados por todos os estados, como é o caso da variação territorial de densidade populacional e concentração de locais de votação.

particularidades do caso. Como o objetivo principal é o desenvolvimento de um método novo, conhecer o caso em profundidade assegura maior controle da superfície modelada.

Os dados considerados são os das eleições gerais de 2018 para deputados federais e estaduais. Somando os dois cargos, foram 1.183 candidaturas para esse pleito, sendo 433 para deputado federal e 750 para estadual. A maior parte dos candidatos não apresenta nenhuma representação efetiva ou condições mínimas de competição — como mostram as estatísticas de distribuição de voto na Tabela 4.2.

Tabela 4.1: Estatísticas descritivas de candidaturas e de votação para os cargos de Deputado Federal e Estadual, 2018, Paraná

Estatísticas	Dep. Federal	Dep. Estadual
Miníma	12	7
Máxima	314.963	427.749
Média	12.326	7.293
1º Quartil	507	315
Mediana	1.631	958
3º Quartil	7.936	5.135

Fonte: O autor (2021), com base no Repositório de Dados Eleitorais do TSE (2021).

Tomei como estratégia analisar o quartil superior de candidatos na quantidade total de votos nominais. No total, analiso uma amostra de 297 candidatos que receberam entre 4.623 (Paulo Demchuk, PROS) e 427.749 votos (Fernando Francischini, PSL) para a disputa Estadual e 6.789 (Sidnei Trevisan, PSD) e 314.963 votos (Gilson Fatur, PSD) para o âmbito Federal. O quartil de candidatos mais votados soma mais de 90% dos votos nominais (91,19% para o pleito Federal e 90,21% para Estadual), o que indica que essa é uma amostra suficientemente representativa do eleitorado. Do ponto de vista prático, esse corte ainda garante um número mínimo de votos para a distribuição nas mais de quatro mil unidades de análise. Do ponto de vista temático, a amostra é formada pelos candidatos que apresentam um grau mínimo de competitividade e estrutura de campanha.

#### 4.2.1.2 *Objetos Espaciais — Geocodificação de Locais de Votação*

Segundo os dados oficiais do TSE, a eleição de 2018 no Paraná contou com um total de 4.723 locais de votação, para um eleitorado de 7.971.087 pessoas. A variação no número de eleitores entre LV foi de 90,10%, com um mínimo de dois eleitores (Centro

Municipal de Educação Infantil Esther Gonçalves Josepetti, Maringá) e máximo de 12.437 (Colégio Anglo Americano, Foz do Iguaçu), com média de 1.626 e mediana de 1.257.

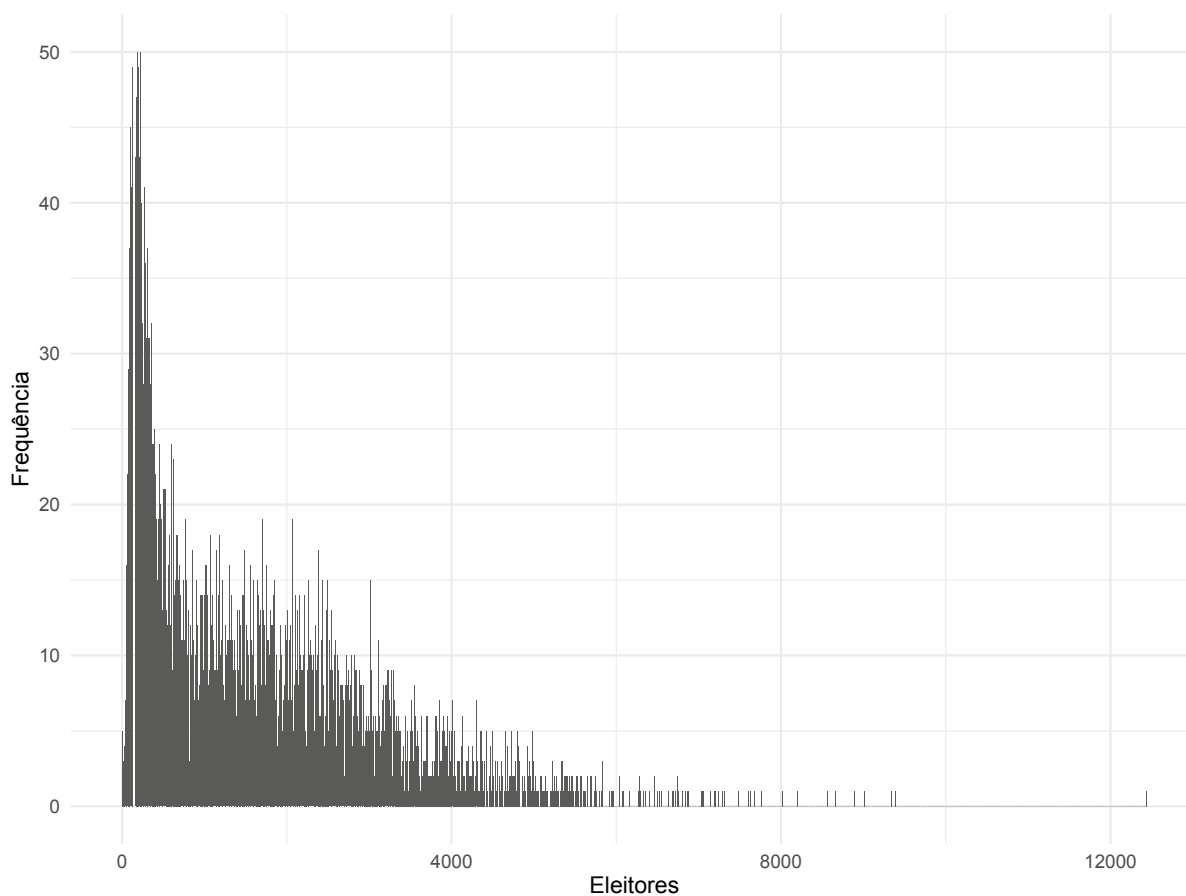


Figura 4.4: Distribuição de eleitores por locais de votação, 2018, Paraná

Fonte: O autor (2021), com base no Repositório de Dados Eleitorais do TSE (2021).

Em 2018, aproximadamente 90% do eleitorado do Paraná passou por recadastramento biométrico. Isso significa que os dados desse pleito são particularmente bons, já que a maior parte do eleitorado teve seu local de votação atualizado com base em comprovantes de endereço recentes. Em outras palavras, os locais de votação nessa disputa representam melhor onde o eleitor mora.

Um dos maiores limitantes para o uso dos dados eleitorais na escala de LV é a falta de dados oficiais de suas coordenadas geográficas. O TSE disponibiliza o endereço e o nome do local de todos os LV. Em linhas gerais, os endereços contam com cidade, bairro, CEP, logradouro, número, além do nome da escola, universidade, unidade de saúde ou outro espaço usado como local de votação. Para endereços centrais nos aglomerados urbanos os sistemas de geocodificação, como a *API* do *Google Maps*, entregam as

coordenadas exatas referentes aos endereços pesquisados. O problema é a performance de *geocoding*<sup>2</sup> em cidades do interior e, sobretudo, em zonas rurais<sup>3</sup>.

Felizmente, o TSE tem disponibilizado essas informações cada vez mais completas e existem trabalhos em desenvolvimento com soluções inovadoras para o problema. Para a construção da base de localização geográfica dos LV do Paraná, utilizei duas dessas soluções. A primeira é a abordagem desenvolvida por Hidalgo (2021)<sup>4</sup> que se baseia em um modelo de *machine learning*. O modelo foi treinado com as coordenadas oficiais do TSE e realiza *string matching* entre endereços nas bases de Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE, Censo 2010 e Censo Agrícola 2017) e no Catálogo de Escolas do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). O segundo método foi desenvolvido por Pindograma (2020)<sup>5</sup> e também utiliza buscas nas bases do CNEFE e INEP, além de Aglomerados Rurais Isolados (BC250) e requisições por *API Google Maps* e *Google Places*. Com isso, foi possível obter mais de 98% do total de coordenadas.

A base de coordenadas de LV que uso aqui tem 4.660 entradas válidas de um total de 4.723. A primeira etapa da construção da base teve a seguinte composição: 3999 localizações foram definidas pelas coordenadas geográficas do banco de dados oficial do TSE (Eleitorado 2018, por seção eleitoral); 559 são coordenadas preditas pelo modelo de Hidalgo (2021)<sup>6</sup>; 150 pertencem a base de dados de Pindograma (2020); 15 coordenadas foram encontradas a partir de coleta manual no Google Maps. Para chegar nesse número, apliquei dois métodos de validação automática para as coordenadas<sup>7</sup>. Primeiro chequei a correspondência entre os limites de fronteira do município sede e a localização geográfica do LV. No segundo teste, busquei duplicatas, LV com identificações de endereço diferentes, mas com coordenadas geográficas iguais. As coordenadas ausentes

---

<sup>2</sup>É comum encontrar trabalhos que se referem à tarefa de determinar as coordenadas geográfica de um ponto como georreferenciamento. Mas se tratam de tarefas diferentes. Diferentemente da geocodificação, o georreferenciamento serve para determinar um sistema de coordenadas geográficas para um vetor ou raster.

<sup>3</sup>E.g., para o LV de número 1066 na cidade de Cruz Machado, com endereço "Estrada Principal, Zona Rural" e nome "Centro Comunitário Rio da Areia", nenhum dos métodos de geocodificação foi capaz de determinar as coordenadas.

<sup>4</sup>Detalhes completos sobre o desenvolvimento e testes do modelo podem ser encontrados aqui: [https://raw.githubusercontent.com/fdhidalgo/geocode\\_br\\_polling\\_stations/master/doc/geocoding\\_procedure.html](https://raw.githubusercontent.com/fdhidalgo/geocode_br_polling_stations/master/doc/geocoding_procedure.html)

<sup>5</sup>Detalhes completos sobre o desenvolvimento podem ser encontrados aqui: <https://github.com/ipeaGIT/geobr/issues/184>

<sup>6</sup>345 apresentam erro menor que 2 km e 134 menor que 6 km, além de 180 locais exclusivos dessa base.

<sup>7</sup>Na conclusão discuto sobre como aumentar a confiabilidade dessa base de coordenadas a partir de testes mais robustos de validação.

(N=63) representam 1,33% do total de LV e, por serem os endereços mais isolados geograficamente, representam aproximadamente 0,01% dos votos dos candidatos da amostra.

#### 4.2.2 *W* — Rotas Transitáveis e Peso Local

Essa seção apresenta o processo de definição da matriz de conexão *W* do modelo construído nesta tese. O objetivo do trabalho foi encontrar um método avançado para representar e analisar a realidade espaço-eleitoral usando como exemplo o caso paranaense. Para isso, testei as opções mais recorrentes na literatura, e desenvolvi um modelo original baseado em distâncias medidas por rotas transitáveis (terrestres) e com peso variável pela densidade local.

O Paraná é um estado que agrega 399 municípios, em uma área de 199.316 km<sup>2</sup> onde se distribuem 9.6 milhões de habitantes, segundo os dados do Censo do Instituto Brasileiro de Geografia e Estatística (IBGE, 2010). Essa distribuição não é uniforme pelo território, como mostram os dados da Tabela 4.2.

Tabela 4.2: Estatísticas descritivas da densidade demográfica do Paraná a partir dos municípios (hab/km<sup>2</sup>)

Propriedade	Valor
Número de unidades territoriais	399
Área total (km <sup>2</sup> )	199.316,4
População total	9.664.226
Densidade demográfica média	62.080
Densidade demográfica mínima	3.31
Densidade demográfica máxima	4.024
Desvio padrão da densidade demográfica	240.180
Coef. de variância da densidade demográfica	3.869

Fonte: Censo IBGE (2010). O autor (2021)

De um lado, o estado concentra aproximadamente 30% da população no aglomerado urbano da grande Curitiba (município que sozinho abriga cerca de 17% da população do Paraná, mais de quatro mil habitantes por km<sup>2</sup>). Também se destacam alguns municípios ao norte do Estado, como Londrina (732 hab/km<sup>2</sup>) e Maringá (306 hab/km<sup>2</sup>), e ao oeste, Foz do Iguaçu (414 hab/km<sup>2</sup>). Do outro lado, municípios como Alto Paraíso no sul e Guaraqueçaba, no litoral, somam menos de 4 habitantes por km<sup>2</sup>.



A distribuição dos locais de votação segue a tendência de alta variação, assim como a densidade demográfica dos municípios. As distâncias mais curtas entre locais de votação variam entre 275 metros e 29,30 km, com média de 2,22 km e desvio padrão de 2,74 km. A figura abaixo apresenta a distribuição dos LV no Paraná.

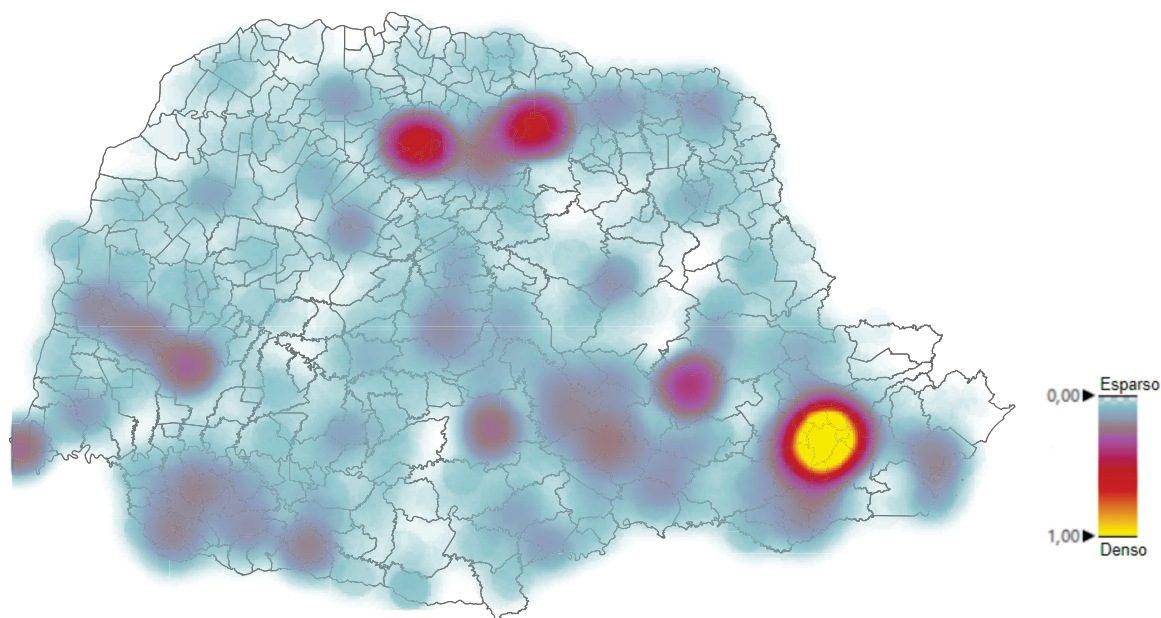


Figura 4.5: Mapa de densidade dos locais de votação, Paraná, 2018

Fonte: O autor (2021);

O histograma das menores distâncias em quilômetros dos locais de votação reflete essa distribuição e mostra dinâmicas claramente diferentes para o mesmo conjunto de dados. Um pico de locais com distâncias de até 1,1 km entre si, seguido de um vale de locais entre 2 km e 3,5 km, que sobe e depois desce até chegar em 10 km. Essa distribuição bimodal ainda apresenta grande amplitude nas distâncias com um conjunto de *outliers* espaciais (valores 3 desvios padrão acima da média) entre 10 km e 29 km de distância do vizinho mais próximo.

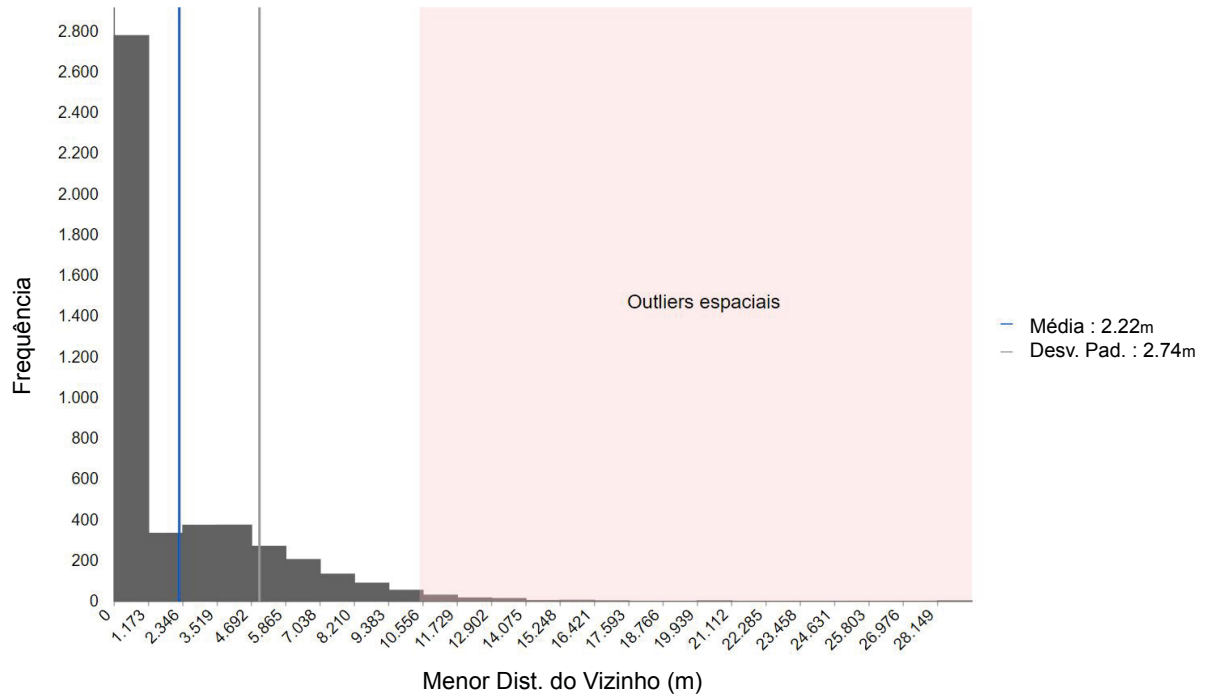


Figura 4.6: Histograma das menores distâncias entre vizinhos dos locais de votação do Paraná, 2018  
Fonte: O autor (2021).

Uma das principais soluções proposta pela literatura é assumir um *threshold* de distância mínima para que nenhum *OE* fique isolado. No caso dos LV do Paraná esse *threshold* é de 29,30 km. Considerando essa definição o número de conexões dos dados apresenta alta variação, entre 1 e 788, com média de 182, mediana de 68 e desvio padrão de 243,355. A distribuição do número de vizinhos, presente no histograma da Figura 4.6, mostra três dinâmicas de vizinhança no mesmo conjunto de dados.

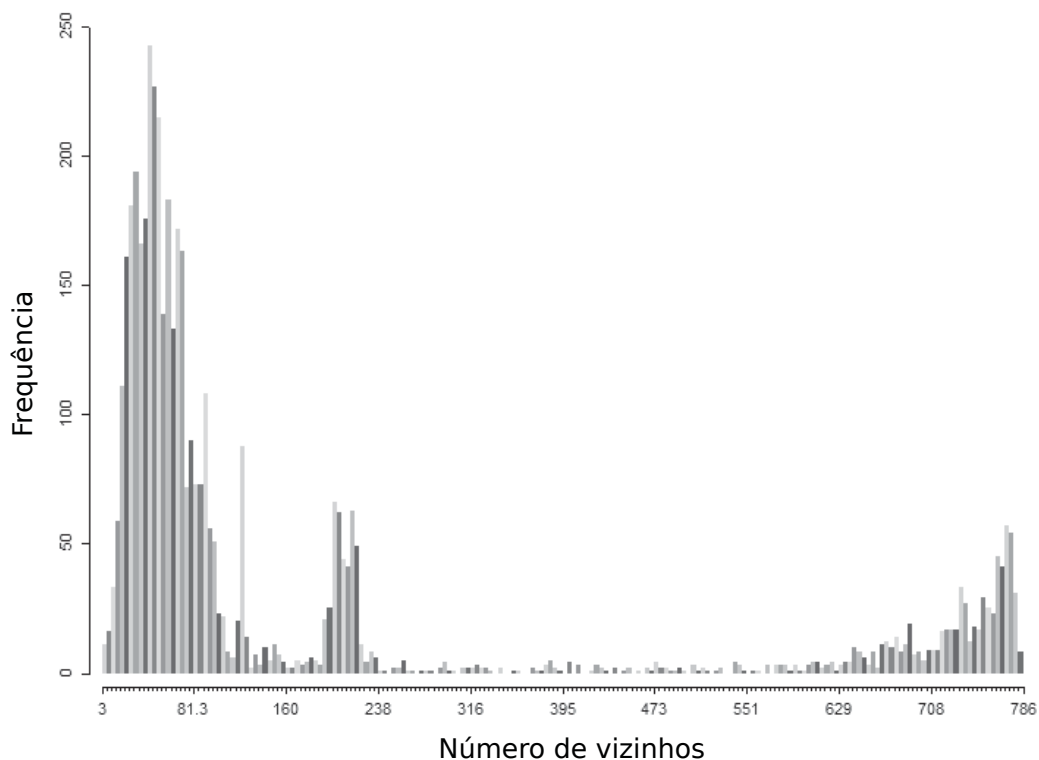


Figura 4.7: Histograma do número de vizinhos por local de votação com *threshold* de 29,30 km, Paraná, 2018

Fonte: O autor (2021).

A matriz com o *threshold* de 29,30 km gera uma densidade de conexões que não corresponde com a realidade. Não é provável que exista relações espaciais entre 68 locais de votação, como é a média para esse grafo, muito menos entre mais de 700, como acontece nas regiões mais populosas (em vermelho na Figura 4.8). Embora esse método crie mais relações espaciais do que é possível esperar na realidade empírica, do lado positivo, usar os dados de ponto para criar as relações fez com que não fossem criados laços em áreas inabitadas, como a área de preservação ambiental da Escarpa Devoniana na região dos Campos Gerais (em verde).

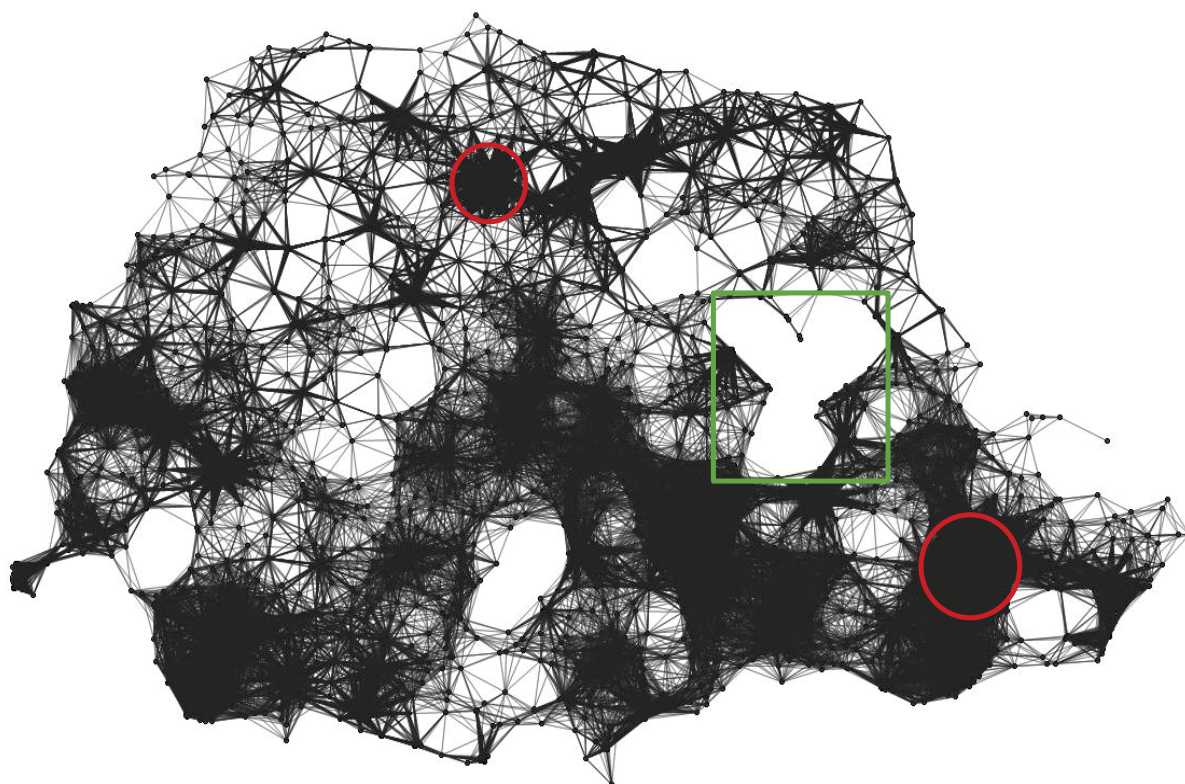


Figura 4.8: Mapa de relações entre Locais de Votação com threshold de 29,30 km, Paraná, 2018  
Fonte: O autor (2021).

Os trabalhos brasileiros usam majoritariamente uma modelagem de relação espacial de distância  $Q1$ . Usando o modelo  $Q1$  a distribuição do número de vizinhos é próxima da normal, com variação entre 1 e 20, média de 5,9 e mediana de 6, desvio padrão de 1,47. À primeira vista, a distribuição dos dados indica que essa matriz é mais adequada para tratar as relações espaciais entre as unidades. Entretanto, do ponto de vista conceitual, esses dados só falam da geometria das fronteiras da transformação de pontos em polígonos com métodos de decomposição, como o de Thiessen. Além disso, ao analisar o grafo de conexões espaciais em  $Q1$ , é possível identificar ruídos, como mostra o exemplo da Figura 4.10 do mapa de conexão entre os pontos após a decomposição.

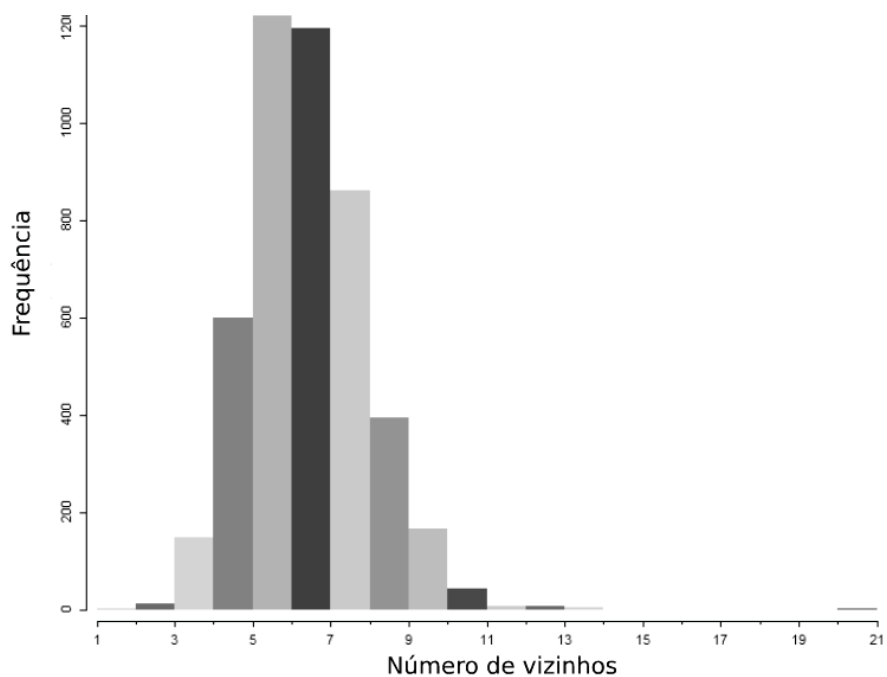


Figura 4.9: Histograma do número de vizinhos por local de votação com contiguidade de ordem 1, Paraná, 2018

Fonte: O autor (2021).

O método força relações entre as unidades sem nenhum sentido temático. Para esse método a mesma área de preservação ambiental citada acima é ignorada (nota em vermelho na Figura 4.10). O processo de decomposição de pontos no espaço ignora toda a presença de montanhas, rios e outros objetos espaciais que tornam a distância real entre pontos — e por conseguinte de eleitores — muito maior. Por isso, a minha proposta, buscando deixar o modelo mais próximo da realidade, considera a distância entre as unidades tomando as rotas terrestres de todo o estado para a definição de vizinhos. Ou seja, uma matriz que leva em conta a circulação *real* entre os LV — que se dá em uma estrutura de rede de ruas, pontes, estradas, viadutos.

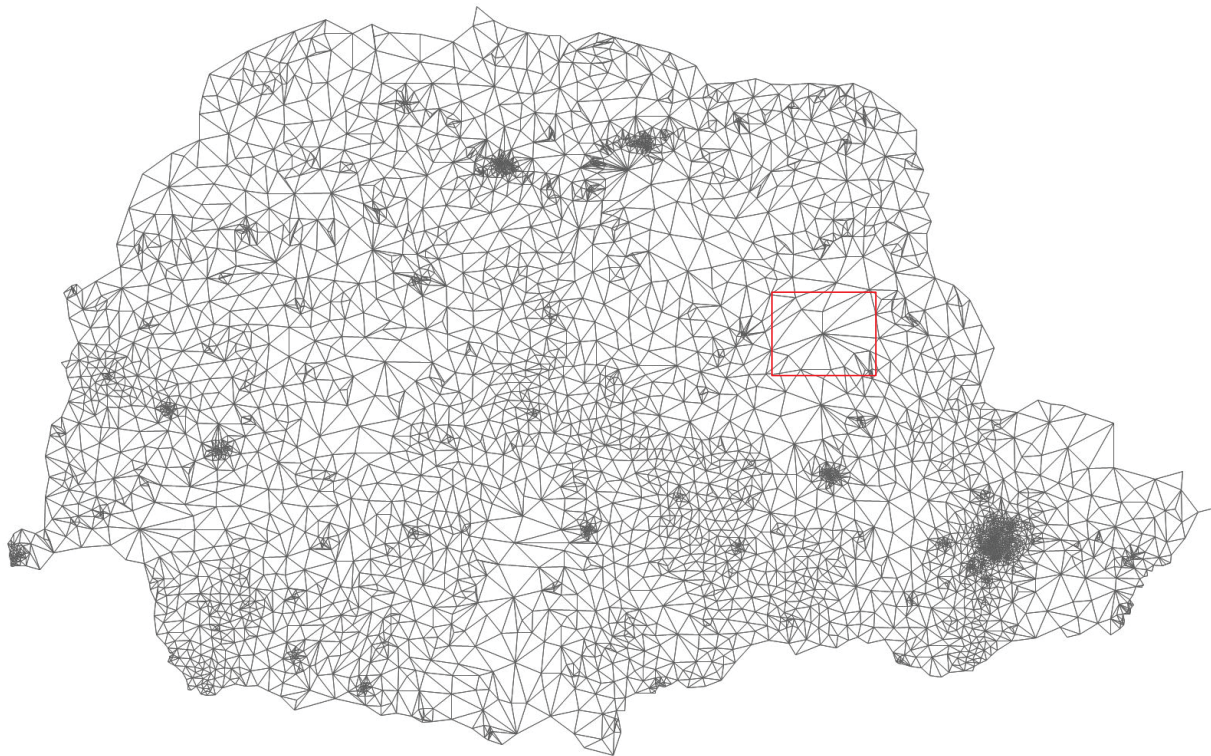


Figura 4.10: Distorções nas conexões entre unidades espaciais de polígonos de Thiessen e matriz de vizinhança Q1, Paraná, LV, 2018

Fonte: O autor (2021).

A matriz de pesos espaciais criada a partir do grafo de estradas permite definir os relacionamentos espaciais por distância, tempo de viagem e, a depender da qualidade da base de dados, até pelo custo da viagem por transporte público. Isso aumenta a liberdade de conceitualização das relações espaciais e permite explorações mais fiéis à realidade. Considerar as distâncias por rotas transitáveis entrega um ganho real para a análise das relações espaciais, sobretudo em áreas grandes e de relevos tão diversos como as do Brasil.



Figura 4.11: Exemplo de viés de distância em linha reta e de Decomposição Voronoi, Paraná, LV, 2018  
Fonte: O autor (2021).

Um exemplo é a relação entre os LV na região do Parque Nacional do Superagui (Figura 4.11). O destaque está na relação entre o posto de saúde de Vila Fátima (LV de número 1325) e a Escola Rural Municipal Padre Mário de Maria (LV 1341), ambas na cidade de Guaraqueçaba, litoral paranaense. Em linha reta a distância entre os dois pontos é de aproximadamente 17 km. Mas, a distância *real* entre esses dois locais é de 21 km em estrada e mais de 29 km de barco. Esse é um exemplo de diversos outros

dados pela geografia do Estado que são ignorados pelos métodos Q1 e os que consideram linhas retas entre os pontos.

Além dos métodos de contiguidade de primeira ordem e *threshold* de 29,30 km, testei diversas combinações para a definição de zona de influência do local de votação. Seguindo algumas opções da literatura foram testados: *threshold* de 15 km, 10 km, distância inversa e distância incremental com pico de  $Z$ . Entretanto, devido às diferentes dinâmicas de concentração de  $LV$ , nenhuma das abordagens globais geraram bons resultados. Considerar o mesmo peso para todas as unidades com uma grande variação de distribuição espacial sub-representa as relações em zonas rurais ou sobrerrepresenta as relações em zonas urbanas. Por isso, incluí no modelo P-DL uma função simples do fator de distância local de cada  $LV$  para pesar as relações espaciais.

A solução que proponho pondera as relações de vizinhança dentro da realidade de distâncias de cada localidade com uma função exponencial decrescente. Esse é um recurso matemático clássico e usual para modelar similaridades locais, variando de 0 a 1 de maneira controlada. Formalmente, incorporei ao modelo uma função exponencial negativa  $e^{-x}$  para chegar em uma matriz de pesos espaciais  $W_{i,j}$  com distâncias ponderadas pela média dos  $k$  vizinhos mais próximos  $t$  de cada objeto espacial  $(i, j)$ :

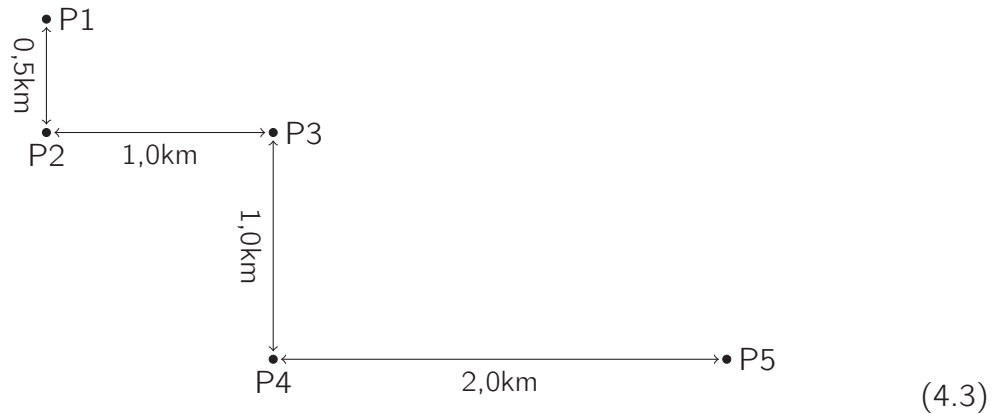
$$w_{i,j} = e^{\left(-\frac{d_{i,j}^2}{t_i^2 + t_j^2}\right)} = e^{\left(-\frac{d_{i,j}^2}{2\sigma^2}\right)} \quad (4.1)$$

$$\sigma^2 = \frac{t_i^2 + t_j^2}{2} \quad (4.2)$$

A Equação 4.3 apresenta um exemplo da aplicação da função para um universo de cinco unidades espaciais conectadas entre si por linhas bidirecionais que representam diferentes distâncias, entre 500 metros e 2 km. A matriz de distâncias par a par é representada por  $d_{P,P}$ , onde a menor distância é entre  $P1$  e  $P2$  (500 metros) e a maior entre  $P1$  e  $P5$  (4,5 km). O cálculo das distâncias locais  $t$  foi definido como a média das



duas<sup>8</sup> unidades espaciais mais próximas  $k=2$ .  $W_{DL}$  corresponde a matriz de distâncias depois da transformação considerando o fator local.



$$d_{P,P} = \begin{matrix} & \begin{matrix} P_1 & P_2 & P_3 & P_4 & P_5 \end{matrix} \\ \begin{matrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \end{matrix} & \begin{pmatrix} 0,0 & 0,5 & 1,5 & 2,5 & 4,5 \\ 0,5 & 0,0 & 1,0 & 2,0 & 4,0 \\ 1,5 & 1,0 & 0,0 & 1,0 & 3,0 \\ 2,5 & 2,0 & 1,0 & 0,0 & 2,0 \\ 4,5 & 4,0 & 3,0 & 2,0 & 0,0 \end{pmatrix} \end{matrix} \quad (4.4)$$

$$t_{k=2} = \begin{matrix} & \begin{matrix} P_1 & P_2 & P_3 & P_4 & P_5 \end{matrix} \\ \begin{matrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \end{matrix} & \begin{pmatrix} 1,0 & 0,75 & 1,0 & 1,5 & 2,5 \end{pmatrix} \end{matrix} \quad (4.5)$$

$$W_{DL} = \begin{matrix} & \begin{matrix} P_1 & P_2 & P_3 & P_4 & P_5 \end{matrix} \\ \begin{matrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \end{matrix} & \begin{pmatrix} 0,000 & 0,852 & 0,325 & 0,146 & 0,061 \\ 0,852 & 0,000 & 0,527 & 0,241 & 0,096 \\ 0,325 & 0,527 & 0,000 & 0,735 & 0,289 \\ 0,146 & 0,241 & 0,735 & 0,000 & 0,625 \\ 0,061 & 0,096 & 0,289 & 0,625 & 0,000 \end{pmatrix} \end{matrix} \quad (4.6)$$

Diferentemente das abordagens universais, que pesam todas as unidades de maneira absoluta, a solução proposta aqui segue uma curva de distância de meia-vida. A Figura 4.12 apresenta uma comparação do decaimento exponencial da relação entre

<sup>8</sup>Para o modelo P-DL, que conta com mais de 4 mil unidades espaciais, a média de distância considera os cinco vizinhos mais próximos ( $k=5$ ).

quatro pares do exemplo anterior. Na abordagem global a linha no gráfico segue uma reta no valor 1 para todos os pares de vizinhança, enquanto no método proposto cada relação espacial entre as unidades tem seu próprio peso. Para as unidades em uma área de maior densidade, como o caso  $\overleftrightarrow{P_1P_2}$ , a função local faz com que o peso da relação caia mais rapidamente ( $e^{-d_{ij}^2/2\sigma^2}$  com  $\sigma=0.658$ ), linha azul. Do outro lado, as unidades em uma área de menor densidade, o ponto  $P_5$  do exemplo, segue uma curva de meia-vida mais alongada (para  $\overleftrightarrow{P_2P_5}$ , a função é pesada com  $\sigma=19,689$ ), linha laranja.

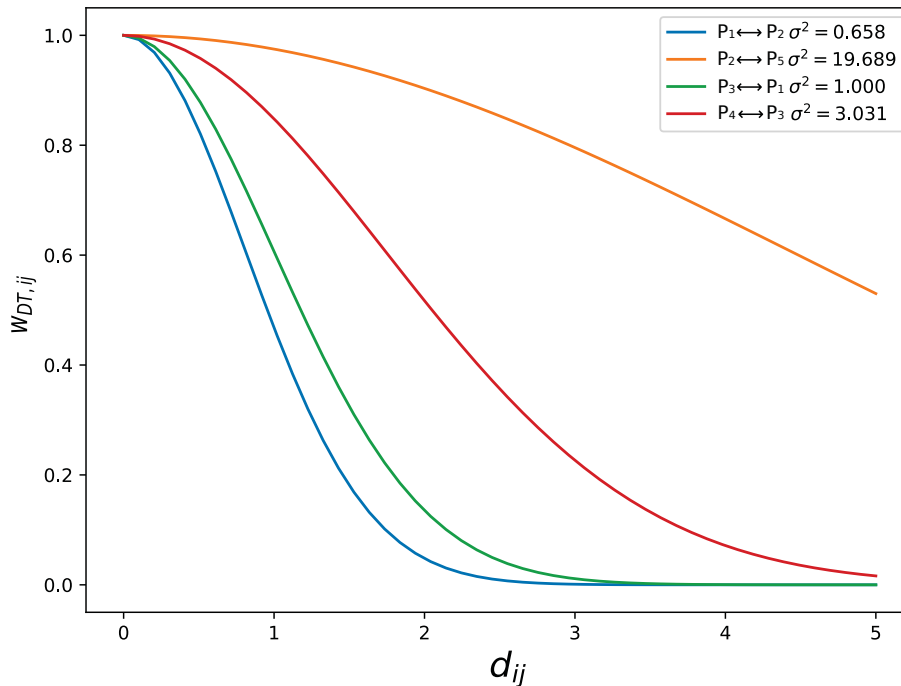


Figura 4.12: Comparação de fatores de distância para relações espaciais com decaimento exponencial com fator local DL

Fonte: O autor (2021).

Em resumo, a forma escolhida para lidar com a realidade da distribuição espacial dos LV do Paraná foi a de considerar as distâncias por rotas transitáveis entre os pontos e o fator de densidade local como ponderação da força das zonas de influência das unidades. Assim, as relações levam em consideração a distribuição heterogênea do eleitorado e a presença de obstáculos espaciais entre os locais de votação.

### 4.2.3 Preparação e Processamento Estatístico de Dados

As ferramentas utilizadas no trabalho foram: para a preparação dos dados o R, o Python, o Qgis e o Geoda; no processamento de estatísticas espaciais o ArcMap e o ArcGIS Pro; e no processamento de estatísticas clássicas o NCSS e o MedCalc.

#### 4.2.3.1 Preparação dos Dados

Eleitorais — do TSE utilizei os dados de resultados eleitorais e de coordenadas geográficas dos LV (a partir do banco de dados do Eleitorado de 2018). O TSE não disponibiliza os dados eleitorais agregados por local de votação e sim por seção eleitoral. A partir do R os dados foram agregados por local de votação e o banco de resultados eleitorais foi unido ao do eleitorado.

$W$  DL — As distâncias transitáveis entre os LV foram medidas utilizando o conjunto de ruas do OpenStreetMap (OSM). Os dados foram calculados com algoritmo de menores distâncias entre origem e destino de pontos com o auxílio da ferramenta de análise de redes do Qgis. O fator local foi aplicado à matriz de distâncias a partir do Python<sup>9</sup>.

$W$  Q1 — A matriz de conexão que considera a contiguidade entre as áreas foi criada através do software Geoda.

#### 4.2.3.2 Processamento de Estatísticas Espaciais

Utilizei dois estimadores de padrões espaciais, um índice de autocorrelação global e um local. Os dois índices são estatísticas univariadas da variável voto percentual. As estatísticas espaciais foram processadas com o auxílio dos softwares ArcMap e o ArcGIS Pro.

Para comparar os métodos utilizo a estatística de autocorrelação espacial  $I$  de Moran, alterando as configurações para os modelos A-Q1, A2-Q1 e P-DL. E para a análise dos territórios eleitorais indígenas o índice aponta a presença e força dos agrupamentos espaciais. Esse é um índice de autocorrelação espacial global, ou seja, analisa o padrão da distribuição dos dados (disperso, aleatório ou concentrado) a partir de todas as relações definidas em  $W$ . O resultado do estimador é o grau de heterogeneidade ou homogeneidade espacial em um coeficiente que varia entre -1 e 1, onde o valor 1 indica a máxima autocorrelação positiva, -1 a mais intensa autocorrelação negativa e 0 indica

---

<sup>9</sup>Agradeço ao Rafael Bernardelli pelo auxílio fundamental no desenvolvimento do código.

total aleatoriedade na distribuição espacial. O I de Moran é o principal estimador da característica do padrão espacial e da força da relação entre objetos espaciais (Darmofal, 2006). Na análise espacial de voto no Brasil este índice pode ser encontrado nos trabalhos de Ames (2003), Soares e Terron (2008), Terron (2009), Terron e Soares (2010), Marzagão (2013) e Gonçalves (2016).

A comparação de mapas é central para a validação de modelos e interpretação de padrões espaciais. Por isso, além da comparação dos resultados das estatísticas de I de Moran global, também faço estimações visuais dos padrões espaciais (VSPC) de cada método. Primeiro, faço as comparações globalmente, buscando padrões gerais nos locais de agrupamentos. Em segundo lugar, comparo os mapas localmente, observando correspondências e diferenças espaciais nos principais agrupamentos a partir da autocorrelação local. A autocorrelação local produz um valor específico para cada OE, permitindo a visualização de agrupamentos de objetos com valores similares (*clusters*) e desviantes (*outliers*). Os padrões locais são interpretados a partir de quatro grupos de autocorrelações: “alto-alto” mostram OE com altas proporções do indicador, cercados de outros OE também com altas proporções do mesmo indicador; “baixo-baixo” OE com baixa proporção, cercados por OE com baixa proporção do mesmo indicador; “alto-baixo” OE com alta proporção, cercados por OE com baixa proporção desse indicador e “baixo-alto” OE com baixa proporção, cercados de OE com alta proporção do mesmo indicador.

Para as concordâncias entre métodos, analiso mapas sobrepostos com a técnica *Translucent Overlay* (OV), onde ambos os mapas são apresentados simultaneamente com transparência. Lobo et al. (2015) comparam as principais técnicas de estimação visual de mapas e os resultados mostram que OV “*is the most versatile technique, featuring good performance for all tasks [...] Translucent Overlay (OV) is the best technique overall*” (Lobo et al., 2015, p. 3580-3581).

Os métodos quantitativos de *Spatial Pattern Comparison* são limitados a comparações “diagonais”, por exemplo como *point-point* e *lattice-lattice*. No meu caso principal, A-Q1 comparado com P-DL, os dados são de tipo *polygon-point*, uma comparação assimétrica. Esse fato inviabiliza a análise quantitativa dos padrões. A próxima seção apresenta os resultados.

### 4.3 CONCORDÂNCIA ENTRE MÉTODOS

Nesta seção, comparo os resultados do índice I de Moran para os métodos A-Q1 e P-DL. No total, a amostra conta com 594 resultados de autocorrelação espacial dos

resultados eleitorais de 297 candidatos. Todos os casos apresentam níveis de significância e valores críticos que permitem rejeitar a hipótese de que os padrões encontrados são efeitos do acaso. Existe menos de 1% de probabilidade de aleatoriedade espacial completa em qualquer um dos resultados dos padrões encontrados na amostra. O valor  $p$  é estatisticamente significativo e o escore  $z$  é positivo e acima do limite crítico, ou seja, a distribuição espacial dos votos é mais concentrada do que o esperado em fenômenos espaciais aleatórios para todos os casos, independentemente do método.

Contudo, existem diferenças significativas entre os resultados. Exploro essas diferenças mais a fundo com base nas seguintes perguntas:

*Perguntas 3: A diferença de concordância entre métodos é sistemática? Como a diferença se manifesta no mapa?*

Para responder a primeira pergunta utilizo o teste de Bland-Altman, também conhecido como Análise de Concordância entre Métodos. Essa é uma ferramenta que mostra a consonância entre dois métodos diferentes que buscam a mesma resolução — no meu caso, dois métodos para avaliar a presença de padrões espaciais de votos com o I de Moran. Os resultados indicam se as diferenças são aleatórias ou se um dos métodos comparados subestima ou superestima as medições de forma reiterada. Para a segunda pergunta utilizo VSPC em mapas que combinam os resultados dos métodos em busca de respostas para as divergências.

#### 4.3.1 Comparação de Índices

A média dos resultados de I de Moran é três vezes maior no método P-DL e a mediana 3,71 vezes. Outro resultado superior para o método P-DL é a amplitude dos resultados  $A=0,774$  contra  $A=0,518$  em A-Q1. O gráfico de violino, Figura 4.13, mostra que as áreas de maiores densidades dos resultados dos índices para o método A-Q1 ficam abaixo do primeiro quartil de P-DL. A maior densidade das observações de A-Q1 está perto do I de Moran em 0,1 — o mesmo número que em P-DL é próximo do valor inferior adjacente<sup>10</sup>. Outra diferença notável está nas distribuições das densidades, A-Q1 apresenta distribuições de característica unimodal, localizada abaixo da mediana. Os resultados do método proposto são mais dispersos, sem uma densidade bem definida, com característica

---

<sup>10</sup>Menor observação que é maior ou igual ao limite interno inferior (UIF), o qual é o primeiro quartil menos  $1.5 \cdot \text{IQR}$ .

multimodal. Não é possível verificar diferenças significativas de resultados entre cargos. Esses primeiros resultados indicam que no método P-DL existe uma maior possibilidade de padrões nos resultados. A-Q1 mostra uma tendência de concentrar os resultados em valores baixos. Fica evidente, também, a diferença na força das autocorrelações encontradas no método P-DL. A maior concentração de resultados de A-Q1 corresponde às menores observações de P-DL. O próximo gráfico mostra como os resultados dos métodos se relacionam.

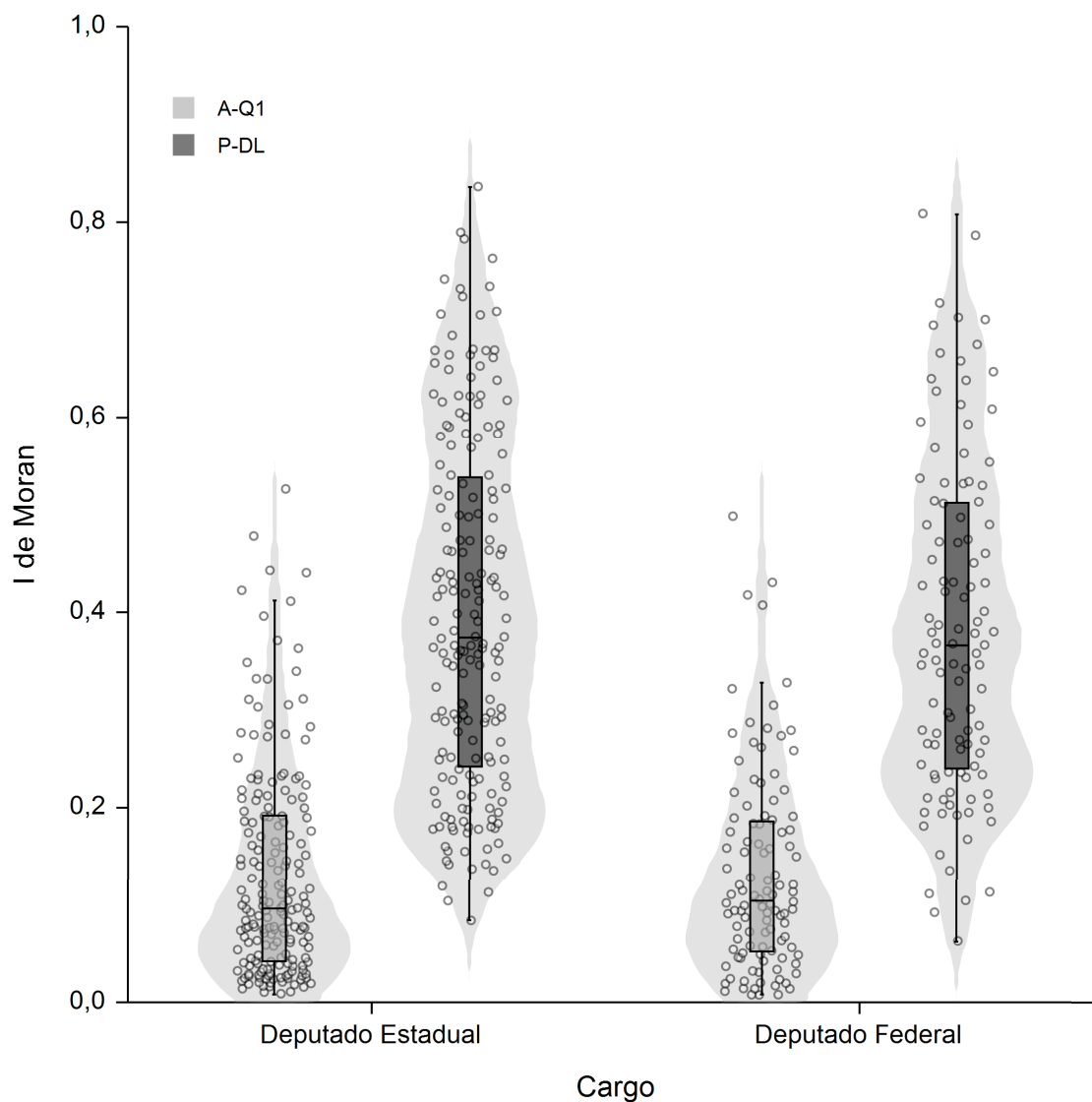


Figura 4.13: Gráfico de Violino com Boxplot de resultados de I de Moran nos métodos P-DL e A-Q1, amostra de candidaturas aos cargos de Deputado Estadual e Federal, 2018, Paraná

Fonte: O autor (2021).

A correlação entre os resultados dos índices é de  $-0,304$  para o cargo de Deputado Federal e de  $-0,403$  para Estadual. O  $R^2$  é de  $0,162$  e  $0,0929$  respectivamente, ambos com significância estatística  $< 0,05$  no teste  $t$ . Entretanto, como é visível na Figura 4.16, para ambos os casos, os resíduos da regressão não seguem uma distribuição normal (hipótese rejeitada pelos testes Shapiro Wilk, Anderson Darling, D'Agostino skewness, D'Agostino Kurtosis e D'Agostino Omnibus). Em suma, não há relação de linha reta entre os resultados dos modelos (rejeitado pelo teste Levene).

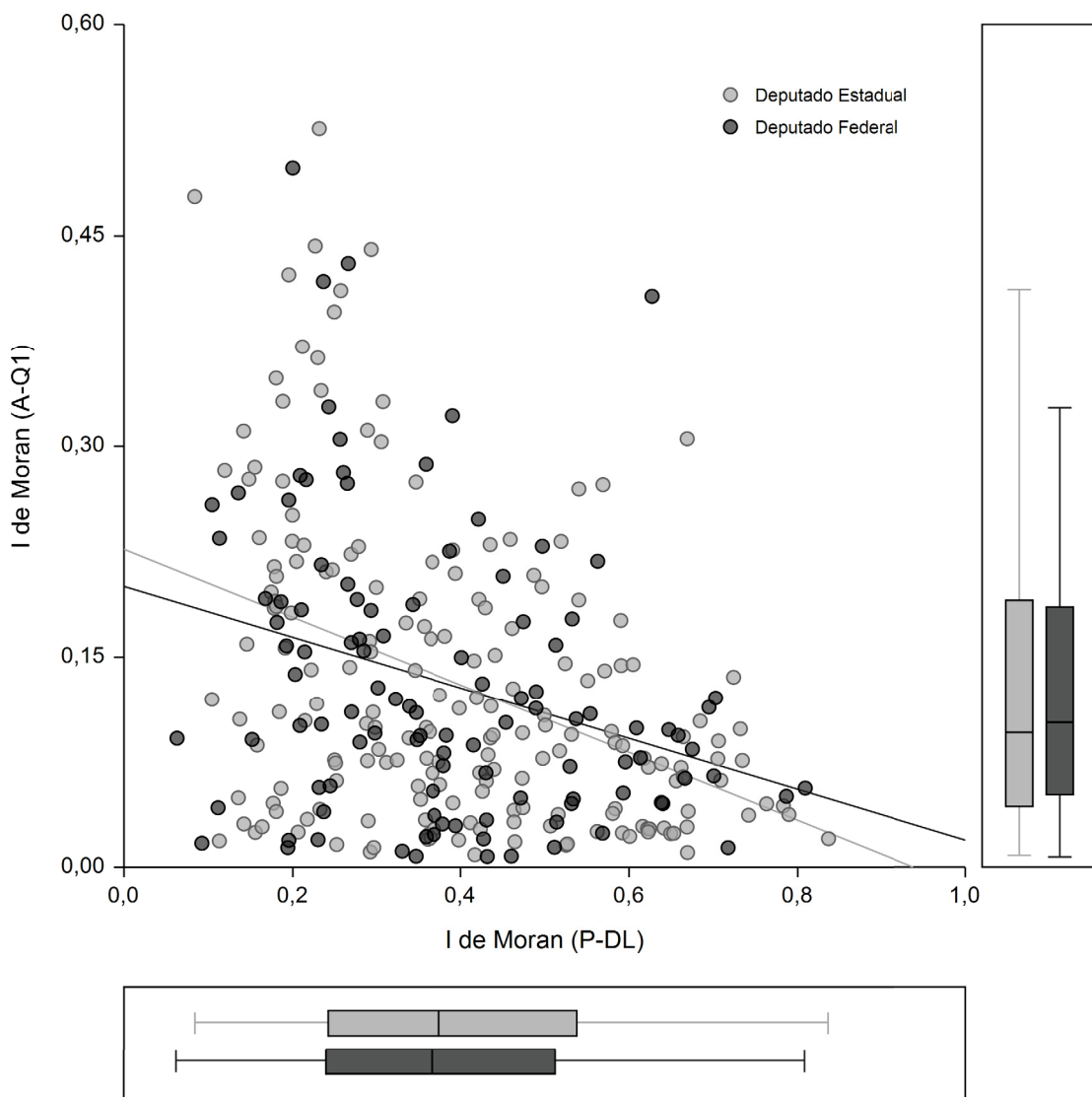


Figura 4.14: Gráfico de Dispersão de resultados de I de Moran nos métodos P-DL e A-Q1, amostra de candidaturas aos cargos de Deputado Estadual e Federal, 2018, Paraná

Fonte: O autor (2021).

Existe uma baixa correlação negativa entre os métodos e não existe relação funcional (no modelo linear simples) entre eles. Esse é um resultado que indica que os métodos não concordam sobre o padrão espacial dos mesmos candidatos. Mas, ainda que o gráfico mostre que diversas observações receberam valores baixos de I de Moran no método P-DL e altos em A-Q1, e vice-versa, não é possível indicar que exista um padrão sistemático a partir desse teste. Para avaliar diferenças sistemáticas entre os métodos, o teste mais indicado é o de Bland-Altman. Como os resultados não mostraram diferenças significativas quanto ao cargo considerado, a partir daqui os resultados de autocorrelação dos votos da amostra de Deputados Federais e Estaduais serão apresentados juntos. A Figura 4.15 apresenta os resultados do teste para a comparação entre A-Q1 e P-DL.



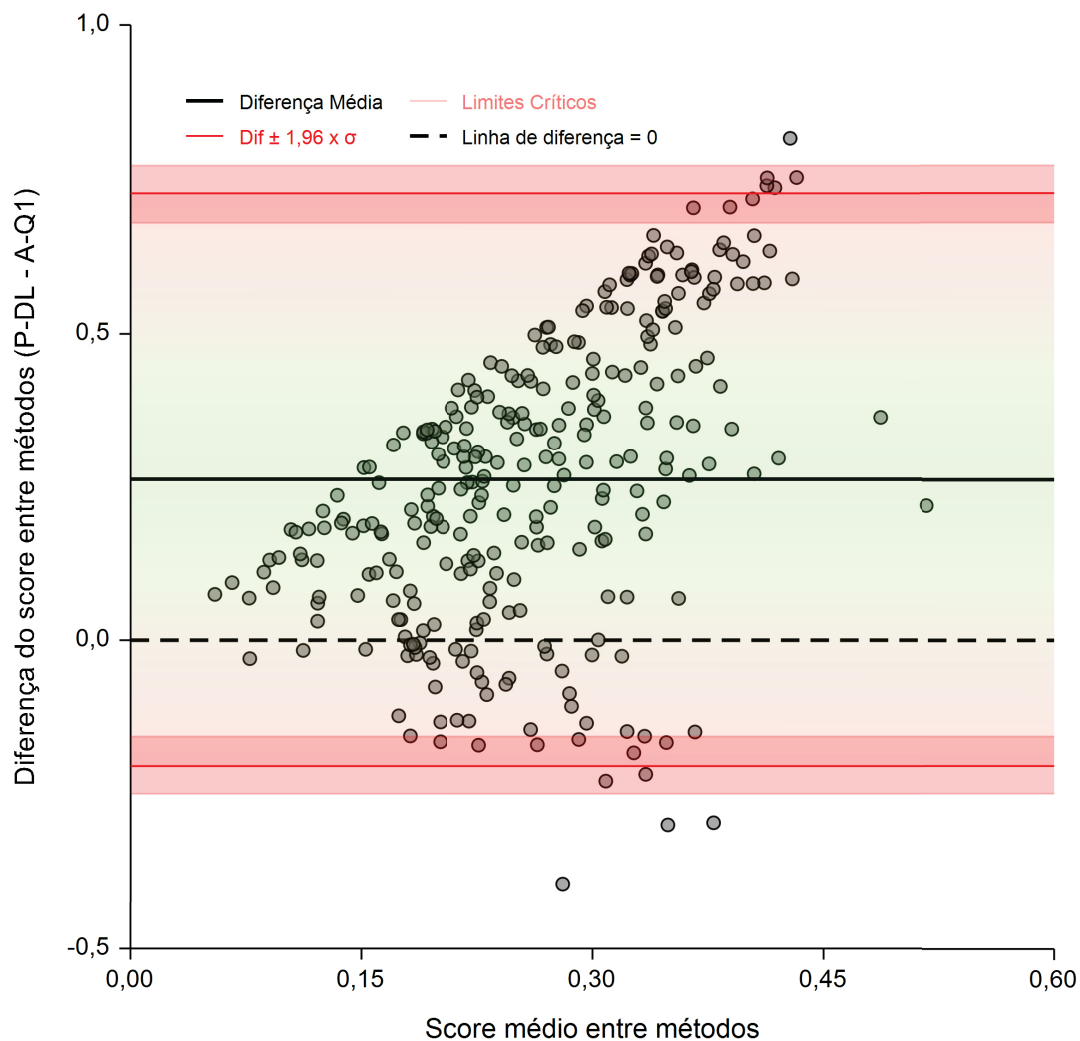


Figura 4.15: Teste de Bland-Altman para concordância entre métodos P-DL e A-Q1 para o índice I de Moran, amostra de candidaturas aos cargos de Deputado Estadual e Federal, 2018, Paraná

Fonte: O autor (2021).

O primeiro resultado do Bland-Altman é o viés entre modelos, a diferença média geral nos valores obtidos com os dois métodos diferentes de medição, representado pela linha preta na Figura 4.15. Quanto mais longe de zero maior é o viés e a discordância entre métodos. O viés nos resultados de I de Moran entre A-Q1 e P-DL é de 0,262. Essa média é mais um indicativo da discordância entre métodos.

Os resultados sugerem que os índices se concentram majoritariamente acima do viés = 0. Isso pode indicar que A-Q1 subdimensiona os padrões espaciais de voto. Para verificar se existe esse subdimensionamento sistemático, o próximo teste aplica uma regressão linear nos dados de diferenças e médias entre métodos.

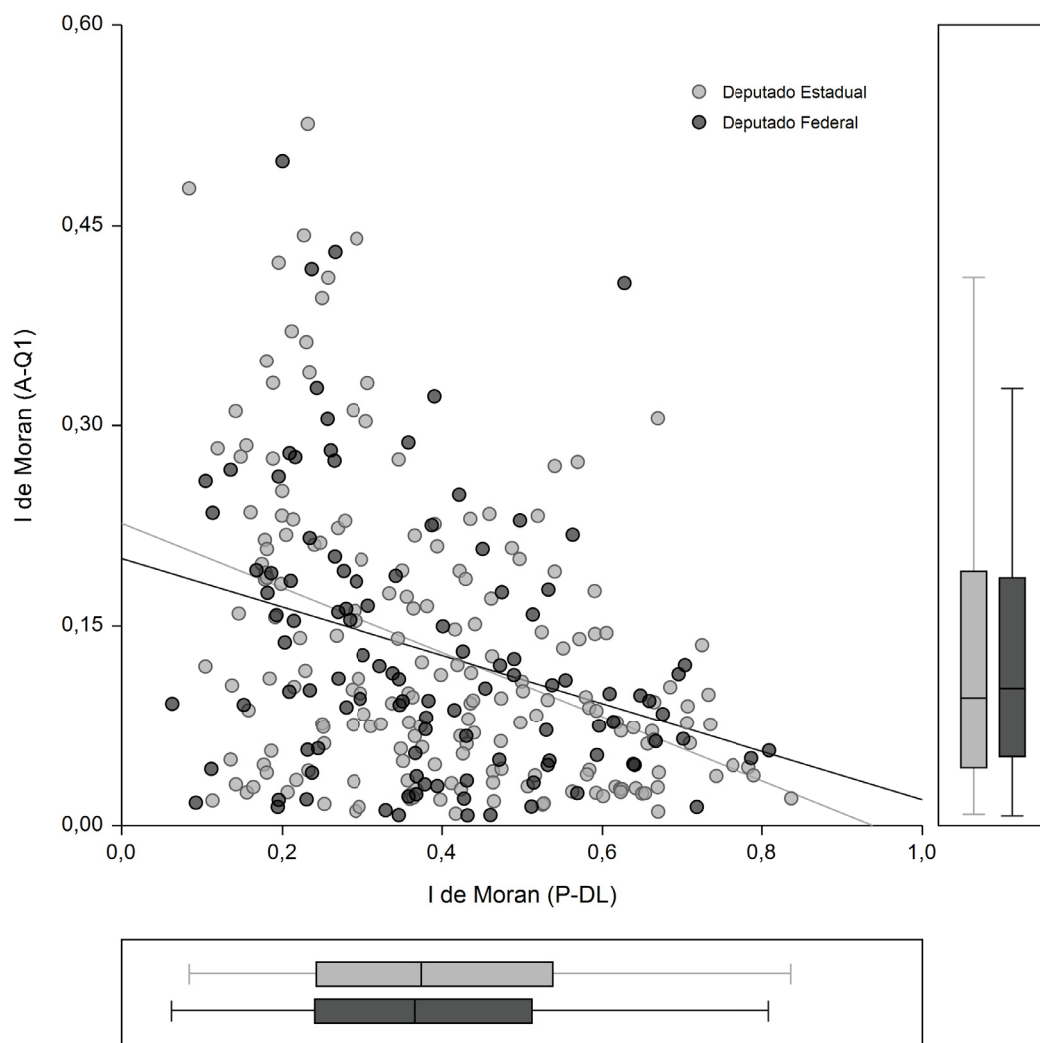


Figura 4.16: Gráfico de Dispersão de médias e diferenças de I de Moran nos métodos P-DL e A-Q1, amostra de candidaturas aos cargos de Deputado Estadual e Federal, 2018, Paraná

Fonte: O autor (2021)

O teste  $t$ , que verifica se as diferenças entre os valores dados pelos métodos são estatisticamente diferentes ( $\mu \neq 0$ ), confirma que não há concordância entre os métodos (t-valor = 9,948, significância = 0,000).

O índice I de Moran é uma estatística dependente do número de unidades de análise e os métodos comparados tem uma diferença de  $n$  maior que 11 vezes (A-Q1 = 399 unidade de análise, P-DL = 4660 unidades de análise). Entretanto, os resultados de comparação entre os modelos A-Q1 e A2-Q1, com a mesma diferença no número de unidades de análise, não indicam viés de proporção entre as médias.

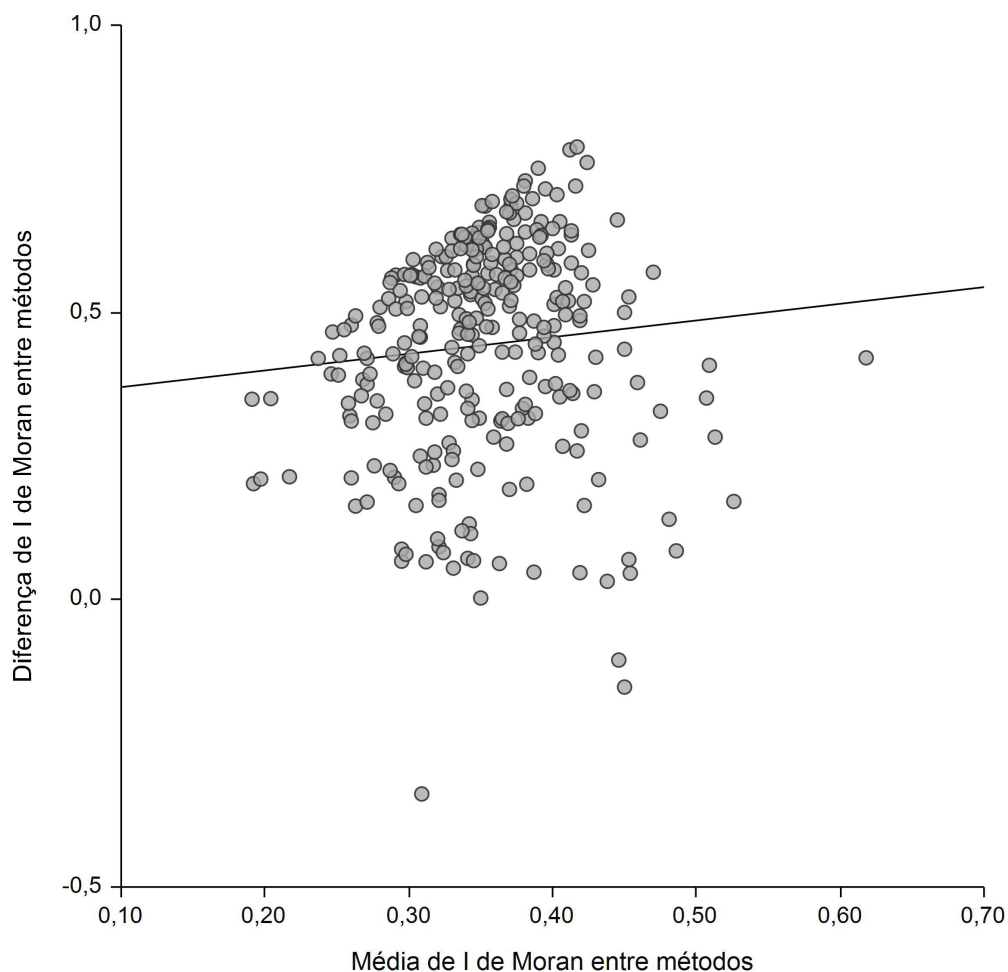


Figura 4.17: Gráfico de Dispersão de médias e diferenças de I de Moran nos métodos A2-Q1 e A-Q1, amostra de candidaturas aos cargos de Deputado Estadual e Federal, 2018, Paraná

Fonte: O autor (2021).

A regressão linear das diferenças de médias nos métodos A-Q1 e A2-Q1 tem um valor de  $R^2$  de 0,008 e correlação de 0,091. Como é visível no gráfico de dispersão, Figura 4.17, as variáveis não apresentam relação de linha reta. Isso impacta o nível de significância do teste  $t$ , resultando em 0,118. Uma vez que o valor é superior ao limite crítico de 95% adotado pelo trabalho, não é possível rejeitar a hipótese nula de dependência entre as variáveis. Dito de outra forma, não há evidência de que haja efeito das médias de resultados de I de Moran sobre as diferenças dos resultados para esses dois métodos.

A Figura 4.18 apresenta as diferenças acumuladas entre os dois métodos LV com o A-Q1 para o índice I de Moran. O método que cria a vizinhança entre locais usando Decomposição Voronoi (Q1) gerou resultados mais distantes dos de A-Q1 comparado aos

do método DT. A diferença entre as medianas é de 0,209. As diferenças no método A2-Q1 foram majoritariamente positivas. Apenas os índices de três candidatos receberam um valor menor comparado ao do A-Q1. A mediana das diferenças para a mesma comparação é de 0,478, com uma maior densidade de casos em 0,46 e 0,6. Os resultados de A2-Q1 são mais concentrados e com menor amplitude.

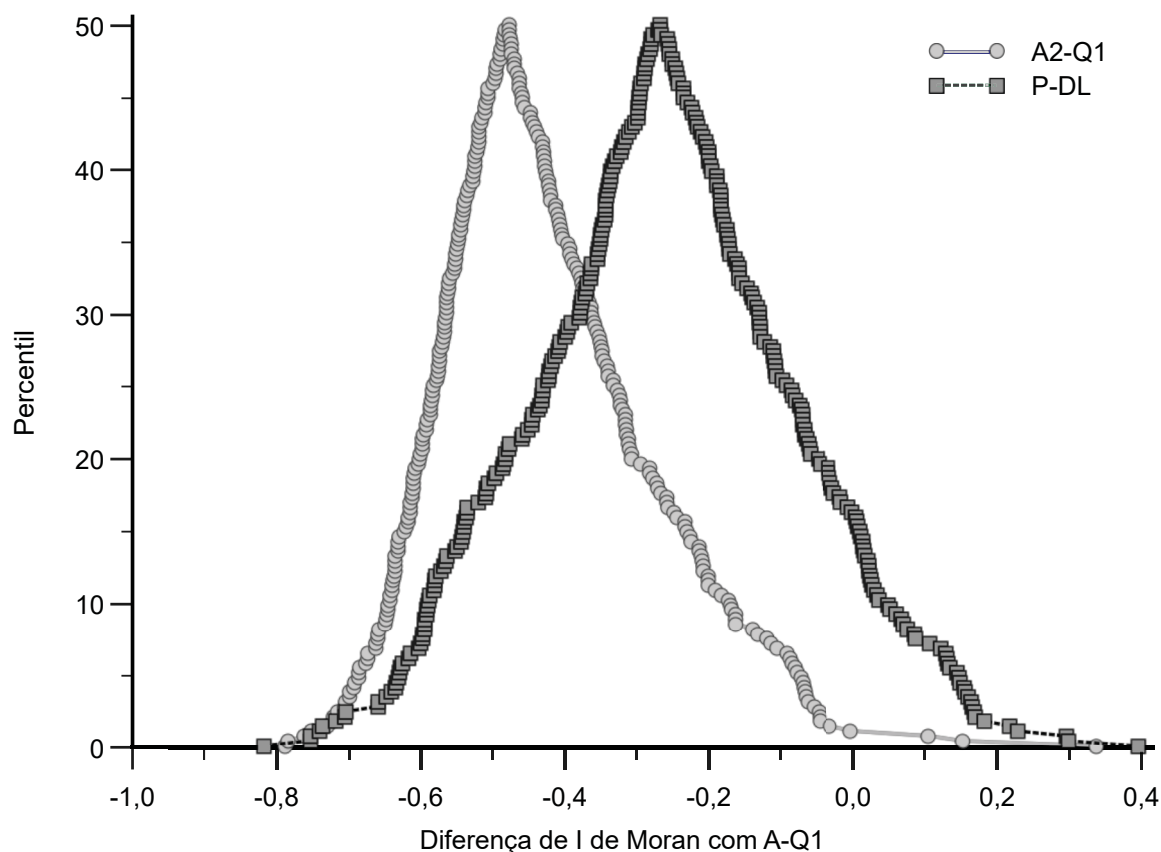


Figura 4.18: Gráfico de Montanha (Função de Distribuição Acumulada Empírica Dobrada) das diferenças de resultados dos índices I de Moran nos métodos A2-Q1 e P-DL em comparação com A-Q1, amostra de candidaturas aos cargos de Deputado Estadual e Federal, 2018, Paraná

Fonte: O autor (2021).

A distribuição das diferenças do modelo P-DL formam uma curva mais próxima da normal, enquanto A2-Q1 concentra os casos à direita, Figura 4.18. Uma interpretação possível desse resultado indica que o método A2-Q1 mostrou uma maior probabilidade de superestimar os padrões espaciais de voto. De sete testes de normalidade aplicados às distribuições, cinco aceitam a normalidade de P-DL, enquanto apenas um aceita a normalidade de A2-Q1.

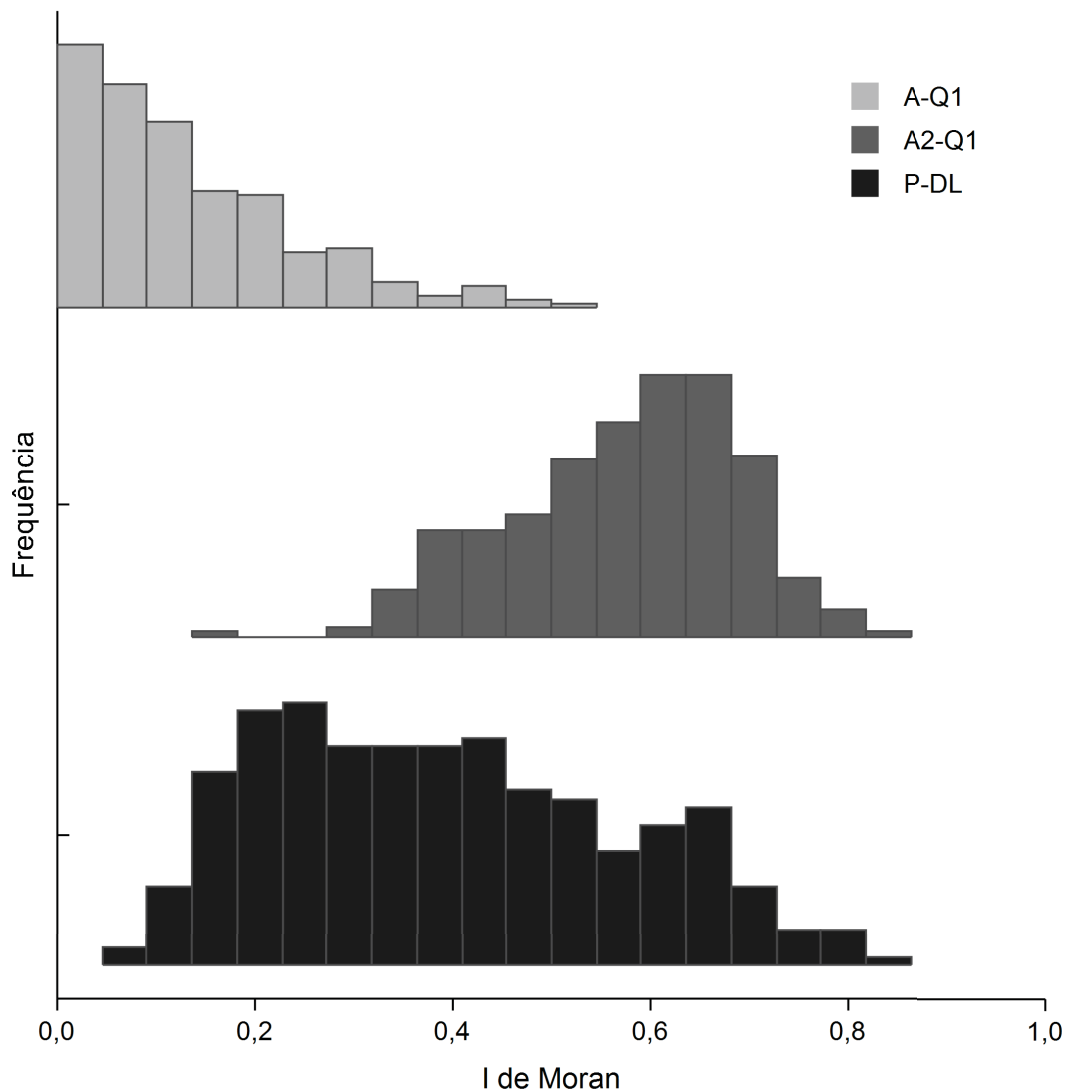


Figura 4.19: Histogramas Comparativos das diferenças de resultados dos índices I de Moran nos métodos A2-Q1 e P-DL em comparação com A-Q1, amostra de candidaturas aos cargos de Deputado Estadual e Federal, 2018, Paraná

Fonte: O autor (2021).

Na próxima subseção exploro as diferenças entre métodos a partir de uma análise visual dos padrões de agrupamento locais.

#### 4.3.2 Comparação Visual de Padrões Espaciais

Como vimos nas comparações entre os resultados dos índices, os métodos não concordam entre si. Nesta seção, comparo os resultados de padrões espaciais em mapas

para as diferentes configurações dos modelos. A análise busca levantar hipóteses das implicações das divergências encontradas.

A diferença mais notável entre os métodos é a maior presença de agrupamentos formados por unidades com votação abaixo do esperado com vizinhança também abaixo do esperado (baixo-baixo) no método P-DL. Pela análise global dos mapas, a diferença dos valores entre métodos parece ser respondida por essa sub-representação de *distritos negativos* no mapa de municípios. O mapa da Figura 4.20 é um exemplo dessa sub-representação. A diferença do índice de autocorrelação para a votação do candidato Gustavo Bonato Fruet é de 0,736. No método que considera o município como unidade espacial de análise e a adjacência imediata como área de influência, o índice é de 0,050. Já no método P-DL o *score* é de 0,787.

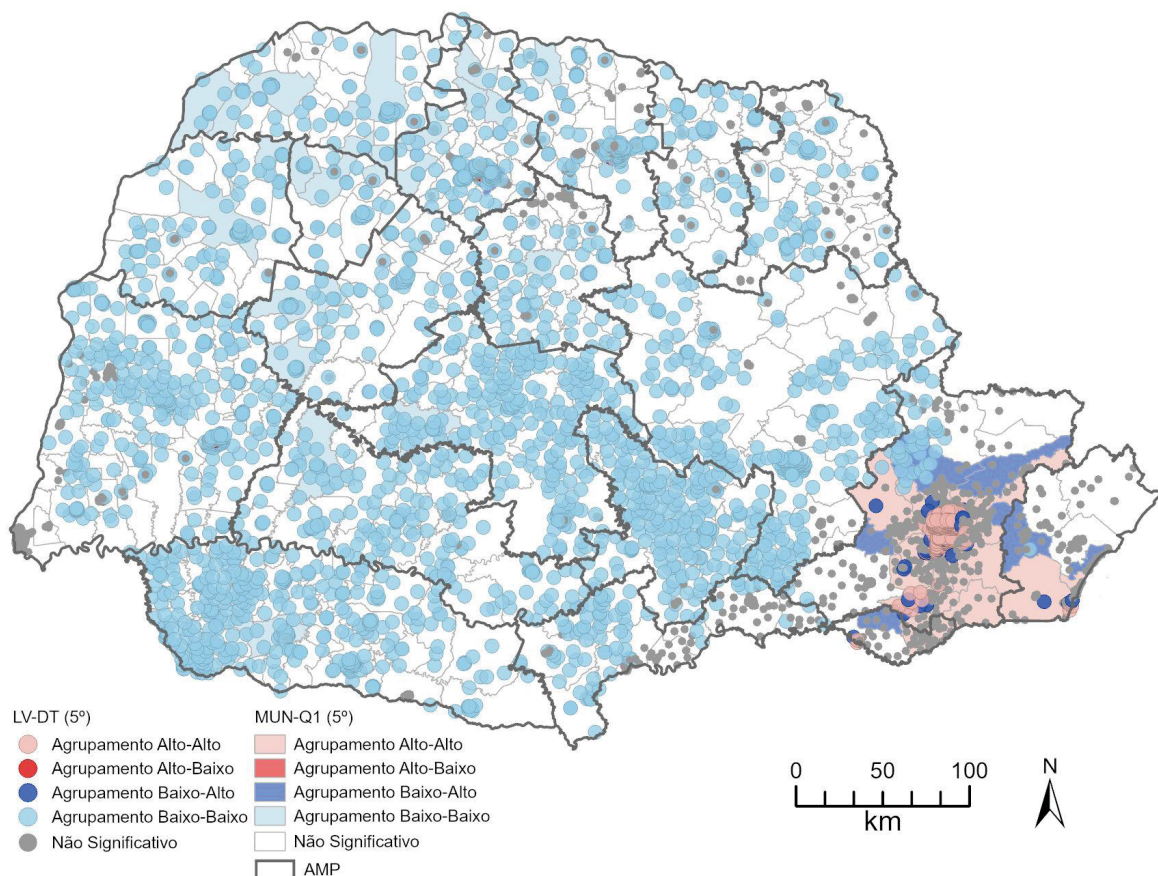


Figura 4.20: Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato Gustavo Bonato Fruet, 2018, Paraná, Brasil.

Fonte: O autor (2021).

A VSPC local dos dados indicam ao menos três tipos de divergências. O primeiro tipo de divergência entre os métodos são as *franjas municipais*, que acontecem no compartilhamento intermunicipal dos padrões espaciais de votação. O segundo vem de padrões intramunicipais de sinais opostos, um *padrão anulado*. O terceiro tipo é o padrão de *simplificação*, dado por reduzir as dinâmicas espaciais complexas. Esses tipos não são “puros”, também são encontrados juntos em uma mesma região.

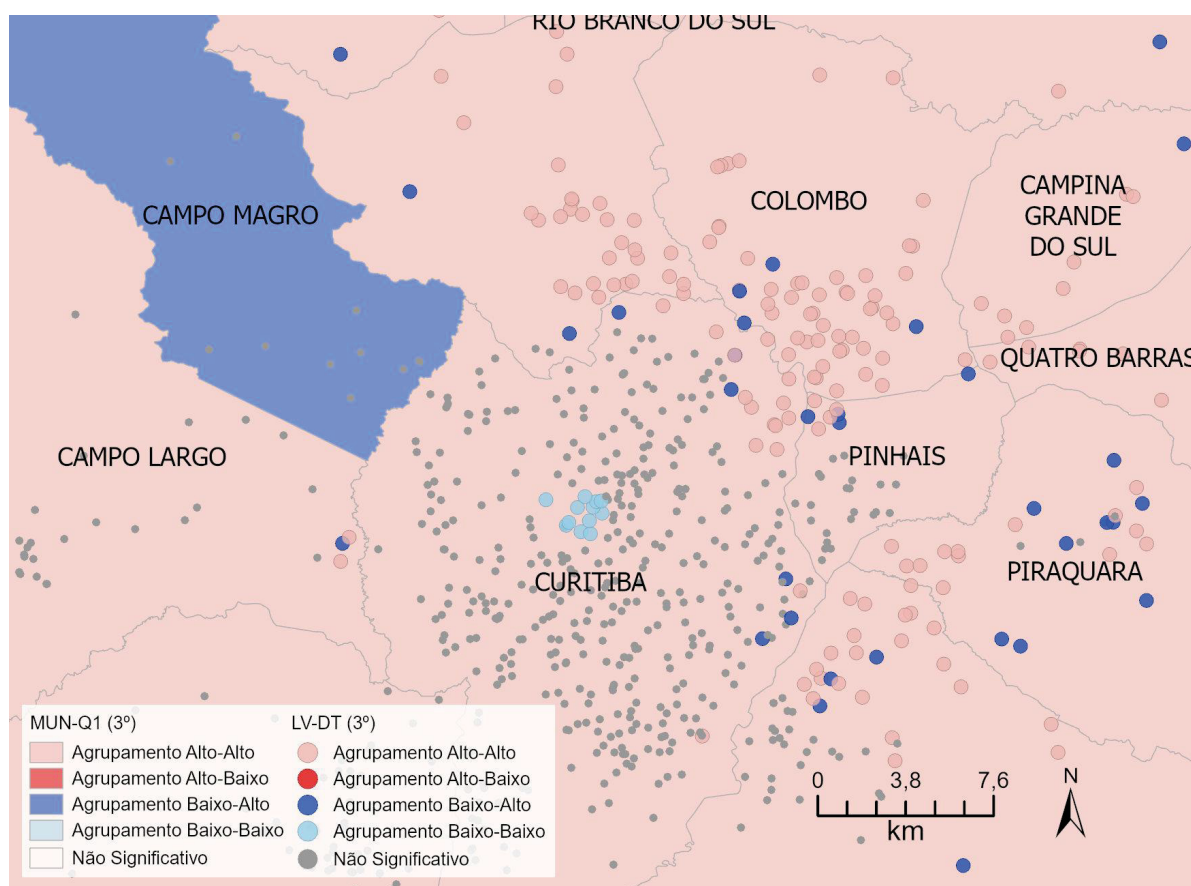


Figura 4.21: Mapa Exemplo de Divergência do tipo Franjas Municipais

Fonte: O autor (2021).

Nota: Mapa referente à votação percentual de Jorge Gomes de Oliveira Brand, diferença de 0,752 no índice I de Moran entre os métodos.

A figura Figura 4.21 apresenta exemplos de divergências do tipo franjas municipais. No método A-Q1, a região representada na imagem concentra majoritariamente unidades de tipo alto-alto. Já no método P-DL, o padrão espacial de votos da região mostra outra realidade. Chamo a atenção para a dinâmica no centro de Curitiba, com agrupamento baixo-baixo, e para a região norte da cidade, nas fronteiras com Colombo e Pinhais, de

agrupamentos alto-alto. O padrão encontrado nos LV das franjas de Curitiba são mais próximos dos padrões dos LV municípios vizinhos do que do centro do próprio município.

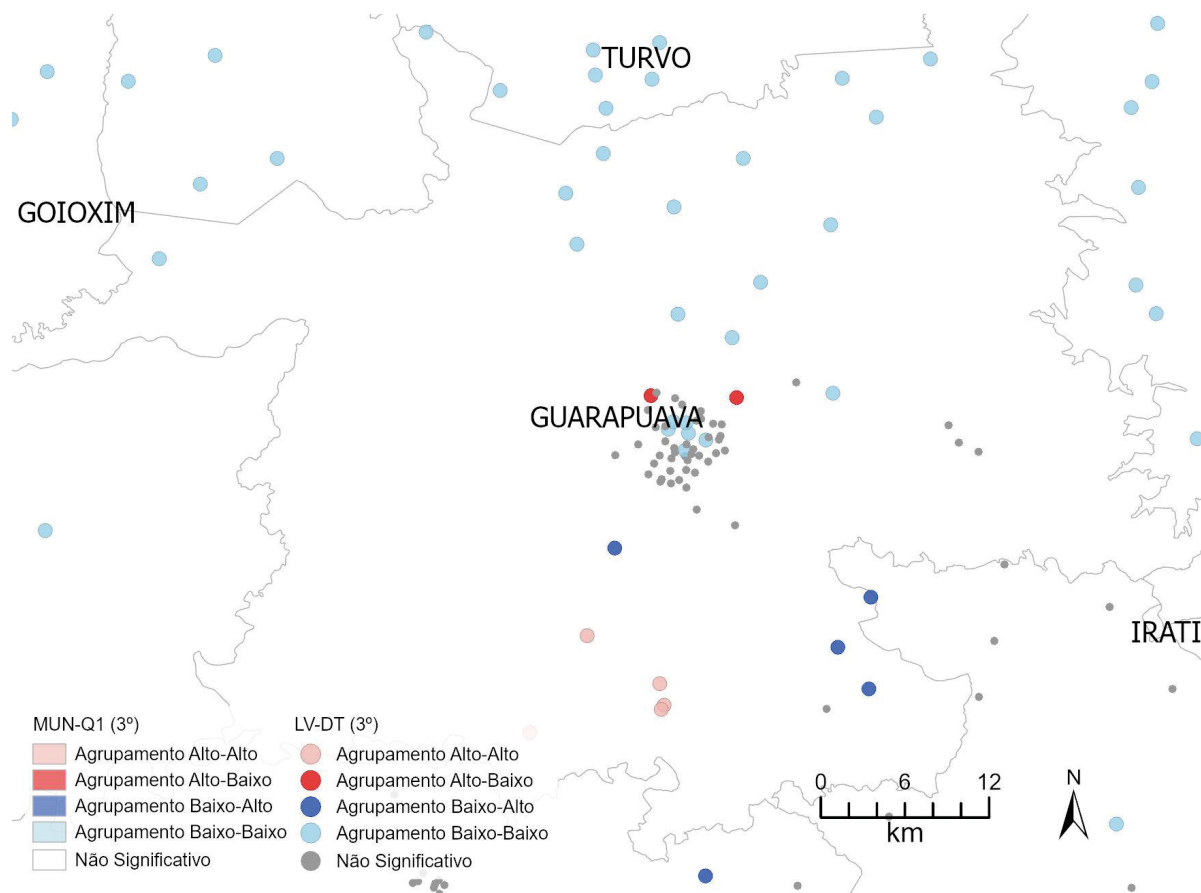


Figura 4.22: Mapa Exemplo de Divergência do tipo Padrão Anulado

Fonte: O autor (2021).

Nota: Mapa referente à votação percentual de Jorge Gomes de Oliveira Brand, diferença de 0,752 no índice I de Moran entre os métodos.

Outra divergência encontrada é a anulação do padrão espacial pelo método A-Q1 dado pela presença de unidades internas com padrões de sentido contrário. No exemplo da Figura 4.22, Guarapuava é uma cidade sem padrão significativo no método A-Q1. Já no método P-DL é possível identificar regiões intramunicipais com padrões espaciais na distribuição de votos. O resultado indica um agrupamento de locais onde o candidato recebeu votos abaixo do esperado na região norte da cidade e acima do esperado na região sul. Entretanto, a agregação dos padrões baixo-baixo com alto-baixo e baixo-alto com alto-alto acabam se anulando na agregação municipal.



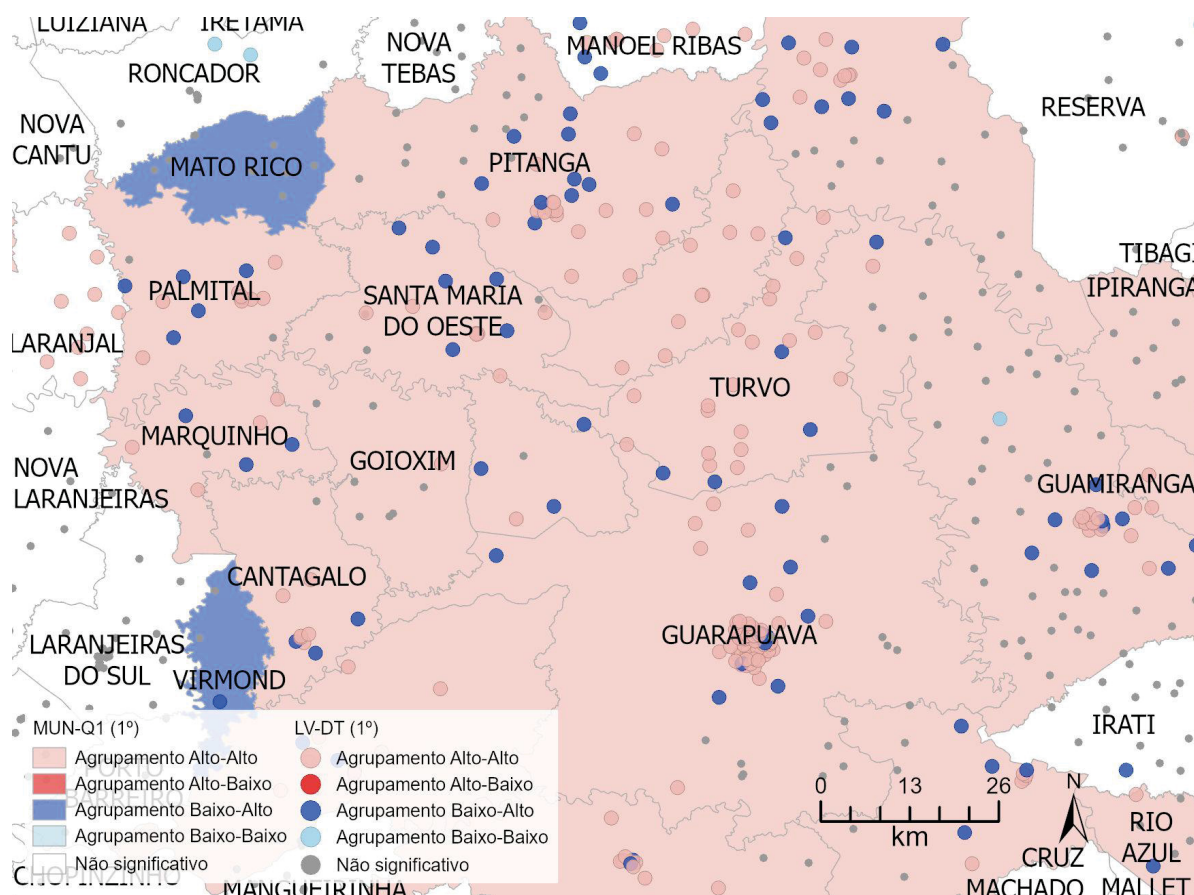


Figura 4.23: Mapa Exemplo de Divergência do tipo Simplificação

Fonte: O autor (2021).

Nota: Mapa referente à votação percentual de Antonio Wandscheer, diferença de  $-0,298$  no índice I de Moran entre os métodos.

O exemplo da divergência de tipo simplificação é de um caso onde o A-Q1 apresentou autocorrelação superior ao P-DL. Na região em destaque na Figura 4.23 o método que agrega os dados por município indicou a presença de um grande agrupamento de tipo alto-alto. O método de granularidade maior mostra unidades espaciais não significativas e de tipo baixo-alto para a mesma região.

#### 4.3.3 Discussão

Em resumo, os testes de concordância realizados indicam que o método P-DL apresentou resultados de autocorrelação de votos que não concordam com os dados por A-Q1. Os resultados mostram que essa discordância entre os métodos tem padrão sistemático onde A-Q1 apresentou tendência de subdimensionar os padrões espaciais

de votos. Os mesmos testes considerando o método A2-Q1 apresentam um viés geral maior (diferença média = 0,447), mas mostram aleatoriedade ao considerar a variação no valor das médias entre métodos. As diferenças encontradas em P-DL são melhor distribuídas do que A2-Q1, que concentra valores altos de diferença. A análise visual das diferenças indicam, ao menos, três tipos de divergências: *franjas municipais*, *padrão anulado* e *simplificação*.

É esperado que exista diferença entre os métodos. O índice testado é dependente dos *OE* e *W* utilizados. Frequentemente o número de casos tem efeito sobre o nível de significância de testes estatísticos. Os índices serem mais altos em P-DL é previsível, em casos sem viés, aumentar o número de observações altera a capacidade dos dados em representar a realidade.

A proposta de analisar os dados eleitorais a partir dos LV amplia as possibilidades de exploração da área com escalas diversas, aprofundando as análises. Aumenta as possibilidades analíticas e permite a identificação de inúmeros padrões espaciais que, usualmente, são classificados de forma genérica como, e.g., concentrada-dominante nas análises de escala de municípios (Fedozzi e Correa, 2015). Na próxima seção apresento um estudo de caso para ilustrar ainda mais os potenciais da utilização dos LV.

#### 4.4 ANÁLISE: UM ESTUDO ORIGINAL DA REPRESENTAÇÃO INDÍGENA

Nesta seção ilustro a potencialidade do método proposto com uma análise simples do padrão de votação de uma candidatura indígena. Utilizo as mesmas definições de P-DL para investigar *territórios eleitorais indígenas* no Paraná. A análise baseia-se em pesquisas anteriores que investigam fenômenos de novas representações na política eleitoral brasileira. Mais especificamente, a exploração de configurações espaciais do fenômeno do voto evangélico (do Nascimento, 2017) e da bancada da bala (Faganello, 2017). Para o primeiro trabalho, o resultado indica que em LV próximos a templos da Igreja Universal do Reino de Deus há incremento de voto no partido apoiado pela igreja. Na segunda pesquisa, que trata do fenômeno relacionado ao crescimento da extrema-direita no país, de representantes ligados aos aparatos militares e policiais, um dos resultados mostra que as bases eleitorais dos políticos analisados estão localizadas ao redor de um complexo de instituições da Polícia Militar. Com base nessas evidências formulo a seguinte pergunta:

*Pergunta 4: A estrutura espacial da votação da candidatura indígena está relacionada à distribuição de Terras Indígenas?*

A Terra Indígena (TI, a partir daqui) é um marco espacial por excelência. São territórios habitados por comunidades que tem etnia em comum, onde são desenvolvidas atividades culturais, produtivas e de reprodução física e cultural <sup>11</sup>. Ainda que existam diversas territorialidades indígenas (e.g., alguns grupos são mais isolados geograficamente, outros mais próximos de centros urbanos, outros que se dividem entre a cidade e a terra indígena), esse parece um caso exemplar para a exploração de efeito de contexto sobre o comportamento.

As teorias contextuais indicam que, grosso modo, interações locais entre pessoas e informações exercem efeito considerável sobre o comportamento eleitoral. Com isso em mente, espera-se que a TI estabeleça um eleitorado regionalmente coeso. Na próxima seção, apresento como exploro o padrão de votos indígenas.

#### 4.4.1 Estratégia Empírica

Os territórios de influência de igrejas e de aparatos policiais, como explorados por do Nascimento (2017); Faganello (2017), representam uma flexibilidade na sistematização conceitual do espaço. Mostram as possibilidades de tratar das dinâmicas territoriais além dos recortes pré-definidos por limites oficiais. No caso dos indígenas, o modelo a partir de LV, permite a análise da representação de um grupo numericamente pequeno e disperso pelo território.

Existem diferentes estratégias empíricas de apreensão de territórios eleitorais ou bases eleitorais. Aqui utilizo a metodologia proposta por Terron (2009), a mais replicada nos casos brasileiros. A metodologia consiste em utilizar as estatísticas de Moran Global e Local. O primeiro índice para determinar a presença e força das bases eleitorais. A segunda para visualizar a localização das bases. Os padrões locais são de três ordem (não significativos, agrupamentos e discrepâncias espaciais), serão interpretados da seguinte forma:

- Agrupamentos espaciais — correlações positivas, região de valores significativamente similares, tanto de valores altos (AA) como baixos (BB). Interpretação dos padrões:

---

<sup>11</sup><http://www.funai.gov.br/index.php/2014-02-07-13-24-32>

- Alto-Alto (AA) — base eleitoral;
  - Baixo-Baixo (BB) — distrito negativo.
- Discrepâncias espaciais — correlações negativas, região de valores significativamente divergentes, tanto de valores altos conectados com baixos (AB) como baixos conectados com altos (BA). Interpretação dos padrões:
    - Alto-Baixo (AB) — Zona de transição ou de vizinhos opostos
    - Baixo-Alto (BA) — Zona de transição ou de vizinhos opostos

O estudo foi conduzido com dados do Estado do Paraná nas eleições gerais de 2018. Das 1.183 candidaturas aos cargos de deputado federal e estadual para esse pleito, apenas uma foi autodeclarada como indígena. O candidato autodeclarado é Ivan Bribis Rodrigues (Ivan Kaingang, na urna), indígena do Povo Kaingang. O candidato é uma liderança política atuante principalmente na TI Apucarantina <sup>12</sup>. A candidatura a deputado estadual obteve um total de 1.766 votos.

No Paraná estão localizadas 23 TI, que somam uma população de cerca de 12.408 habitantes. Se considerarmos que todos os votos do candidato foram dados por indígenas habitantes de TI, os votos representariam cerca de 14% da população total. É um grupo relativamente pequeno e disperso pelo território. Considerando toda a extensão das terras 1253 km<sup>2</sup>, a densidade demográfica é de 9,9 hab/km<sup>2</sup>. Aproximadamente 70% dos indígenas do Estado pertencem ao povo Kaingang <sup>13</sup>, a etnia do candidato analisado aqui. A próxima seção apresenta os resultados da distribuição espacial.

#### 4.4.2 Resultados

Os votos da candidatura indígena analisada apresentam padrão concentrado. O escore  $z$  é positivo e acima do limite crítico ( $z = 27,50$ ) e o  $p$  valor indica relação estatisticamente significativa com margem de 99% de confiança. Embora os votos não sejam dispersos aleatoriamente no território, a autocorrelação espacial é baixa:  $I = 0,161$ .

<sup>12</sup><http://campanhaindigena.org/2020/10/09/ivan-kaingang/>

<sup>13</sup><http://www.educadores.diaadia.pr.gov.br/modules/conteudo/conteudo.php?conteudo=554>

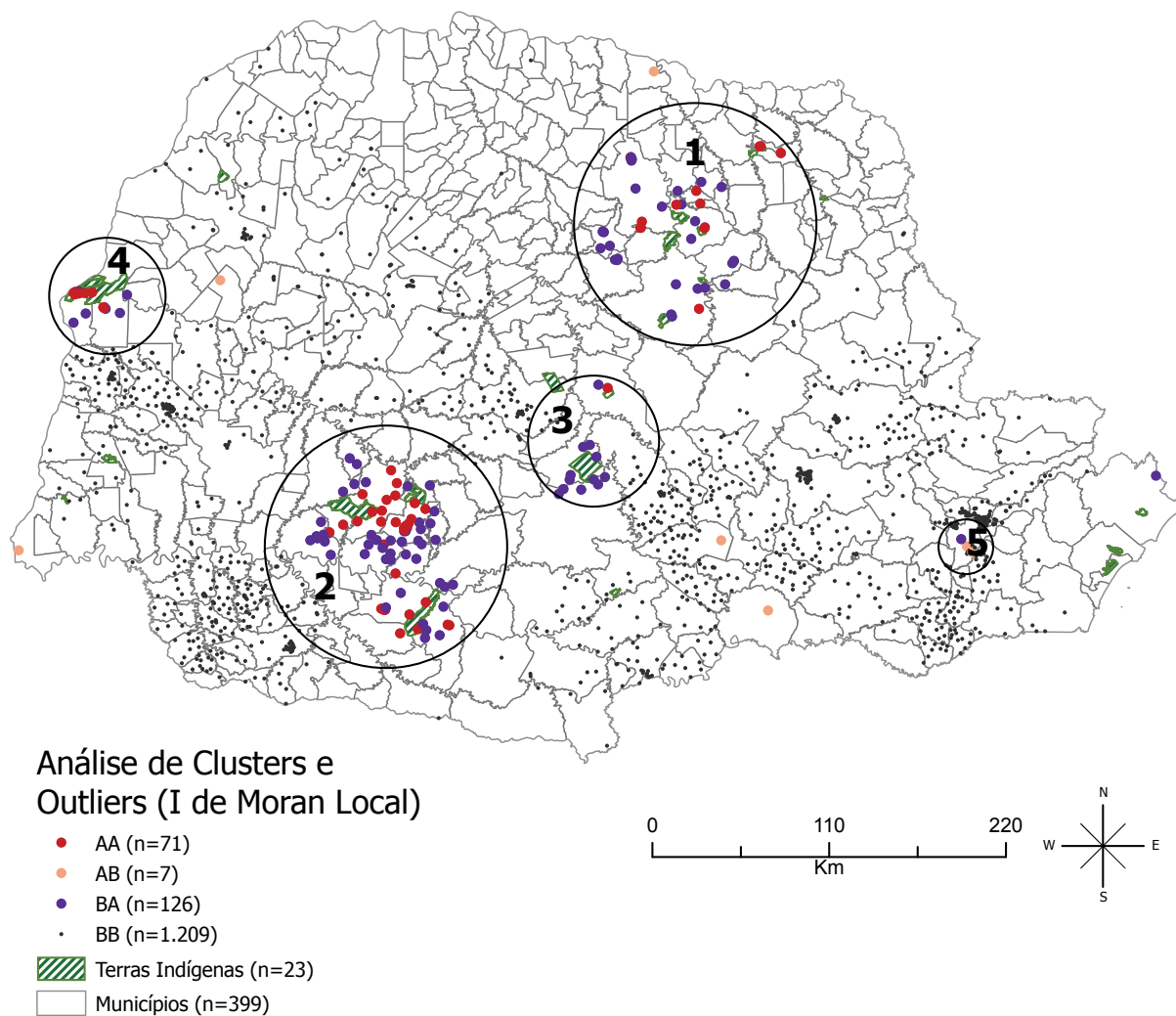


Figura 4.24: Indicador Local de Associação Espacial de candidatura indígena, Ivan Kaingang, Paraná, 2018  
Fonte: O autor (2021).

Uma visão global do mapa do indicador local de associação espacial, Figura 4.24, sugere que as marcas territoriais da votação estão associadas às Terras Indígenas. Mas mostram uma dinâmica espacial que é não comum nas explorações (ver, e.g., os resultados de Soares e Terron (2008); Faganello (2017); do Nascimento (2017)): as áreas de divergência são maiores que as bases eleitorais. Parece refletir uma estrutura espacial de "*vizinhos indesejados*", zonas de transição brusca entre padrões do tipo AA e BB que resultam, sobretudo, em BA. Os próximos mapas focam nos cinco agrupamentos principais para uma exploração mais detalhada desse padrão.

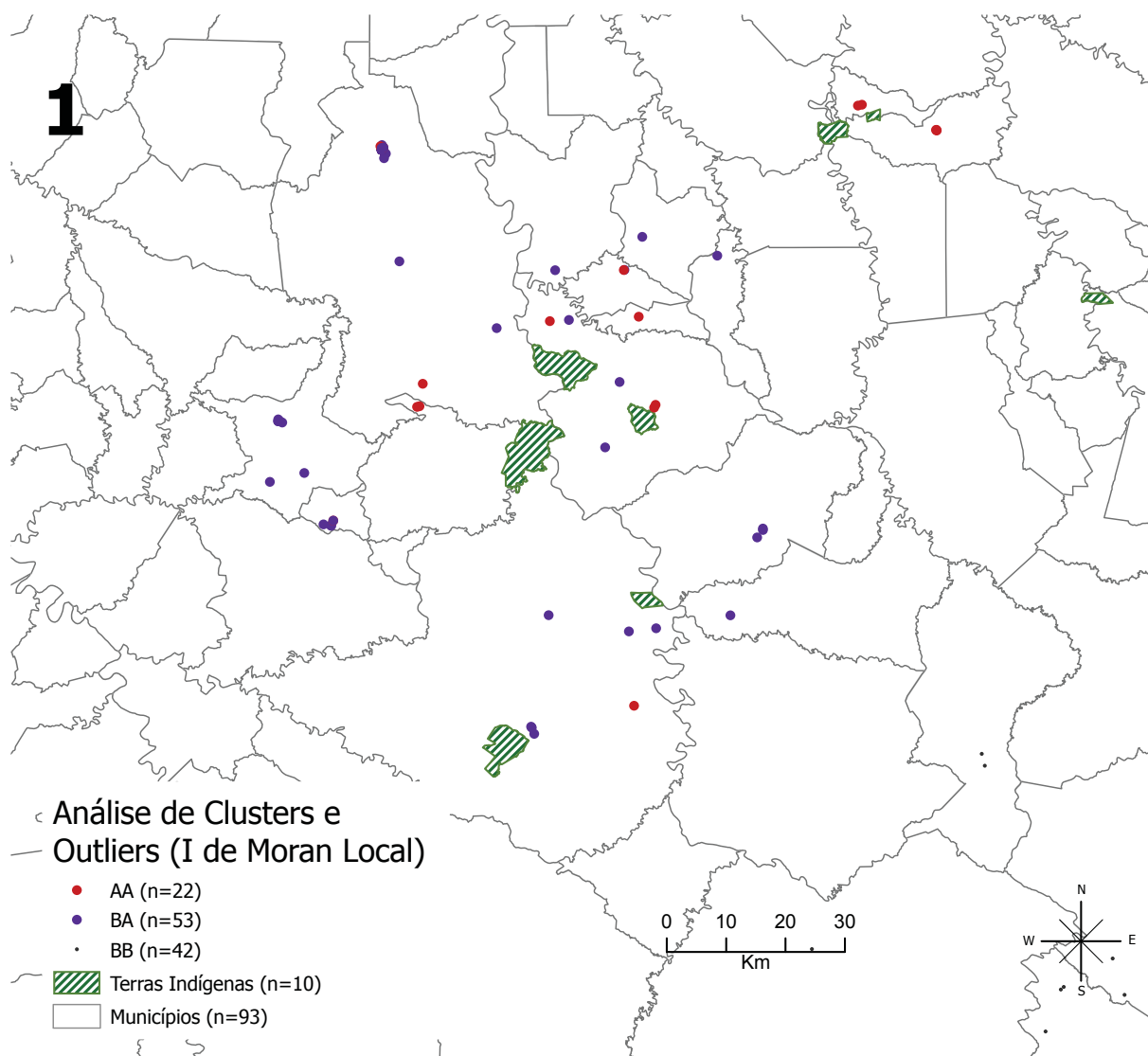


Figura 4.25: Indicador Local de Associação Espacial de candidatura indígena, Ivan Kaingang, Paraná, 2018 (1)

Fonte: O autor (2021).

(1) O local que agrupa o maior número de TI, Figura 4.25, soma 3.964 habitantes<sup>14</sup>. Essa é a região onde o candidato Ivan Kaingang nasceu e onde atua mais fortemente. O território em destaque é a região metropolitana de Londrina, a segunda maior cidade do estado. Nessa região a base eleitoral da candidatura indígena é formada por 22 LV, enquanto os "anti-clusters" somam 53 unidades. Esse é o maior exemplo de quebras bruscas. O padrão pode ser explicado pela maior dispersão da população indígena nessa área. As 10 TI na região variam entre 129 e 1.752 habitantes ( $\sigma = 568$ ). Podemos

<sup>14</sup>Essa é a fonte principal dos dados sobre população e etnias indígenas a partir daqui: <https://terrasindigenas.org.br/pt-br/brasil>

interpretar que a região forma diversas pequenas bases eleitorais que são cercadas por zonas de divergência.

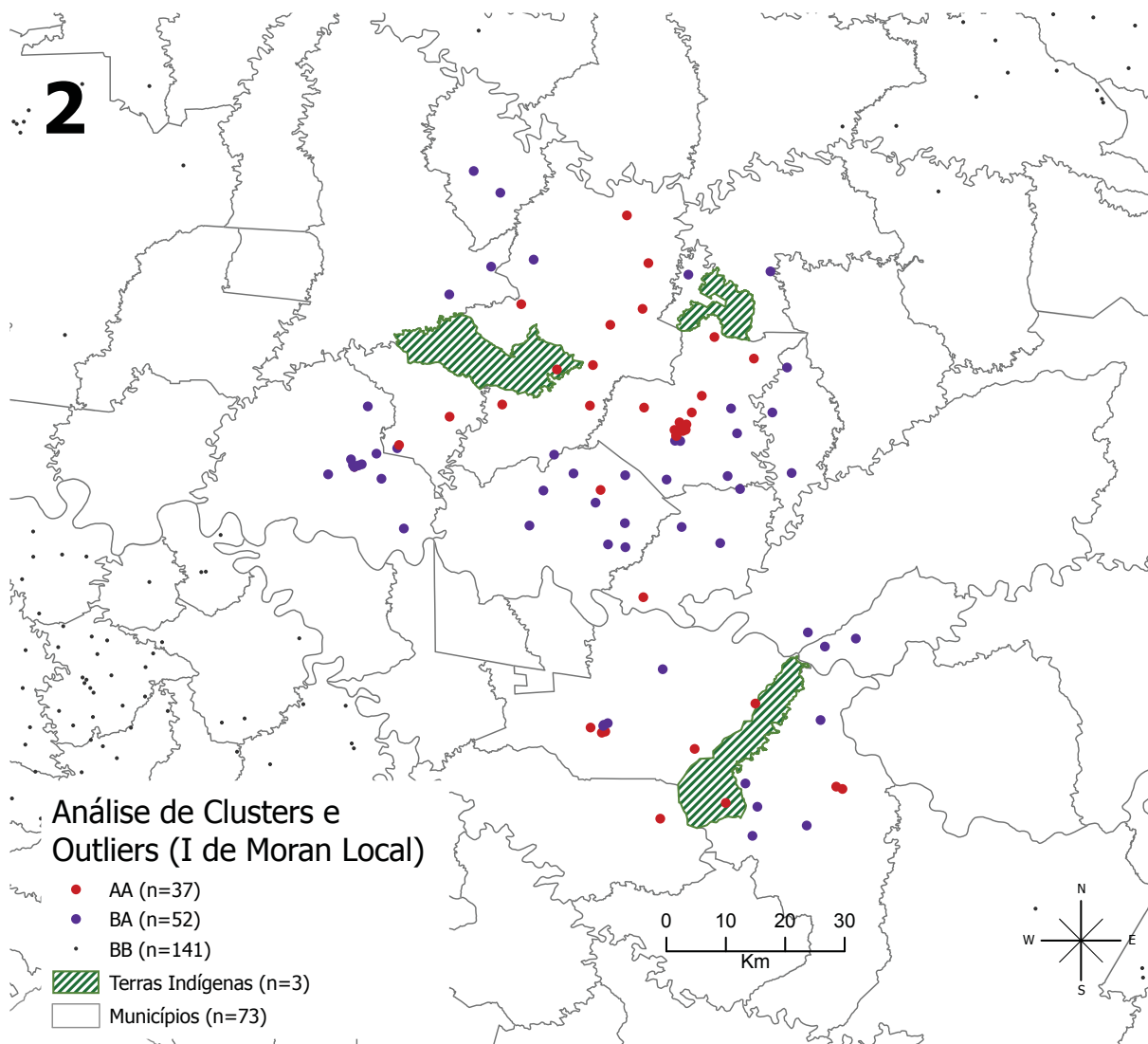


Figura 4.26: Indicador Local de Associação Espacial de candidatura indígena, Ivan Kaingang, Paraná, 2018 (2)

Fonte: O autor (2021).

(2) A Figura 4.26 destaca as TI entre a região do Centro e Centro-Sul do Paraná. Essa é uma região de características mais rurais e é onde se encontra o maior número de habitantes de terras indígenas. As três TI na região somam 4.184 habitantes, com alta concentração na TI Rio das Cobras (a região mais a oeste no mapa). O número de OE com padrão de zona de transição também é superior as de padrão de base eleitoral, 52 e 37, respectivamente. Mas, diferentemente do padrão da Figura 4.25, as bases aparecem mais concentradas.

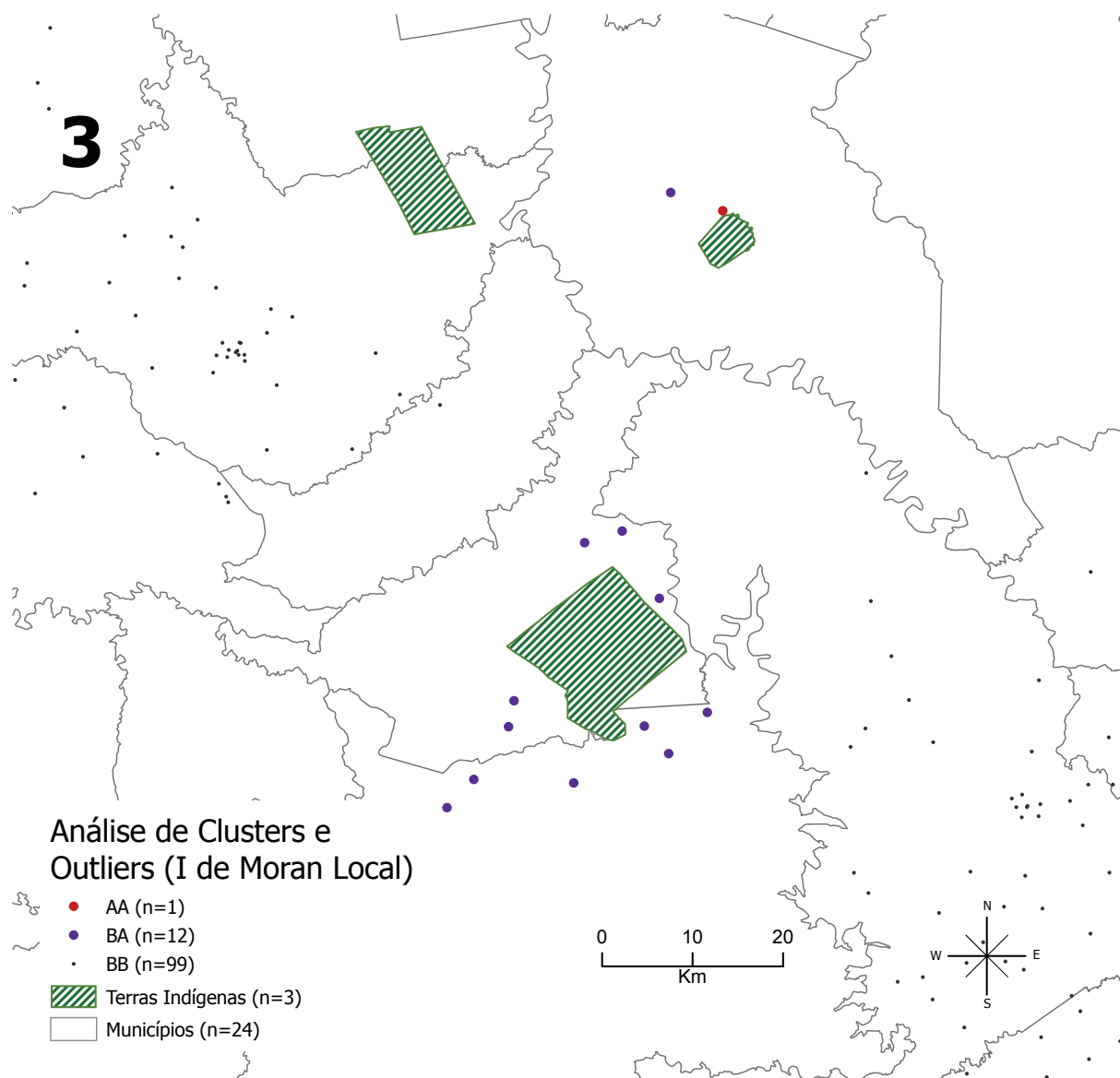


Figura 4.27: Indicador Local de Associação Espacial de candidatura indígena, Ivan Kaingang, Paraná, 2018 (3)

Fonte: O autor (2021).

(3) A Figura 4.27 destaca as TI da região central do estado. Nas três TI nesse espaço vivem 2.900 indígenas. Essa região apresenta o padrão de maior isolamento de base do candidato. Apenas uma unidade pode ser considerada AA e 12 BA. Nessa região o candidato apresenta um baixo desempenho, ainda que as três TI sejam da mesma etnia do candidato (Kaingang).



## 4

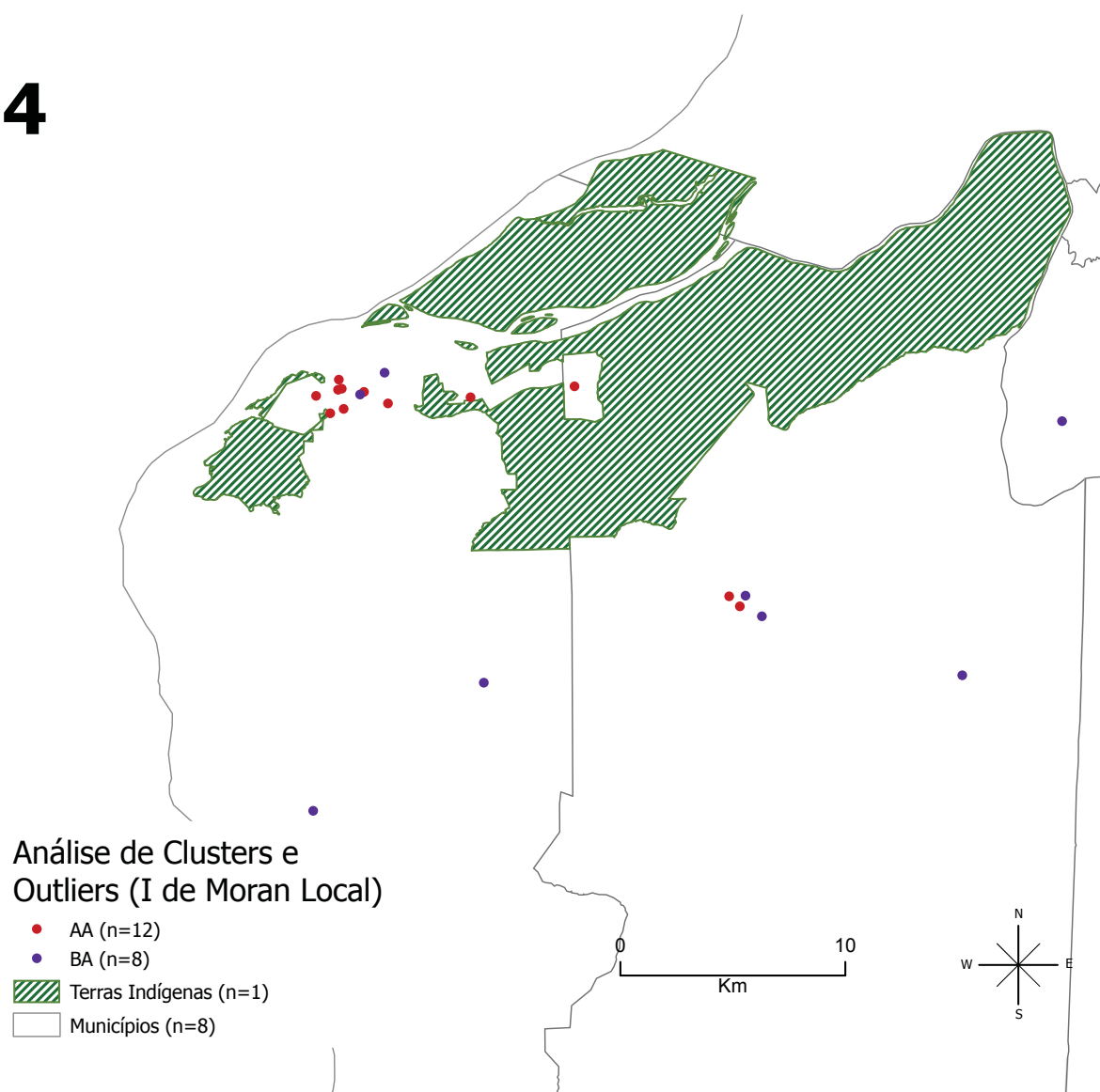


Figura 4.28: Indicador Local de Associação Espacial de candidatura indígena, Ivan Kaingang, Paraná, 2018 (4)

Fonte: O autor (2021).

(4) A quarta região destacada, Figura 4.28, é a da TI Tekoha Guasú Guavirá onde vivem 1.360 indígenas. Essa TI fica no Noroeste do Paraná, na divisa com Mato Grosso do Sul. Essa é a única região de concentração de votos onde o número de unidades que formam a base eleitoral supera as unidades divergentes. Isso surpreende pelo fato da região não ter terras da etnia do candidato. Essa TI é da etnia Nhandevas, relacionada ao povo Guarani. O resultado indica que, embora a estrutura espacial de concentração de votos se relacione claramente à presença de TI, não há uma relação evidente entre a etnia do candidato e das terras onde estão as suas bases eleitorais.

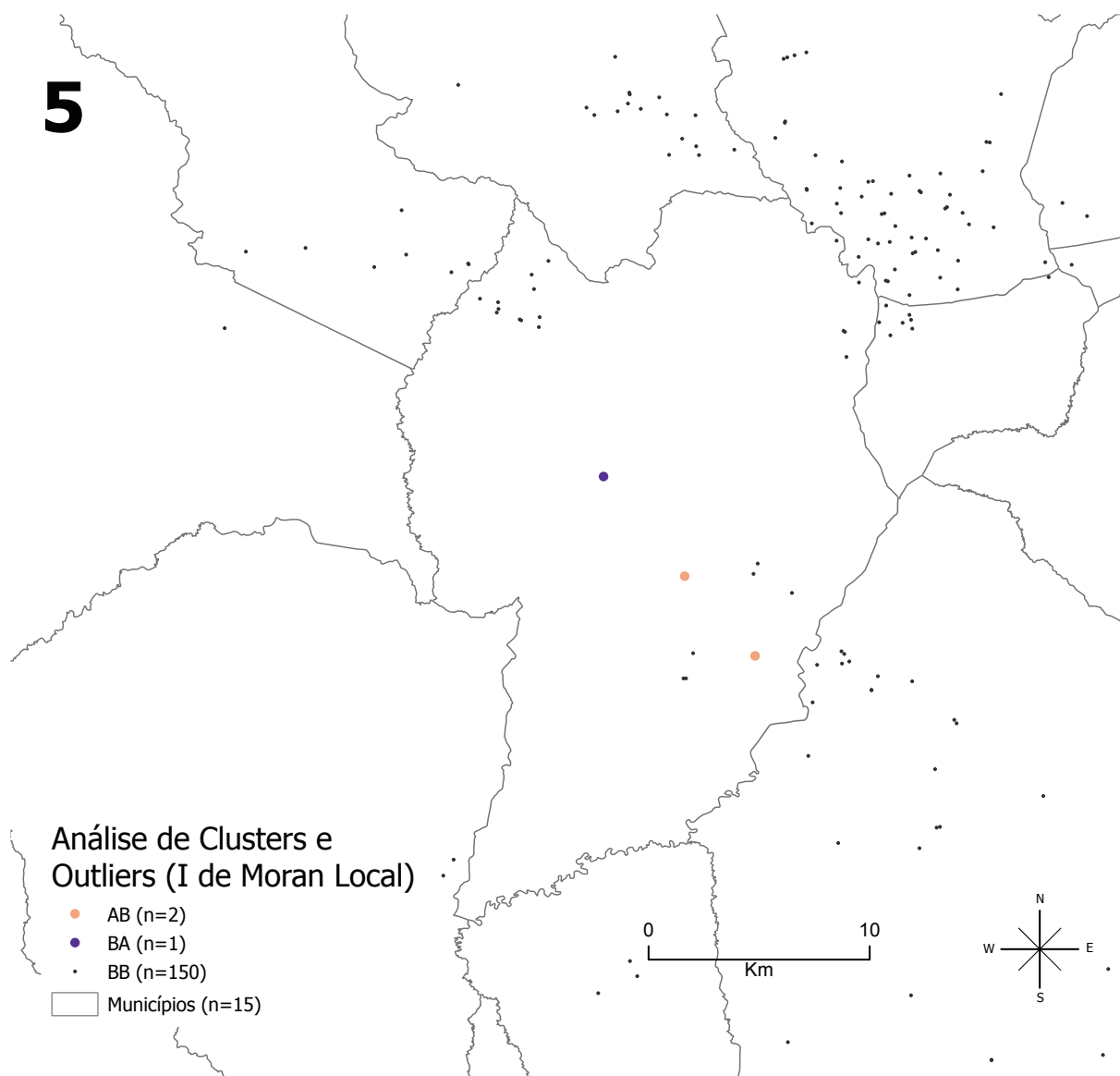


Figura 4.29: Indicador Local de Associação Espacial de candidatura indígena, Ivan Kaingang, Paraná, 2018 (5)

Fonte: O autor (2021).

(5) A região metropolitana de Curitiba, Figura 4.29, não possui nenhuma TI, mas a capital conta com uma aldeia urbana, a Kakané Porã (Fachini, 2018). Os padrões espaciais de votação na cidade não mostram uma base eleitoral bem delimitada, apenas unidades desviantes. Dois dos OE desviantes são do tipo alto com entorno baixo, mas a localização desses pontos não corresponde à localização da aldeia Kakané Porã.

Os resultados indicam uma estruturação espacial da votação da candidatura do representante indígena. Essa análise inicial abre diversas perguntas sobre a dinâmica de representação político-eleitoral dos indígenas. Por que em algumas TI o candidato é bem

votado e em outras não? Qual o papel das relações entre lideranças das etnias? Qual a lógica dos mecanismos de coordenação de campanha? Existe uma estrutura de *brokerage* indígena?

#### 4.4.3 Discussão

No Brasil de Bolsonaro na presidência com seu discurso "lamentável", "ofensivo", "racista" e "paranoico" em relações aos indígenas — segundo líderes de instituições indígenas <sup>15</sup> — a representatividade parlamentar do grupo é um tema caro. Os indígenas representam 0,42% da totalidade da população brasileira. Ao todo, as candidaturas indígenas representam 0,47% em 2018. Em candidaturas autodeclaradas a proporção é justa. Entretanto, dos 120 candidatos apenas dois foram eleitos, a representação cai para 0,009%.

Junto aos trabalhos sobre os territórios de influência de igrejas e de aparatos policiais, como feitos por do Nascimento (2017); Faganello (2017), a análise dos resultados eleitorais indígenas representam uma flexibilidade na operacionalização do espaço. Evidenciam as possibilidades de tratar das dinâmicas territoriais além dos recortes pré-definidos por limites oficiais. No caso dos indígenas, o modelo a partir de LV, permite a análise da representação de um grupo numericamente pequeno e disperso pelo território.

Diversas estratégias metodológicas podem surgir desses dados. E.g. nos nomes dos LV de 2018 constam 543 identificações de "Aldeia", "Terra Indígena", "Comunidade Indígena" e "Reserva Indígena" — como o exemplo ilustrado na Figura 4.30. Dados inexplorados que podem ser utilizados para o desenvolvimento de desenhos de pesquisas para se pensar temas como a representação política desse grupo. Uma aproximação simples que pode aumentar o número de unidades de análise para recursos estatísticos.

---

<sup>15</sup><https://www.bbc.com/portuguese/brasil-49819045>

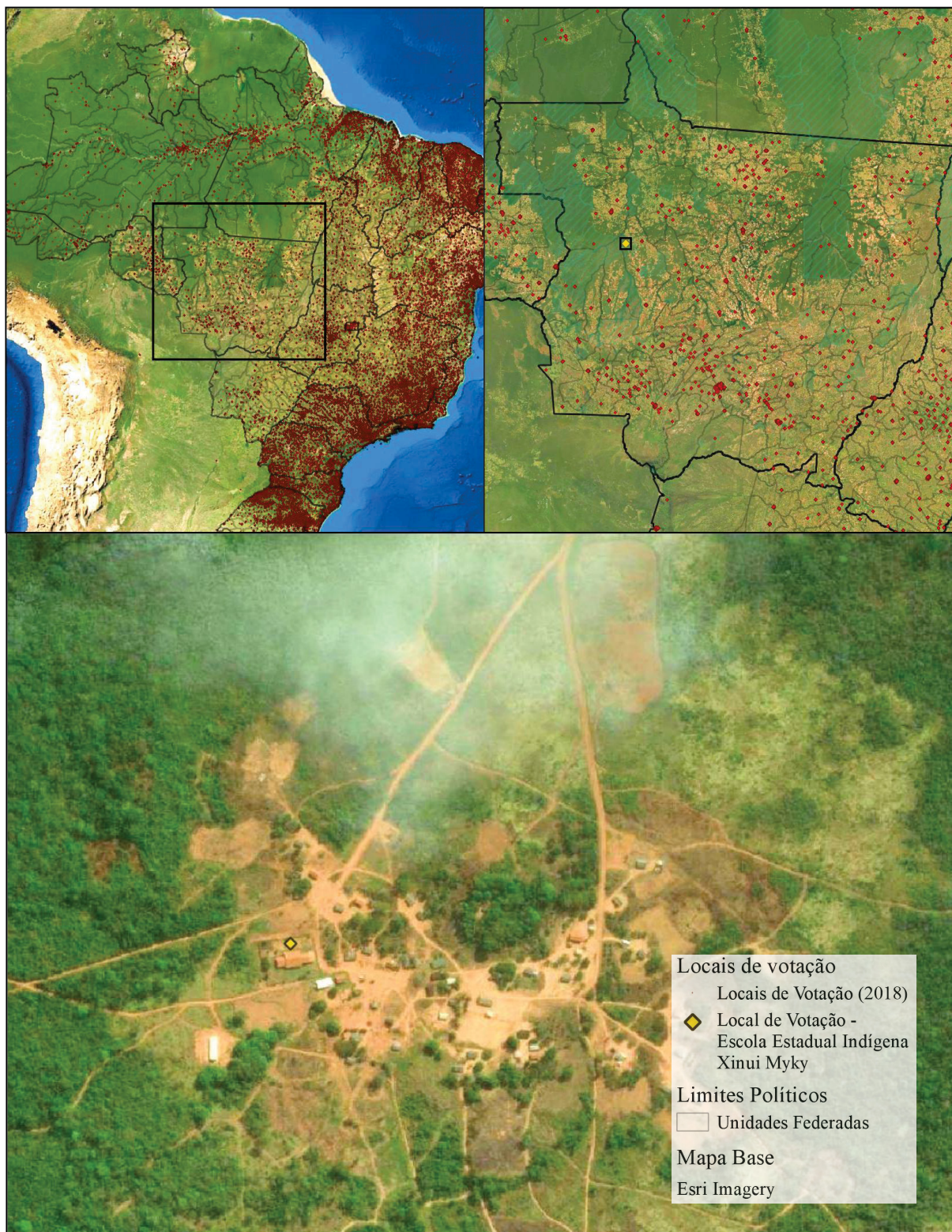


Figura 4.30: Locais de Votação, 2018, Brasil, Mato Grosso, Aldeia Japuira  
Fonte: O autor (2020).

Os estudos pós-estruturalistas na Geografia Humana, de maneira mais ampla, são baseados em métodos qualitativos. No caso específico da GE os pesquisadores adotaram métodos quantitativos para explorar as mesmas ideias (Leib e Quinton, 2011, p.20). Na análise ilustrada aqui, com o indicador local de associação espacial de votos, interpretei as zonas de quebra brusca, ou os *outliers* espaciais, como "*vizinhos indesejados*". Essa é uma dinâmica espacial documentada na literatura como uma espécie de incômodo local por coexistência de espacialidades distintas entre os moradores da cidade e moradores de terras indígenas (da Silva e Helfenstein, 2019). Essas conclusões, até onde consta, são feitas exclusivamente a partir de análises qualitativas (e.g. no caso de da Silva e Helfenstein (2019) em que o material analisado são textos e desenhos de alunos indígenas).

De fato, os métodos de análise quantitativos dominam as explorações da GE (Johnston e Pattie, 2006; Leib e Quinton, 2011). O que é vantagem já que sabemos que a metodologia qualitativa tem pouco espaço na CPBR (o que, certamente, responde em parte pela ausência de ideias pós-estruturalistas e de identidade). O método que illustrei é um caminho para explorar as temáticas de forma quantitativa e diminuir a lacuna entre a produção brasileira e a internacional.

#### 4.5 CONSIDERAÇÕES FINAIS

Neste capítulo desenvolvi uma operacionalização original de análise espacial de voto no Brasil (P-DL). As definições de *OE* e *W* foram pensadas considerando aprimoramentos metodológicos dos modelos atuais — em especial, a relação por rotas transitáveis, o fator local de ponderação e *OE* pontual. Comparei esse modelo com a operacionalização mais comum na área (A-Q1), em busca de resposta para as seguintes perguntas:

*Perguntas 3: A diferença de concordância entre métodos é sistemática? Como a diferença se manifesta no mapa?*

Para formular a resposta utilizei o teste de Bland-Altman e VSPC de mapas de autocorrelação local. O resumo dos resultados é:

*Resposta 3:* Os métodos não concordam e há indicação de subestimação sistemática dos padrões espaciais de voto dos deputados. É possível notar, ao menos, três divergências nos mapas: apagamento da diversidade nas franjas municipais, padrões

anulados por unidades de sinais opostos e simplificações de dinâmicas regionais complexas.

Para ilustrar o potencial do método proposto apresentei uma análise exploratória das bases eleitorais de uma candidatura indígena. Essa temática foi escolhida por aproximar a GE da Geografia Humana, uma lacuna da produção brasileira. A pergunta desenvolvida para a demonstração foi:

*Pergunta 4: A estrutura espacial da votação da candidatura indígena está relacionada à distribuição de Terras Indígenas?*

Analisei a estrutura da votação do candidato Ivan Kaingang no pleito de 2018 no Paraná. Utilizei VSPC de mapas de autocorrelação local como método de exploração. Os resultados indicam:

*Resposta 4: A votação da candidatura indígena apresenta padrão espacial significativo e a distribuição das bases eleitorais está relacionada a *algumas* Terras Indígenas. Os resultados também mostram uma estrutura espacial com grande presença de zonas de transição brusca que podem evidenciar uma relação de "vizinhos indesejados" nessas áreas.*

## 5 CONCLUSÃO

Apontar caminhos de desenvolvimento para a GE no Brasil foi o foco desta pesquisa. Mostrei que precisamos de reflexão mais aprofundada sobre as operacionalizações de análise espacial do voto, principalmente quanto à definição de relações espaciais. Constatei que a maior lacuna da área é em relação à falta de conexão com a Geografia Humana moderna. Desenvolvi, comparei e illustrei um modelo original de análise espacial do voto.

### 5.1 SUMÁRIO DE RESULTADOS

A seguir, resumo os resultados da análise a partir das perguntas de pesquisa em que este estudo foi baseado.

*Pergunta de pesquisa 1:* Como podemos aprimorar a operacionalização dos componentes principais da análise espacial dos votos no Brasil?

- A principal limitação das operacionalizações dos modelos de análise espacial da GE brasileira é em relação as *W*. As justificativas são escassas e pouco trabalhadas. Podemos aprimorar o modelo com maior atenção aos pressupostos teóricos e a superfície modelada. Quanto aos *OE*, os padrões intramunicipais podem revelar padrões desconhecidos, principalmente a partir de *LV*, sem limites de área pré-estabelecidos.

*Pergunta de pesquisa 2:* Quais os principais desalinhamentos entre a produção da Geografia Eleitoral brasileira e as tendências internacionais?

- O maior desalinhamento é a ausência de conexão com a Geografia Humana e, mais especificamente, com as pesquisas com abordagens pós-estruturalistas e de identidade. Um desalinhamento secundário é o de explorações de novos fenômenos espaciais como a gentrificação e as novas direitas e extremas-direitas.

*Perguntas 3:* A diferença de concordância entre métodos é sistemática? Como a diferença se manifesta no mapa?

- Os métodos não concordam e há indicação de subestimação sistemática dos padrões espaciais de voto dos deputados. É possível notar, ao menos, três

divergências nos mapas: apagamento da diversidade nas franjas municipais, padrões anulados por unidades de sinais opostos e simplificações de dinâmicas regionais complexas.

*Pergunta 4: A estrutura espacial da votação da candidatura indígena está relacionada à distribuição de Terras Indígenas?*

- A votação da candidatura indígena apresenta padrão espacial significativo e a distribuição das bases eleitorais está relacionada a *algumas* Terras Indígenas. Os resultados também mostram uma estrutura espacial com grande presença de zonas de transição brusca que podem evidenciar uma relação de "vizinhos indesejados" nessas áreas.

## 5.2 POSSÍVEIS IMPLICAÇÕES

Quais são as possíveis consequências das contribuições desta tese? Nesta seção, esboço alguns encadeamentos possíveis dos achados.

- Possíveis implicações para a **disponibilidade e qualidade de dados eleitorais**:

O trabalho de Ames (2003) é o primeiro, até onde consta, a utilizar estatística espacial para explorar os dados eleitorais brasileiros. Em 2003, devido a falta de dados digitalizados, a criação da matriz de conexão era uma tarefa muito complexa. Os mapas eleitorais foram construídos com uma mesa de digitalização e Autocad, a partir de mapas rodoviários. Para o trabalho, o autor contou com o auxílio de Paul Sampson, da Universidade de Washington, que desenvolveu um *software* para criar as matrizes. E, ainda antes disso, as cartografias eleitorais com Zonas Eleitorais realizadas por Fleischer (1976) foram totalmente feitas à mão. Ou seja, uma barreira enorme para a utilização da metodologia. De lá pra cá, essa barreira vem sendo diminuída e mais pesquisadores adotam o método. Mas os dados ainda tem muito o que melhorar.

Cientistas são bem familiarizados com o princípio de Pareto do trabalho com dados. O nosso tempo em uma pesquisa é dividido em 80% para coleta, limpeza, validação, correção e transformação de dados e apenas 20% na análise de fato (Snyder, 2019). Um problema de analisar dados de resultados eleitorais no Brasil é que esses 80% são mais concentrados em limpeza e correção do que em modelagem estatística. Por exemplo, os dados de coordenadas dos LV, essenciais para a proposta desenvolvida aqui, somam um



total de 49,38% ausências. Dos 96.259 LV totais na base de dados oficial do TSE, para o pleito de 2018, 47.536 não possuem coordenadas geográficas.

Trabalhar individualmente com as ausências dos dados de hoje é um limitador do potencial de desenvolvimento dos modelos e das análises. Certamente, se os meses dedicados a esse problema para a elaboração desta tese fossem dedicados ao desenvolvimento de modelo de estatística espacial as propostas seriam mais sofisticadas.

Uma possível implicação da divulgação do potencial, das lacunas e limitações da GE no Brasil, e mais especificamente, da indicação do uso dos dados de LV é aumentar a demanda por essas informações. A Assessoria de Gestão Estratégica do TSE tem um canal aberto para sugestões ([estatistica@tse.jus.br](mailto:estatistica@tse.jus.br)) e podem ser movimentados por essas solicitações.

- Possíveis implicações para **campanhas eleitorais**:

A Geografia Eleitoral pode fornecer informações relevantes para as estratégias de campanha eleitoral (Terron, 2012), principalmente para os pleitos proporcionais. O distrito eleitoral das disputas para deputado é o estado. Não é viável, nem possível, que o candidato considere todo o território para distribuir as ações da campanha. Uma estratégia baseada em uma análise espacial de votos mais adequada para a realidade pode aumentar a efetividade da campanha. Meu modelo pode ser usado como base para ferramentas como o *Densify Sampling Network* e, assim, determinar localizações de forma otimizada para ações de campanha.

- Possíveis implicações para a **conexão eleitoral**:

O resultado da comparação entre métodos indica uma subestimação sistemática dos padrões espaciais de voto pelo modelo a partir de municípios. Se pesquisas adicionais aumentarem a validade desta abordagem, isso pode significar que a literatura subestima sistematicamente os padrões espaciais de voto dos deputados.

Não é segredo que uma das tarefas que mais ocupa os deputados brasileiros é conseguir recursos para determinadas localidades. Se testemunhos devem ser levados em conta, nada menos que metade deles considera a atividade essencial para suas eleições. Quem entra em seus gabinetes dificilmente nota outra coisa. Nas paredes é possível ver mapas cravejados de alfinetes indicando seus municípios de atuação (Meireles, 2019, p. 137).

O que a maior parte dessa literatura não tem notado na localização do alfinete é que, quando uma emenda orçamentária é destinada para um município para, e.g., a construção de uma estrada, esse é um bem que não pode ser distribuído uniformemente no espaço. Essa estrada vai ser construída em um ponto específico do município, em detrimento de outro, em uma decisão que é, sobretudo, política. E nessa decisão o prefeito não é a única voz. Esse é um fenômeno de dinâmicas de poder complexas que se manifestam espacialmente de forma multi-escalar através de, e.g., comunidades, bairros, associações, representantes distritais, donos de terra, todos competindo pelo bem para sua região.

Discussões semelhantes podem ser feitas para, e.g., o estudo de competição intrapartidária e comportamento eleitoral. Vinculações, alianças, campanhas eleitorais não se distribuem uniformemente no município, assim como as votações não se distribuem aleatoriamente dentro dos municípios. Os padrões espaciais se repetem em escalas intramunicipais, nos distritos administrativos (Zolnerkevic, 2019), nos bairros (Nicolau e Terron, 2012) e locais de votação (do Nascimento, 2017; Faganello, 2017; Gonçalves, 2016). Por isso, a recomendação geral é explorar os padrões intramunicipais.

- Possíveis implicações para os  **cursos da disciplina**:

Em 1991 uma *força tarefa* de representantes da Ciência Política Norte-Americana foi montada para produzir um relatório com o objetivo de desenvolver e fortalecer o ensino dos/as graduandos/as na área. Esse grupo, encabeçado pela *American Political Science Association*, elaborou uma série de recomendações para maximizar a capacidade dos/as estudantes de analisar e interpretar o significado e a dinâmica dos eventos políticos e processos governamentais. A recomendação de número três diz:

That departments reexamine their conceptions of "minor field" and "cognate disciplines," and seek to insure that (1) students' minor-field courses do in fact contribute to their political understanding and analytic competence, and (2) important knowledge from other disciplines is not ignored. In particular, they should: [...] encourage instructors to incorporate in each of their courses the teaching of whatever facts and principles of geography are essential to mastery of the course material (Wahlke, 2020, p.53).

Mais recentemente, Whitten (2016) publicou a seguinte afirmação, no periódico *Political Science Research and Methods*, o *journal* oficial da *European Political Science*

*Association*, um dos mais relevantes da área: teorias espaciais e modelos de estatística espacial são tecnicamente complexos e irão ocupar papel central em todas as linhas de pesquisa da Ciência Política nos próximos anos. Evidências de que a área reconhece que a Ciência Política tem muito a aprender com a Geografia Eleitoral.

No Brasil, é preciso reconhecer a iniciativa do *Grupo de investigación en Análisis Espacial en América Latina* (Espacio Alacip), que promove oficinas gratuitas de análise espacial com foco em estudos sociopolíticos. Entretanto, de forma geral, a oferta de disciplinas específicas de teorias e métodos da GE nos programas de Ciência Política é exceção. O convite feito aqui, para uma reflexão mais aprofundada sobre as operacionalizações dos modelos espaciais de voto no Brasil, pode estimular a entrada de novos pesquisadores na área.

### 5.3 LIMITAÇÕES DO ESTUDO

Em tempo, cabe atentar para algumas limitações da pesquisa:

- Capítulo 3:
  - O material analisado pode ser enriquecido com outras bases de dados, como a Scopus, ou por método de bola de neve, buscando outros artigos a partir das citações da amostra. Outro ponto que reconheço é que, como resultado de decisões metodológicas, o trabalho ignora parte importante da literatura. A amostra não inclui livros, e por isso, trabalhos clássicos como Ames (2003) e de Carvalho (2003) não participaram da análise. Assim como, a dissertação do Faganello (2017), que trabalha com fenômeno de representação da direita e preenche uma das lacunas citadas na revisão;
  - As palavras-chave são, em parte, subjetivas (Callon et al., 1983), e podem ser consideradas *proxies* inadequadas para representar a produção da área;
  - Os dados ainda possuem muito para serem explorados. A produção pode ser analisada com diversos outros filtros, e.g., considerando apenas os países da América Latina, ou isolando apenas a Geografia e a Ciência Política. Não encontrei significado (epistemológico, histórico e/ou político) para alguns dos agrupamentos, principalmente entre países. Essas conexões podem ser exploradas com maior detalhe em pesquisas futuras;
  - Não trato de questões de gênero, raça e identidade das autorias da produção da área. Há evidências de que a produção da área é dominada por autores:

de 224 artigos da GE publicados entre 1990 e 2007 apenas cinco (ou 3,9%) foram escritos por mulheres na primeira autoria. Na minha pesquisa não encontrei nenhuma referência sobre raça e o campo da Geografia Eleitoral.

- Capítulo 4:

- O modelo só foi aplicado em um estado e em uma eleição. Entretanto, mesmo que os resultados tratem apenas do caso do Paraná, existe a possibilidade de generalização simples do experimento. O que foi desenhado para o Paraná pode ser feito para qualquer unidade da federação. A ideia por trás do desenvolvimento do modelo pode ser facilmente replicada. A discussão realizada aqui, utilizando o estado do Paraná, pode ser aproveitada para pensar nas outras unidades federativas. Ao tratar especificamente da matriz de vizinhança, o padrão de alta concentração em aglomerados urbanos e longas distâncias entre LV no interior se repete pelo território brasileiro. O método proposto pode servir de base para a área lidar de maneira mais apropriada com a distribuição multimodal e de grande amplitude nas distâncias entre LV;
- A base de coordenadas geográficas não passou por validação rigorosa. Esses dados possuem um erro desconhecido. A base pode ser validada com outros métodos, e.g., por amostra estratificada;
- Não existem dados sócio-econômicos na mesma simetria espacial de Locais de Votação. Na prática, as análises de GE são multivariadas e não univariadas, como o caso de bases eleitorais indígenas ilustrado aqui. O problema, já documentado na literatura, é que os conjuntos de dados eleitorais e socio-econômicos que desejamos combinar não são da mesma escala (Rodrigues-Silveira, 2013);
- Pontos não são boas representações visuais dos fenômenos;
- LV são *OE* homogêneos;
- A solução proposta para a definição de peso local é simples. Uma solução mais sofisticada pode definir o fator local com algoritmo não supervisionado como o *local outlier factor* (LOF) (Breunig et al., 2000).
- Distância não é o mecanismo causal da relação espacial prevista nas teorias. Estou a par das críticas de Neumayer e Plümper (2016) sobre *W* definida por

distância não representar o efeito causal da dependência espacial. Distância não é o ideal, assim como nenhuma proxy é ideal. Oxalá tivéssemos acesso a dados que descrevem as conexões de aprendizado, interação, comunicação de cada ator para definir  $W$ . E.g. resultados eleitorais são espacialmente autocorrelacionados não porque municípios contíguos influenciam-se reciprocamente, mas sim porque municípios próximos são influenciados por fatores comuns (como campanhas eleitorais) (Marzagão, 2013, p. 281). Não há dados disponíveis sobre a localização das campanhas eleitorais.

#### 5.4 RECOMENDAÇÕES PARA PESQUISAS FUTURAS

*Perhaps the geographic world is better seen as continuous, with reporting zone boundaries somewhat arbitrarily imposed on it* (Goodchild e Janelle, 2010, p. 2).

O modelo desenvolvido não é final. Não existe modelo final, não há uma forma normatizada de fazer análise espacial. Mas ele representa um avanço na direção de operacionalizar a realidade com maior precisão. Tornar o modelo mais robusto com interpolações espaciais parece ser a evolução natural das ideias apresentadas aqui.

A interpolação espacial, transformação de dados espaciais discretos em escalares, resolve uma série das limitações do modelo desenvolvido. Os pontos de LV são de uso exclusivo do TSE, para os quais só existem informações eleitorais. Essa incompatibilidade de níveis é tratada pelas pesquisas com uma agregação em comum entre os dados, e.g., LV e setores censitários agregados em bairros (Rodrigues-Silveira, 2013). Essa não é a única solução e, pela proposta de Tobler (1989), nem a melhor. Métodos de interpolação, como o *Pycnophylactic*, para dados de área, ou *Kriging*, para pontos, são soluções mais robustas para o problema.

As explorações da GE são dominadas por objetos discretos, geralmente áreas, simples e homogêneas. Essas inferências se limitam a áreas definidas, muitas vezes, arbitrariamente. Os modelos discretos são usados, principalmente, por razões práticas, como a disponibilidade de dados (Câmara et al, 2004, p. 19). Mas, um único objeto espacial homogêneo é simplista para representar entidades geográficas, na prática, precisamos de classes mais complexas de dados (Câmara et al, 2000, p. 20).

Contágio e difusão são conceitos fundamentais para as teorias da GE. Esses fenômenos são melhor representados em estruturas contínuas. No polígono fechado o objeto é considerado com total homogeneidade interna e mudança brusca nas fronteiras (Câmara et al, 2004). Essa não é a premissa de um efeito de contágio por fluxos locais

de informações e pessoas. Mas, nas palavras de Tobler (1989, p.1): "*It is not the areal units that are to blame*". Na visão do autor, o inadequado é o método, não o dado. E a solução proposta é o desenvolvimento e aplicação de modelos de interpolação.

Antes de finalizar, vale enfatizar aqui que não há uma forma canônica, correta, de se fazer pesquisas. Há uma série de decisões metodológicas que dependem diretamente da pergunta, do objeto, dos dados disponíveis. Por outro lado, é preciso reconhecer que há decisões que representam melhor a realidade empírica, enquanto outras são documentadamente problemáticas para as análises — o caso de  $W$  com definições topológicas e peso único me parecem exemplos disso.

## REFERÊNCIAS

- Agnew, J. (1990). *From political methodology to geographical social theory? A critical review of electoral geography, 1960–87*. Routledge.
- Agnew, J. (1996). Mapping politics: how context counts in electoral geography. *Political Geography*, 15.
- Ames, B. (2003). *Os entraves da democracia no Brasil*. FGV.
- Anselin, L. e Bera, A. K. (1998). *Spatial Dependence in linear Regression Models with an Introduction to Spatial Econometrics*. CRC Press.
- Avelar, L. e Walter, M. I. M. T. (2008). Lentas mudanças: o voto e a política tradicional. *Opinião Pública*, 14.
- Begu, E. (2007). Elections in a spatial context: a case study of albanian parliamentary elections, 1991-2005.
- Bertus, Z. (2017). Investigating the background of radical right-wing mobilization in hungary with regional statistical methods. *Regional Statistics*.
- Blay, S. K. N. (1988). Self-determination and the crisis in new caledonia: The search for a legitimate self. *Asian Survey*, 28.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R. e Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008.
- Borges, A., de Paula, C. e da Nóbrega Silva, A. (2016). Eleições legislativas e geografia do voto em contexto de preponderância do executivo. *Revista de Sociologia e Política*, 24.
- Borges, A., Filho, A. S. e Rocha, C. (2011). La conexión subnacional: geografía del voto y coaliciones electorales en el federalismo brasileño. *Revista SAAP*, 5.
- Breunig, M. M., Kriegel, H.-P., Ng, R. T. e Sander, J. (2000). *Lof*. ACM Press.

- Callon, M., Jean-Pierre Courtial, W. A. T. e Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, 22:191–235.
- Cazzolato, J. D. (2005). Os bairros como instância territorial local - contribuição metodológica para o caso de são paulo. Dissertação de Mestrado, Faculdade de Filosofia, Letras e Ciências Humanas - Universidade de São Paulo.
- Cervi, E. U. (2009). Produção legislativa e conexão eleitoral na assembléia legislativa do estado do paran . *Revista de Sociologia e Pol tica*, 17.
- Clem, R. S. (2011). From the arab street to the silk road: Implications of the unrest in north africa for the central asian states. *Eurasian Geography and Economics*, 52.
- Codato, A., Lobato, T. e Castro, A. O. (2017). Vamos lutar, parentes!" as candidaturas ind genas nas elei es de 2014 no brasil. *Revista Brasileira de Ci ncias Sociais*, 32.
- da Silva, G. P. (2014). Para al m de s o paulo: Medindo a concentra o eleitoral em todo o brasil.
- da Silva, S. R. e Helfenstein, A. M. (2019). Pr ticas espaciais de jovens ind genas no cotidiano da cidade de dourados (ms). *Confins*.
- Darmofal, D. (2006). The political geography of macro-level turnout in american political development. *Political Geography*, 25.
- de Carvalho, N. R. (2003). *E no in cio eram as bases: geografia pol tica do voto e comportamento legislativo no Brasil*. Revan.
- de  vila, R. P. e Monasterio, L. (2009). o maup e a an lise espacial: um estudo de caso para o rio grande do sul (1991-2000). *An lise Econ mica*, 26.
- Delaney, D. e Leitner, H. (1997). The political construction of scale. *Political Geography*, 16.
- do Nascimento, C. C. (2017). Igreja como partido: capacidade de coordena o eleitoral da igreja universal do reino de deus.
- do Socorro Braga, M. e Zolnerkevic, A. (2020). Padr es de vota o no tempo e no espa o: classificando as elei es presidenciais brasileiras. *Opini o P blica*, 26.



- Druck, S., Carvalho, M. S., Câmara, G. e Monteiro, A. V. M. (2004). *Análise Espacial de Dados Geográficos*. EMBRAPA.
- Druckman, J. N., Kuklinski, J. H. e Sigelman, L. (2009). The unmet potential of interdisciplinary research: Political psychological approaches to voting and public opinion. *Political Behavior*, 31.
- Elazar, D. J. (1999). Political science, geography, and the spatial dimension of politics. *Political Geography*.
- Ethington, P. J. e McDaniel, J. A. (2007). Political places and institutional spaces: The intersection of political science and political geography. *Annual Review of Political Science*, 10.
- Fachini, M. W. (2018). Kakané porã: Um documentário sobre a primeira aldeia urbana do sul do Brasil.
- Faganello, M. A. (2017). O voto na bancada da bala: estudo de geografia eleitoral na cidade de São Paulo (2012/2016).
- Faganello, M. A. e Fernandes, J. L. M. (2018). Migração partidária nos municípios brasileiros (2000-2016). *Revista de Sociologia e Política*, 26.
- Fedozzi, L. e Correa, F. (2015). *Conexões eleitorais e geografia social do voto na Região Metropolitana de Porto Alegre: das debilidades da "polity" aos déficits da "policy"*, páginas 254–294. Letra Capital/Observatório das Metrôpoles.
- Feretti, V. E., Junckes, I. J. e Clemente, A. J. (2018). Ciência política e análise de redes: uma metodologia para o mapeamento de comunidades temáticas. *Guaju*, 4.
- Fleischer, D. (1976). Concentração e dispersão eleitoral: um estudo da distribuição geográfica do voto em Minas Gerais 1966/1974. *Revista Brasileira de Estudos Políticos*.
- Flint, C. (2006). *Introduction to Geopolitics*. Routledge.
- Geara, L. E., Junckes, I. J., Horochovski, R. R. e Clemente, A. J. (2018). O financiamento eleitoral nas comissões permanentes da Câmara dos Deputados: Uma análise do meta-relacionamento entre empresas e parlamentares. *E-Legis*, 11.
- Gonçalves, R. D. (2016). Onde agrego os votos? contribuições à geografia eleitoral aplicada a problemas político-eleitorais brasileiros.

- Gonçalves, R. D. (2020). 99 mil unidades espaciais de análise para explorar problemas político-eleitorais do brasil: Potenciais da geografia eleitoral.
- Goodchild, M. F. e Janelle, D. G. (2010). Toward critical spatial thinking in the social sciences and humanities. *GeoJournal*, 75.
- Gorecki, M. A. e Marsh, M. (2012). Not just 'friends and neighbours': Canvassing, geographic proximity and voter choice. *European Journal of Political Research*, 51.
- Grofman, B. e Handley, L. (1989). Minority population proportion and black and hispanic congressional success in the 1970s and 1980s. *American Politics Quarterly*, 17.
- Grácio, M. C. C. (2020). *Análises relacionais de citação para a identificação de domínios científicos: uma aplicação no campo dos Estudos Métricos da Informação no Brasil*. Cultura Acadêmica.
- Hidalgo, F. D. (2021). Geocoding brazilian polling stations with administrative data sets. [https://github.com/fdhidalgo/geocode\\_br\\_polling\\_stations](https://github.com/fdhidalgo/geocode_br_polling_stations). Acessado em 10/04/2021.
- Horochovski, R. R., Junckes, I. J., Serafim, T. R. e Neto, P. B. (2017). Redes de financiamento eleitoral e grupos de capital na comissão de meio ambiente e desenvolvimento sustentável da câmara dos deputados. *REDES - Revista do Desenvolvimento Regional*, 22:145–169.
- Horochovski, R. R., Junior, A. C., Sampaio, R. C. e Mendonça, R. F. (2019). Democracia deliberativa no brasil: A expansão de um campo concentrado. *Civitas - Revista de Ciências Sociais*, 19:583–604.
- Hoyler, T. (2020). "eu não preciso de waze" – desvendando a estrutura de brokerage em são paulo.
- Jacob, C. R., Hees, D. R., Waniez, P. e Brustlein, V. (2011). A eleição presidencial de 2010 no brasil: continuidade política e estabilidade na geografia eleitoral. *Revista Alceu*, 12:189–229.
- Johnston, R. e Pattie, C. (2006). *Putting Voters in their Place: Geography and Elections in Great Britain*. Oxford University Press.

- Johnston, R. J., Pattie, C. J. e Dorling, D. F. L. (2001). Social locations, spatial locations and voting at the 1997 british general election: evaluating the sources of conservative support. *Political Geography*, 20:85–112.
- Jones, A. e Clark, J. (2020). Political geography and political science: Common territory? *Geopolitics*, 25.
- Jones, M., Jones, R. e Woods, M. (2004). *An Introduction to Political Geography*. Routledge.
- Kainz, W. (2010). *The Mathematics of GIS*. University of Vienna.
- Kim, J., Elliott, E. e Wang, D.-M. (2003). A spatial analysis of county-level outcomes in us presidential elections: 1988–2000. *Electoral Studies*, 22.
- Kim, M. C., Zhu, Y. e Chen, C. (2016). How are they different? a quantitative domain comparison of information visualization and data visualization (2000–2014). *Scientometrics*, 107.
- King, G. (1996). Why context should not count. *Political Geography*, 15:159–164.
- Leib, J. e Quinton, N. (2011). *On the shores of the "moribund backwater"? Trends in electoral geography research since 1980*. Ashgate.
- Lobo, M.-J., Pietriga, E. e Appert, C. (2015). An evaluation of interactive map comparison techniques. ACM.
- Long, J. e Robertson, C. (2018). Comparing spatial patterns. *Geography Compass*, 12.
- Mancosu, M. e Ladini, R. (2020). The red and the black: neo-fascist inheritance in the electoral success of the lega in tuscany, umbria, and marche. *Journal of Modern Italian Studies*, 25.
- Marques, F. M. (2018). The electoral connection in an open-list pr system: Informal constituencies, budget amendments and public policies in rio de janeiro. *Revista Brasileira de Ciências Sociais*, 33.
- Marston, S. A., Jones, J. P. e Woodward, K. (2005). Human geography without scale. *Transactions of the Institute of British Geographers*, 30.

- Marzagão, T. (2013). A dimensão geográfica das eleições brasileiras. *Opinião Pública*, 19.
- Matsakis, P., Wendling, L. e Ni, J. (2010). *A General Approach to the Fuzzy Modeling of Spatial Relationships*.
- McGing, C. F. (2014). Towards a feminist electoral geography. *SSRN Electronic Journal*.
- Meireles, F. (2019). A política distributiva da coalizão.
- Meneguello, R. e Neto, F. B. (2012). Contexto e competição na política paulista. *Dados*, 55.
- Mongeon, P. e Paul-Hus, A. (2016). The journal coverage of web of science and scopus: a comparative analysis. *Scientometrics*, 106.
- Neumayer, E. e Plümpner, T. (2016). *W. Political Science Research and Methods*, 4.
- Nicolau, J. M. e Terron, S. L. (2012). Uma cidade partida? as eleições para prefeito do rio de janeiro em 2008.
- O'Loughlin, J. (2008). *Spatial Analysis in Political Geography*. Basil Blackwell.
- Pattie, C. e Johnston, R. (1999). Context, conversation and conviction: Social networks and voting at the 1992 british general election. *Political Studies*, 47.
- Pattie, C. e Johnston, R. (2000). 'people who talk together vote together': An exploration of contextual effects in great britain. *Annals of the Association of American Geographers*, 90.
- Pindograma (2020). A precise, nation-wide brazilian electoral map. <https://github.com/pindograma/mapa>. Acessado em 10/04/2021.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15.
- Rodrigues-Silveira, R. (2013). Território, escala e voto nas eleições municipais no brasil. *Cadernos Adenauer*, XVI:167–192.
- Rodrigues-Silveira, R., Terron, S. e Sonnleitner, W. (2017). Política, espacialidad y territorialidad. hacia un análisis espacial de los procesos socio-políticos en américa latina. *América Latina Hoy*, 75.

- Russo, L. e Beauguitte, L. (2014). Aggregation level matters: evidence from french electoral data. *Quality & Quantity*, 48.
- Salvatore, J. D. e Ruggeri, A. (2021). Spatial analysis for political scientists. *Italian Political Science Review/Rivista Italiana di Scienza Politica*.
- SEADE (2016). Divisão regional do estado de são paulo evolução histórica 1967 – 2014.
- Silotto, G. (2019). A relevância regional nas estratégias partidárias: evidências das listas de candidatos de são paulo. *Revista de Sociologia e Política*, 27.
- Simoni, S. J. (2017). Política distributiva e competição presidencial no brasil: Programa bolsa-família e a tese do realinhamento eleitoral.
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24:265–269.
- Smith, M., Goodchild, M. e Longley, P. (2015). *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*. The Winchelsea Press, 5 edition.
- Snyder, J. (2019). Data cleansing: An omission from data analytics coursework. *Information Systems Education Journal*, 17:22–29.
- Soares, G. A. D. e Terron, S. L. (2008). Dois lulas: a geografia eleitoral da reeleição (explorando conceitos, métodos e técnicas de análise geoespacial). *Opinião Pública*, 14.
- Soares, L. B. (2018). Idle no more: sobre a mobilização indígena no Canadá. *Revista Brasileira de Ciências Sociais*, 33.
- Sonnleitner, W. (2013). Explorando las dimensiones territoriales del comportamiento político: Reflexiones teórico-metodológicas sobre la geografía electoral, la cartografía exploratoria y los enfoques espaciales del voto. *Estudios Sociológicos*, 31:97–142.
- Sonnleitner, W. (2017). Rastreando las dinámicas territoriales de la fragmentación partidista en México (1991-2015). *América Latina Hoy*.
- Strozzi, F., Colicchia, C., Creazza, A. e Noè, C. (2017). Literature review on the 'smart factory' concept using bibliometric tools. *International Journal of Production Research*, 55.

- Sánchez-Lozano, J. M., Teruel-Solano, J., Soto-Elvira, P. L. e García-Cascales, M. S. (2013). Geographical information systems (gis) and multi-criteria decision making (mcdm) methods for the evaluation of solar farms locations: Case study in south-eastern spain. *Renewable and Sustainable Energy Reviews*, 24.
- Terron, S. (2009). A composição de territórios eleitorais no brasil: Uma análise das votações de lula (1989 - 2006).
- Terron, S. (2012). Geografia eleitoral em foco. *Revista Em Debate*, 4:8–18.
- Terron, S. L. e Soares, G. A. D. (2010). As bases eleitorais de lula e do pt: do distanciamento ao divórcio. *Opinião Pública*, 16.
- Thomaz, P. G., Assad, R. S. e Moreira, L. F. P. (2011). Uso do fator de impacto e do índice h para avaliar pesquisadores e publicações. *Arquivos Brasileiros de Cardiologia*, 96.
- Tobler, W. (1989). *Frame independent spatial analysis*. CRC Press, 1 edition.
- Vieira, J. V. B. (2012). A geografia eleitoral: padrões de votação para deputados federais do rio grande do sul e em pernambuco nas eleições de 2006 e 2010.
- Wacquant, L. (2010). Ressituando a gentrificação: a classe popular, a ciência e o estado na pesquisa urbana recente. *Caderno CRH*, 23.
- Wahlke, J. C. (2020). Liberal learning and the political science major: A report to the profession. *PS: Political Science and Politics*, 24.
- White, A. C. (2015). How ethnicity matters in russian elections: ethnic minorities and support for the united russia party. *Eurasian Geography and Economics*, 56.
- Whitten, G. D. (2016). Spatial models of politics. *Political Science Research and Methods*, 4.
- Yao, J. e Murray, A. T. (2013). Continuous surface representation and approximation: spatial analytical implications. *International Journal of Geographical Information Science*, 27.
- Zavala, R. B. (2012). Génesis de la geografía electoral. *Espacialidades*, 2:80–95.
- Zolnerkevic, A. (2019). Contexto social de vizinhança: percepções políticas na cidade de são paulo. *Revista Brasileira de Ciência Política*.

## **APÊNDICE A – MÉTODOS E PROCEDIMENTOS DO MAPEAMENTO DA GEOGRAFIA ELEITORAL**

Para mapear o estado da arte e identificar lacunas e tendências de pesquisa da GE e, mais especificamente, da posição do Brasil nesse cenário, realizei uma análise bibliométrica e cientométrica da produção. Essa é uma metodologia caracterizada por tratar a literatura a partir de protocolos que tornam a análise transparente e replicável.

Os procedimentos técnicos foram realizados em cinco fases: I) Definição de base de dados; II) Mineração e tratamento de dados; III) Processamento estatístico de dados; IV) Análise de parâmetros bibliométricos; e V) Análise de redes bibliométricas. A Figura A.1 sumariza os procedimentos.

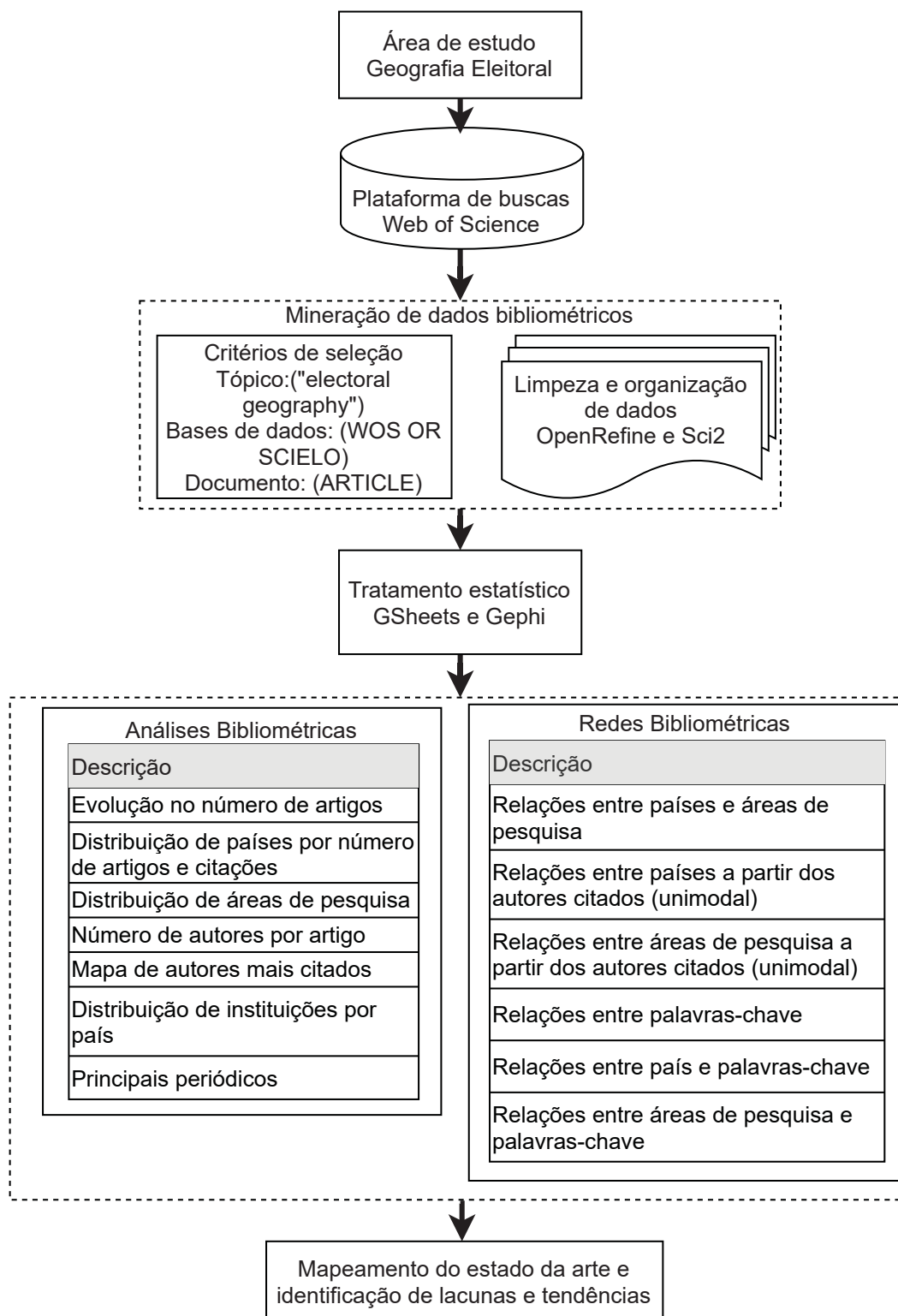


Figura A.1: Diagrama das etapas do método de mapeamento do estado da arte e identificação de lacunas e tendências de pesquisa

Fonte: O autor (2021).

Nota: Desenvolvido a partir do modelo proposto por Oliveira et al, 2019.



### A.0.1 Definição de base de dados

Os dados analisados aqui tem como fonte a plataforma online *Web of Science* (WoS). Essa é uma base de dados que agrega mais de 23 mil periódicos técnicos e científicos, com mais de 40 milhões de arquivos que datam desde 1900. É citada como uma das principais fontes de dados para explorações bibliométricas, sobretudo para pesquisas sociais e multidisciplinares (Harinayana, 2015. p. 6) — como é o caso da GE.

Das bases de dados que compõem a WoS foram selecionadas a SciElo Citation Index e toda a Coleção Principal (Composta por Social Sciences Citation Index, Science Citation Index Expanded, Art Humanities Citation Index e Emerging Sources Citation Index). Essas bases indexam um grande conjunto de periódicos e impõem uma série de requisitos rigorosos de qualidade. A SciElo foi escolhida como fonte adicional à coleção principal pois representa um conjunto de bibliotecas de países importantes para o objetivo de uma revisão abrangente (como África do Sul, Argentina, Bolívia, Chile, Colômbia, Costa Rica, Cuba, México, Peru, Uruguai, Venezuela e — o mais relevante para a exploração em tela — o Brasil).

A WoS indexa os periódicos de maior peso na produção científica e é uma fonte usual de análises bibliométricas e cientométricas, mas é preciso reconhecer que existem outras fontes (como a Scopus) e essa escolha não é livre de viés (como a sobrerrepresentação de periódicos em inglês e das Ciências Naturais nesta plataforma). (Mongeon e Paul-Hus, 2016).

#### A.0.1.1 Mineração e tratamento de dados

Os artigos foram requisitados em 25 de maio (Coleção Principal) e 18 de julho de 2020 (SciElo) usando a plataforma online de buscas. Foram incluídos todos os artigos (Documento:[*article*]) em que constava o termo "*electoral geography*" no título, resumo ou palavras-chave. A busca por "Tópico" na plataforma WoS utiliza tanto as palavras-chave definidas pelos autores quanto as dadas pelo sistema automático Keywords Plus. O sistema automático utiliza a frequência de palavras e frases dos títulos das referências de cada artigo para criar novas categorias de indexação da literatura. O principal critério de exclusão foi o tipo de documento, limitado a artigos, por serem o principal meio de comunicação científica. Além desse primeiro filtro, os critérios de inclusão são amplos: sem limitação temporal, pois é do interesse traçar uma linha do tempo da área; sem limitação de área, entendendo que a Geografia Eleitoral é uma área naturalmente interdisciplinar e;

sem limitação de língua, já que é de interesse da exploração um mapa abrangente dos metadados das pesquisas da área.

O termo de busca, "*electoral geography*", foi definido depois de diversos testes e considerações. As primeiras tentativas partiram de uma busca ampla combinando objetos de pesquisa da Ciência Política com identificadores geográficos. A estratégia visava explorar ao máximo a produção científica que associa, implícita ou explicitamente, fenômenos político-eleitorais e localização. O objetivo era identificar as pesquisas de Ciência Política que de alguma forma se utilizam do ferramental teórico, conceitual e metodológico da Geografia Humana e Análise Espacial de forma geral. Esse desenho de pesquisa se mostrou inviável resultando entre 2.679 e 2.686 artigos, um número de casos que sobrerrepresenta o campo de foco desta análise bibliométrica. i.e., o resultado trataria de um mapa maior do que os limites da área de interesse da exploração. As tabelas abaixo resumem os testes de palavras-chave e bases de dados para coleta do material analisado.

Tabela A.1: Testes de construção de corpus analítico por palavras-chave e base de dados - Web of Science, 06/02/2020

Termos de busca	Refino de busca	Resultado
1 TÓPICO: (election*) OR (electoral) OR (voting) OR ("political participation") OR ("political behaviour") OR ("electoral systems") OR ("presidential system") OR ("political parties") OR ("party system")	DOMÍNIOS DE PESQUISA: (SOCIAL SCIENCES ) TIPOS DE DOCUMENTO: (ARTICLE ) Bases de dados= WOS, DIIDW, KJD, RSCI, SCIELO	80.709
2 TÓPICO: (geography) OR (spatial analysis) OR (spatial)	Tempo estipulado=Todos os anos Idioma da pesquisa=Auto	411.857
1 AND 2		2.679
3 TÓPICO: (election*) OR ("electoral") OR (voting*) OR (vote*)	TIPOS DE DOCUMENTO: (ARTICLE) Índices: SCI-EXPANDED, SSCI, A HCI,	78.753
4 TÓPICO: (geography) OR (spatial analysis) OR (spatial)	CPCI-S, CPCI-SSH, ESCI. Tempo estipulado=Todos os anos Idioma da pesquisa=Auto	832.555
3 AND 4		2.686

Fonte: O autor (2021)

Tabela A.2: Testes de construção de corpus analítico por palavras-chave e base de dados - Scopus, 06/02/2020

Termos de busca	Refino de busca	Resultado
TITLE-ABS ( ( "election" OR "electoral" OR "voting patterns" OR "voting behaviour" OR "political participation" OR "political parties" OR "electoral systems" OR "party system" OR "presidential system") ) AND ( "geography" OR "spatial analysis" ) )	( LIMIT-TO ( SRCTYPE , "j" ) ) AND ( LIMIT-TO ( PUBSTAGE , "final" ) ) AND ( LIMIT-TO ( DOCTYPE , "ar" ) ) AND ( LIMIT-TO ( SUBJAREA , "SOCI" ) )	643
TITLE-ABS ( ( "election" OR "electoral" OR voting* OR vote* ) ) AND ( "geography" OR "spatial analysis" ) )		694

Fonte: O autor (2021)

Tabela A.3: Testes de construção de corpus analítico por palavras-chave e base de dados - Web of Science, 25/05/2020

Termos de busca	Refino de busca	Resultado
TÓPICO: ("electoral geography")	TIPOS DE DOCUMENTO: (ARTICLE) Índices: SCI-EXPANDED, SSCI, A HCI, CPCI-S, CPCI-SSH, ESCI. Tempo estipulado=Todos os anos Idioma da pesquisa=Auto	202
TÓPICO: ("electoral geography")	TIPOS DE DOCUMENTO: (ARTICLE) Bases de dados= SCIELO Tempo estipulado=Todos os anos Idioma da pesquisa=Auto	24

Fonte: O autor (2021)

O termo "*electoral geography*" deixa claro os limites do campo tratado e gera um *corpus* bem delimitado para a revisão. Em inglês o termo garante representação dos trabalhos, já que é regra para os artigos indexados pela plataforma um *abstract*, a versão de resumo em inglês. O termo resulta em dois tipos de pesquisas: as que os próprios autores identificam o artigo como pertencente à área em seus títulos, resumos e/ou palavras-chave ou as que são indexadas na área pela alta presença do termo nas referências citadas.

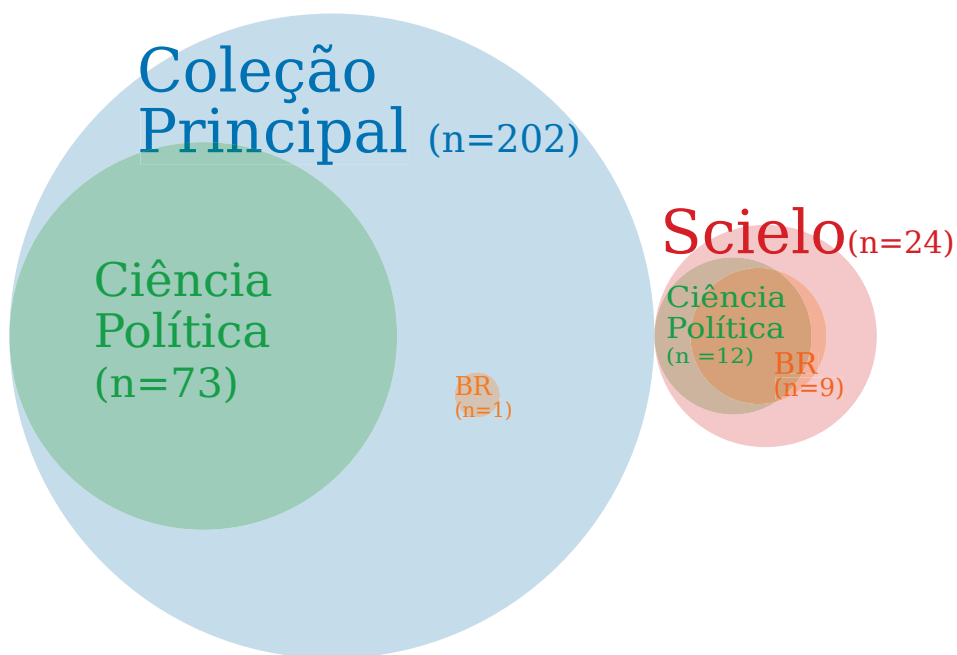


Figura A.2: Diagrama de Venn proporcional da distribuição e interseção dos artigos entre bases de dados (Coleção Principal WoS e SciELO), área (Ciência Política) e país (Brasil)

Fonte: O autor (2021).

O resultado das buscas seguindo esses critérios totalizou 226 artigos, distribuídos conforme apresentado na Figura A.2. A análise destaca essas duas agregações principais, por área de pesquisa e por país, isolando a Ciência Política e o Brasil — focos de interesse do *rationale* da revisão.

Todos os dados passaram por pré-processamento de tratamento de inconsistências. Para garantir a confiabilidade dos testes os dados de nomes dos autores, referências citadas, nomes dos periódicos, países, instituições e palavras-chave foram normalizados, principalmente a partir de análise de cluster com algoritmos *key collision* e *nearest neighbor* com auxílio do *software* OpenRefine (v. 3.3). Mais informações sobre métodos e teorias das funções de agrupamento utilizados pelo OpenRefine disponíveis em <https://github.com/OpenRefine/OpenRefine/wiki/Clustering-In-Depth>. É comum encontrar inconsistências nos dados bibliométricos. No caso da pesquisa em tela alguns exemplos para a variável *Authors* são:

- Discrepância entre letras maiúsculas e minúsculas

SONNLEITNER, W (2 rows)  
Sonnleitner, W (2 rows)

- Erro de encoding

HERNANDEZ-HERNANDEZ, V (1 rows)  
Hernández-Hernández, V (1 rows)

- Mesmo caso com valores inconsistentes

MACALLISTER, I (2 rows)  
MCALLISTER, I (1 rows)

Todas as variáveis processadas passaram por testes de consistência para lidar com os problemas como os indicados acima. Além da limpeza dos dados, a variável *Cited References* foi transformada em *Autores citados*. A variável de maior inconsistência é a *Cited References*, existe uma grande variação em como os mesmos autores são citados. Em uma primeira análise os dados seguem o padrão “sobrenome primeira letra do nome, nome do periódico/livro, edição, páginas, DOI”, como o exemplo: *Agnew J, 1996, POLIT GEOGR, V15, P129, DOI 10.1016/0962-6298(95)00076-3*. Entretanto, as mesmas citações aparecem de diversas formas, como apresentam os exemplos abaixo.

- Mesmo autor, com espaço, sem espaço, com duas letras de nome do meio e com uma letra, com ponto e sem ponto.

- Cliff A D (4 rows)
- CLIFF AD (3 rows)
- Cliff A (2 rows)
- CLIFF A (1 rows)
- Cliff A. (1 rows)
- Cliff A. D. (1 rows)
- Cliff AD (1 rows)

O'Loughlin J (23 rows)  
OLOUGHLIN J (21 rows)  
OLoughlin J (9 rows)  
O'Loughlin J. (6 rows)  
O'LOUGHLIN J. (2 rows)

- Primeiro nome sem abreviação, com ponto e sem ponto.

- Grofman Bernard (3 rows)
- Grofman Bernard. (2 rows)

O primeiro passo para lidar com essas inconsistências foi separar o nome dos autores usando expressões regulares no Google Sheets, usando o exemplo citado acima:

```

1 =REGEXEXTRACT("Agnew J, 1996, POLIT GEOGR, V15, P129, DOI 10.1016/0962-6298",".*?,"")
2 Agnew J, // saída
3
4 Edit > Find and replace > Find , > Replace with > Replace all
5 Agnew J // saída

```

O próximo passo foi normalizar todos os nomes a partir de análise de *cluster* com algoritmos *key collision* e *nearest neighbor*. Foi definido o padrão “SOBRENOME Primeira letra do nome”. Todos os casos com mais de uma presença seguem esse padrão, alguns casos com uma citação também conta a “primeira letra do nome do meio”.

#### A.0.2 Processamento estatístico de dados

Foram utilizadas dez variáveis bibliométricas: autoria, ano de publicação, país, número total de citações, periódico, instituição de vinculação dos autores, departamento de vinculação dos autores, área de pesquisa do periódico e palavras-chave. Também foram incluídas duas variáveis *dummy* como forma de isolar “Ciência Política” e “Brasil” para a análise dos dados.

Os dados foram processados por dois conjuntos de estatísticas. O primeiro, uma abordagem descritiva dos atributos bibliométricos. O segundo, explora as relações da área a partir de redes bibliométricas, elaboradas com o auxílio dos *softwares* Gephi (v. 0.9.2) e Sci2 (v. 1.3).

Na exploração cientométrica com redes bibliométricas, os países, áreas de pesquisa, autores citados e palavras-chave, convertem-se em vértices conectados entre si por arestas. Isso permite a exploração de como a área se articula, ou de outra forma, como esses atributos bibliométricos estão conectados. O foco da exploração é identificar lacunas e tendências de pesquisa da área mostrando a posição do Brasil e da Ciência Política, por isso as unidades de observação agrupam os países e áreas de pesquisa.

Foram aplicadas cinco estatísticas de rede:



- Modularidade: medida da relação de vizinhança na rede, divide os vértices em grupos a partir da relação de proximidade entre eles (mais detalhes sobre o algoritmo Louvain de clusterização utilizado pelo Gephi podem ser encontrados em Blondel et al. (2008));
- Centralidade de Grau: mede a quantidade de vínculos estabelecidos pelo vértice;
- Centralidade de intermediação: mede a quantidade de vezes em que o nó é usado como ponte para outros dois nós dentro da rede;
- Grau de entrada: é dado pela quantidade de arestas que se conectam a um nó (da rede para o nó);
- Grau de saída: mede a quantidade de conexões que um nó estabelece com o resto da rede (do nó para a rede).

A construção da análise de *co-word* é feita pela conexão entre atores (aqui, países e áreas de pesquisa) e palavras-chave. As palavras-chave foram normalizadas e homogeneizadas através da eliminação de palavras sem significados e *stopwords* (para o processamento de linguagem natural considere como palavras vazias as de características funcionais, como “a”, “e”, “o”, “em” e “no”), seguindo a proposta de Feretti et al. (2018). Mas, diferentemente dos autores, optei por não utilizar as formas canônicas e radicais das palavras-chave por considerar a perda de precisão dos termos excessiva. O sentido básico das palavras acaba por encobrir a heterogeneidade dos conceitos. Tomando como exemplo o canônico-radical de *space* que agruparia *spatial* dentro dos contextos: *spatial analysis*, *spatial concentration*, *spatial distribution*, *spatial statistics*, *spatial models*, *spatial clustering*, *spatial heterogeneity*, *socio-spatial inequalities*, entre outros.

Ainda sobre o tratamento das redes bibliométricas, foi realizado um procedimento de conversão multimodal de redes bimodais, formadas entre países e autores citados e áreas de pesquisa e autores citados, em rede unimodais (*dual mode - one mode*). Esse procedimento possibilita a análise da relação de um tipo de nó a partir das conexões compartilhadas com outro tipo de nó. Uma metaleitura que funciona para evidenciar as relações entre países formada pelos autores citados em comum e entre as áreas de pesquisa a partir desses autores. Maiores explicações sobre a aplicação do procedimento podem ser encontradas em Horochovski et al., 2017; Geara et al., 2018

## APÊNDICE B – MAPAS DO INDICADOR LOCAL DE ASSOCIAÇÃO ESPACIAL

B.1 CANDIDATOS COM MAIORES DIFERENÇAS DE SCORE ENTRE MÉTODOS:

*A – Q1 - P – DL*

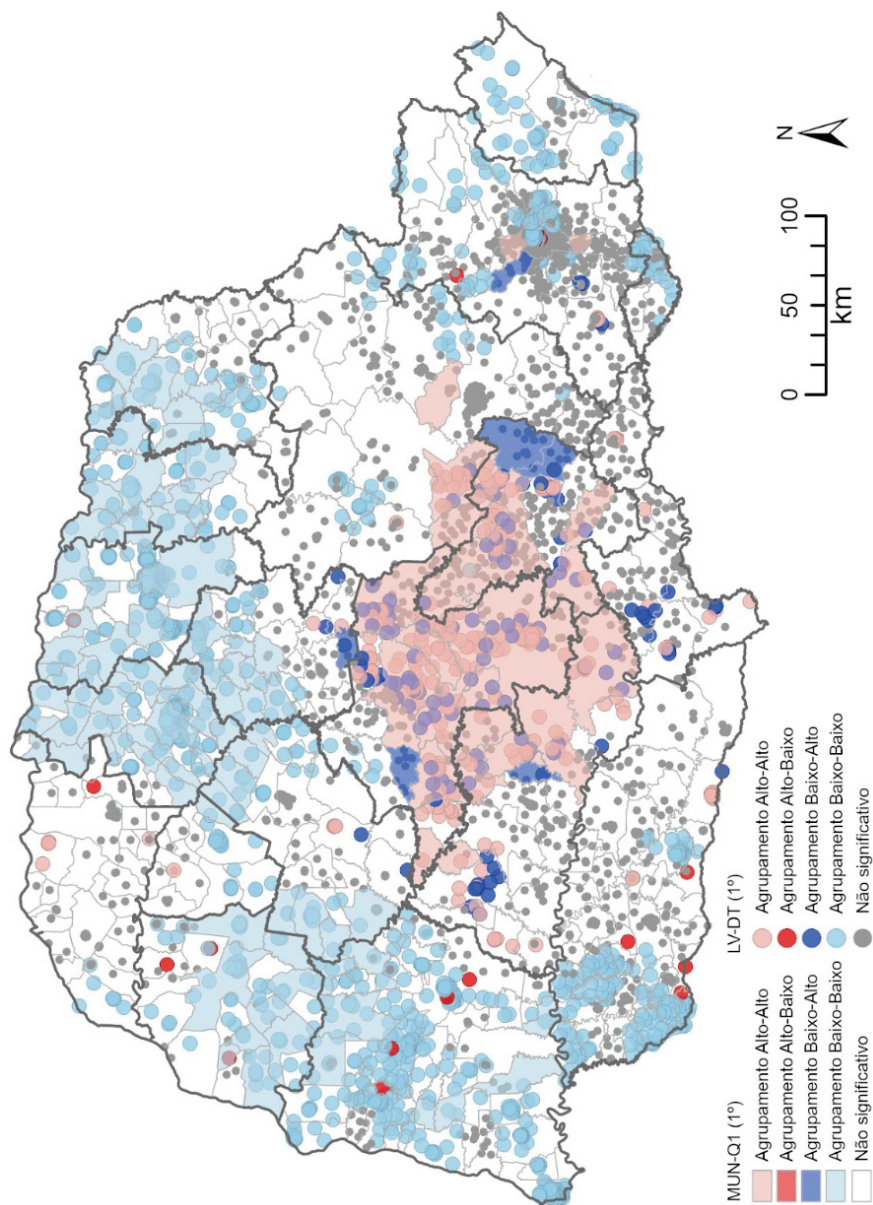


Figura B.1: Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato ARTIDOR JOSÉ DELCI SOARES, 2018, Paraná, Brasil.

Fonte: O autor (2021)

Nota: Candidato 90090. Diferença de I de Moran ( $A - Q1 - P - DL$ ): 0,394 ( $A - Q1 = 0,478$ ;  $P - DL = 0,084$ )

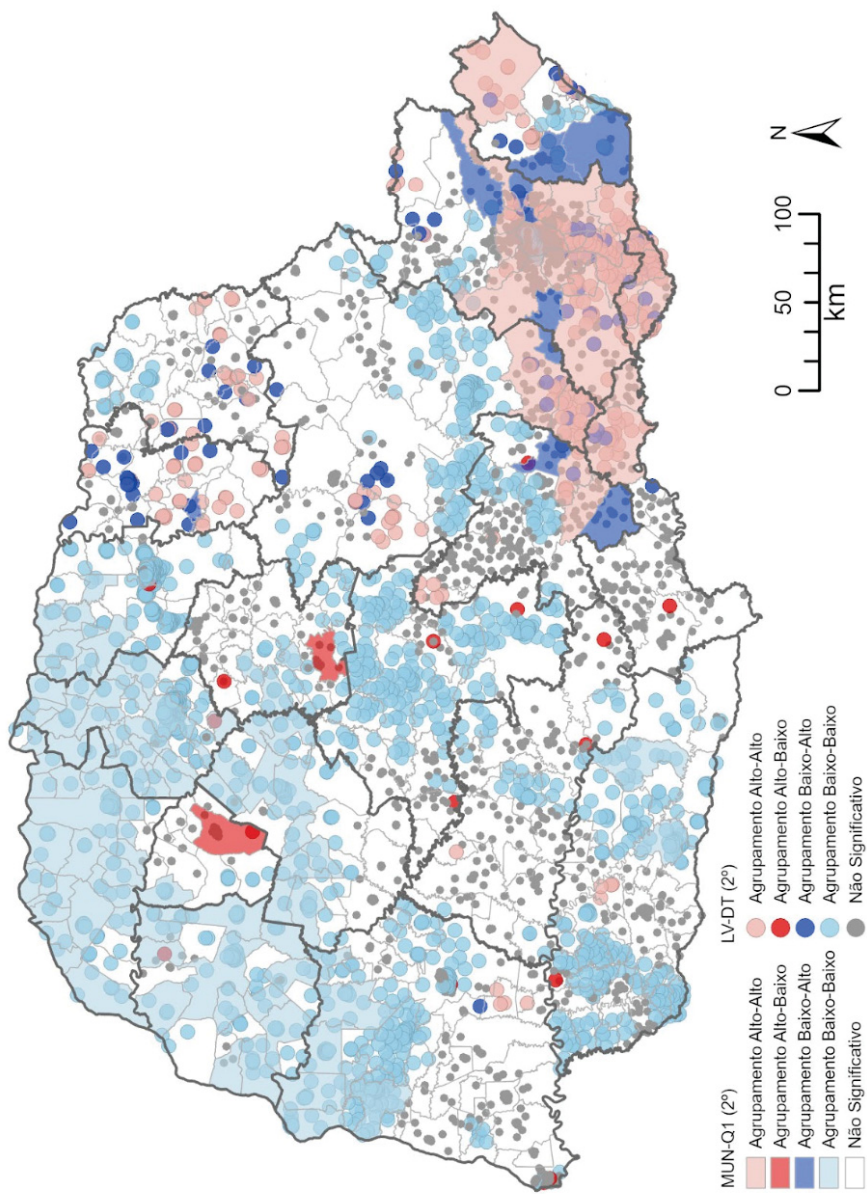


Figura B.2: Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato ANTONIO WANDSCHEER, 2018, Paraná, Brasil.

Fonte: O autor (2021)

Nota: Candidato 9090. Diferença de I de Moran(A - Q1 - P - DL): 0,298 (A - Q1 = 0,498; P - DL = 0,200)

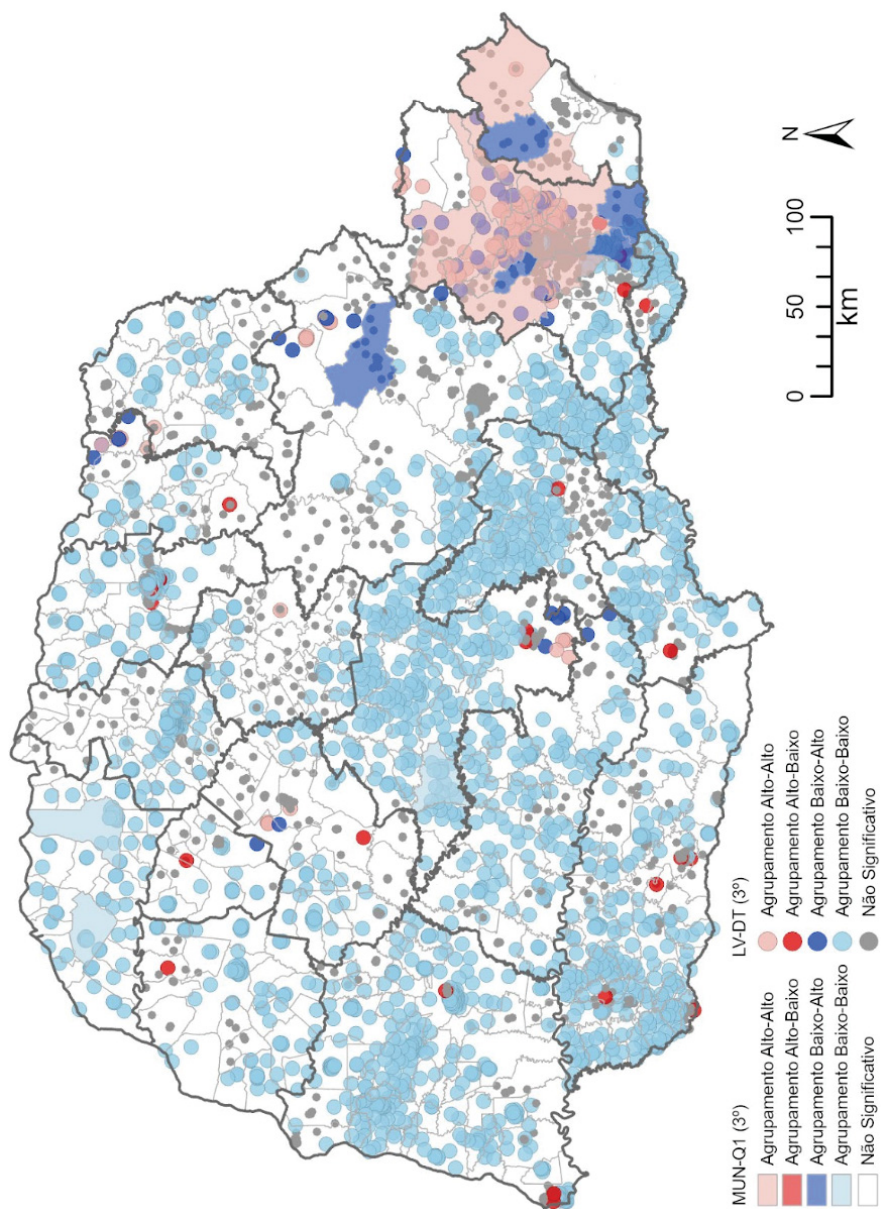


Figura B.3: Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato LUIZ ALSIONI MARTINS DA SILVA, 2018, Paraná, Brasil.

Fonte: O autor (2021)

Nota: Candidato 20222. Diferença de I de Moran: 0,295 (A-Q1 = 0,526; P-DL = 0,231)

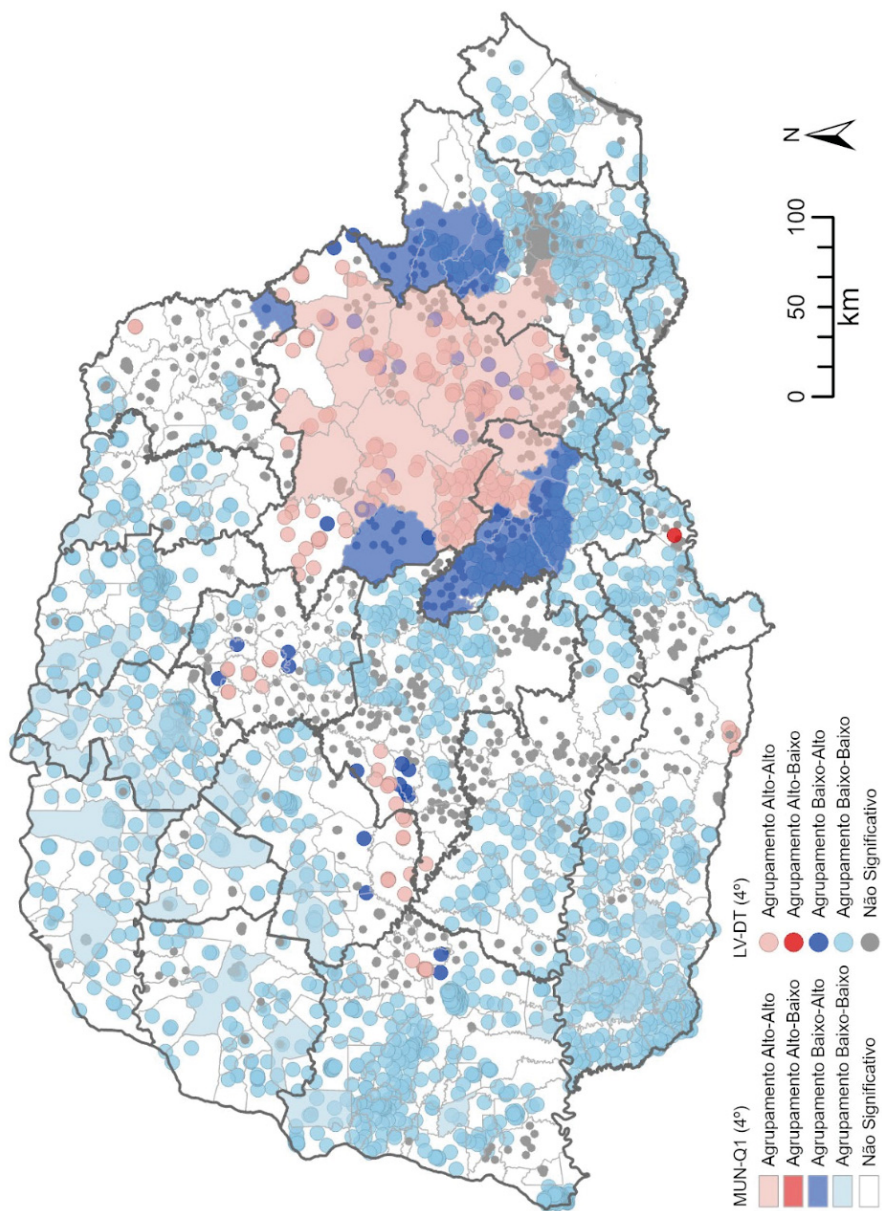


Figura B.4: Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato PLAUTO MIRO GUIMARÃES FILHO, 2018, Paraná, Brasil.

Fonte: O autor (2021)

Nota: Candidato 25110. Diferença de I de Moran: 0,227 (A-Q1 = 0,423; P-DL = 0,195)

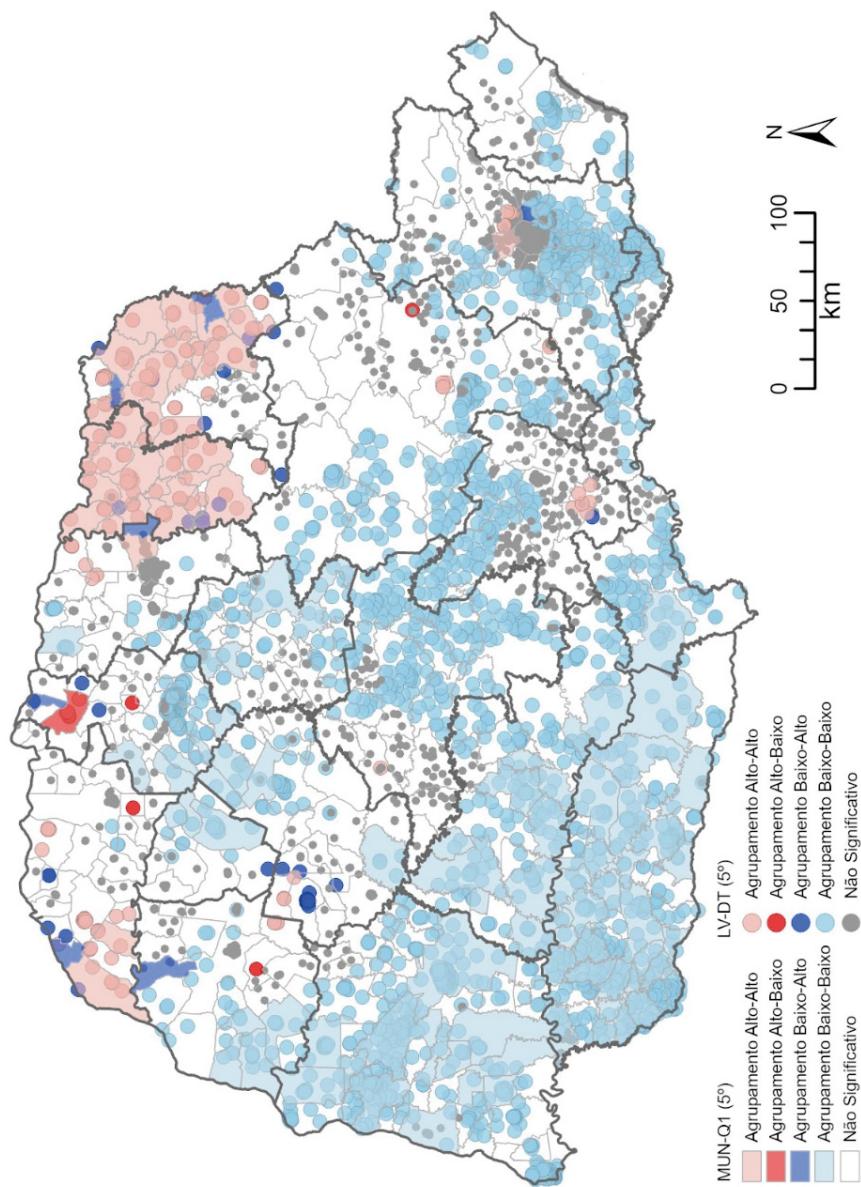


Figura B.5: Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato LUIZ CLAUDIO ROMANELLI, 2018, Paraná, Brasil.

Fonte: O autor (2021)

Nota: Candidato 40678. Diferença de I de Moran: 0,216 (A-Q1 = 0,443; P-DL = 0,227)

## B.2 CANDIDATOS COM MAIORES DIFERENÇAS DE SCORE ENTRE MÉTODOS: P – DL – A – Q1

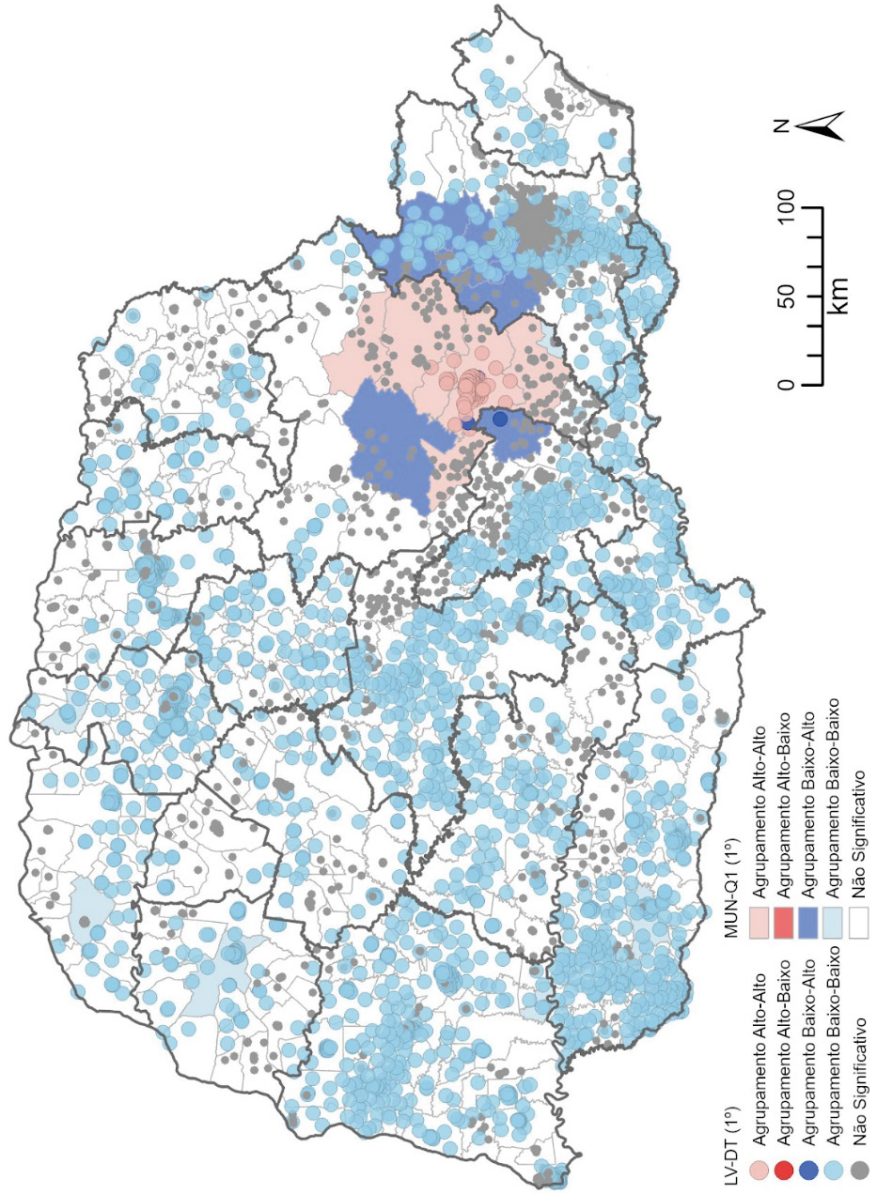


Figura B.6: Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato JOSE CARLOS SAHAGOFF RAAD, 2018, Paraná, Brasil.

Fonte: O autor (2021)

Nota: Candidato 23456. Diferença de I de Moran:  $-0,816$  (A-Q1 =  $0,020$ ; P-DL =  $0,837$ )



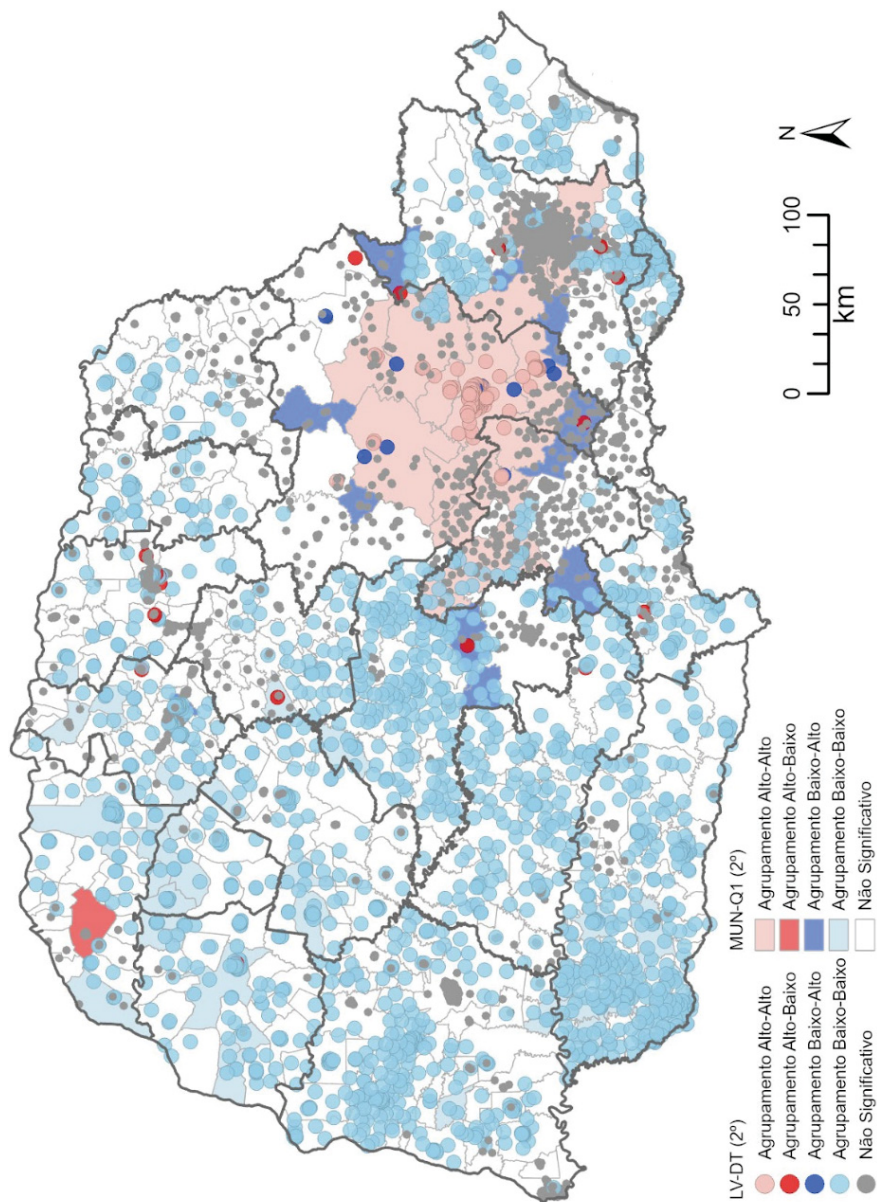


Figura B.7: Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato FELIPE RAMON DOS PASSOS, 2018, Paraná, Brasil.

Fonte: O autor (2021)

Nota: Candidato 4512. Diferença de I de Moran:  $-0,753$  (A-Q1 =  $0,056$ ; P-DL =  $0,809$ )

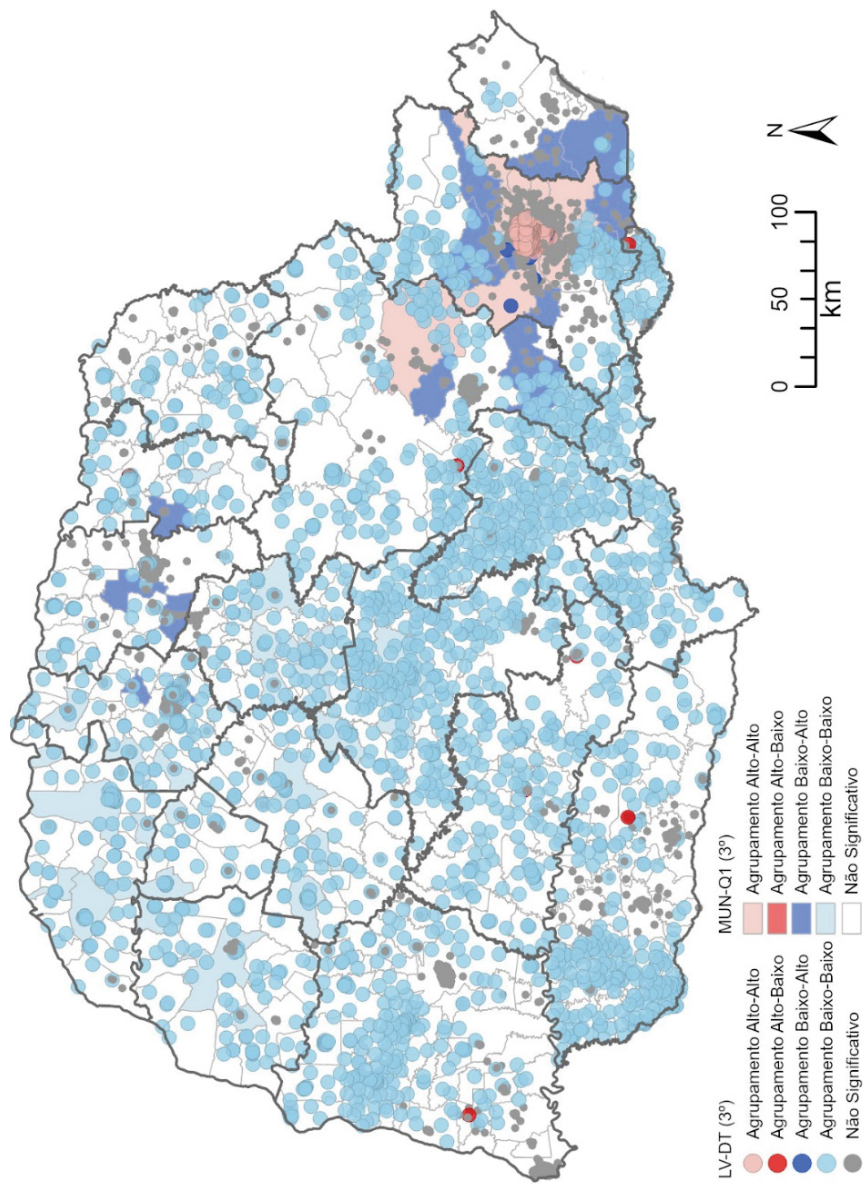


Figura B.8: Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato JORGE GOMES DE OLIVEIRA BRAND, 2018, Paraná, Brasil.

Fonte: O autor (2021)

Nota: Candidato 12108. Diferença de I de Moran: -0,752 (A-Q1 = 0,037; P-DL = 0,790)

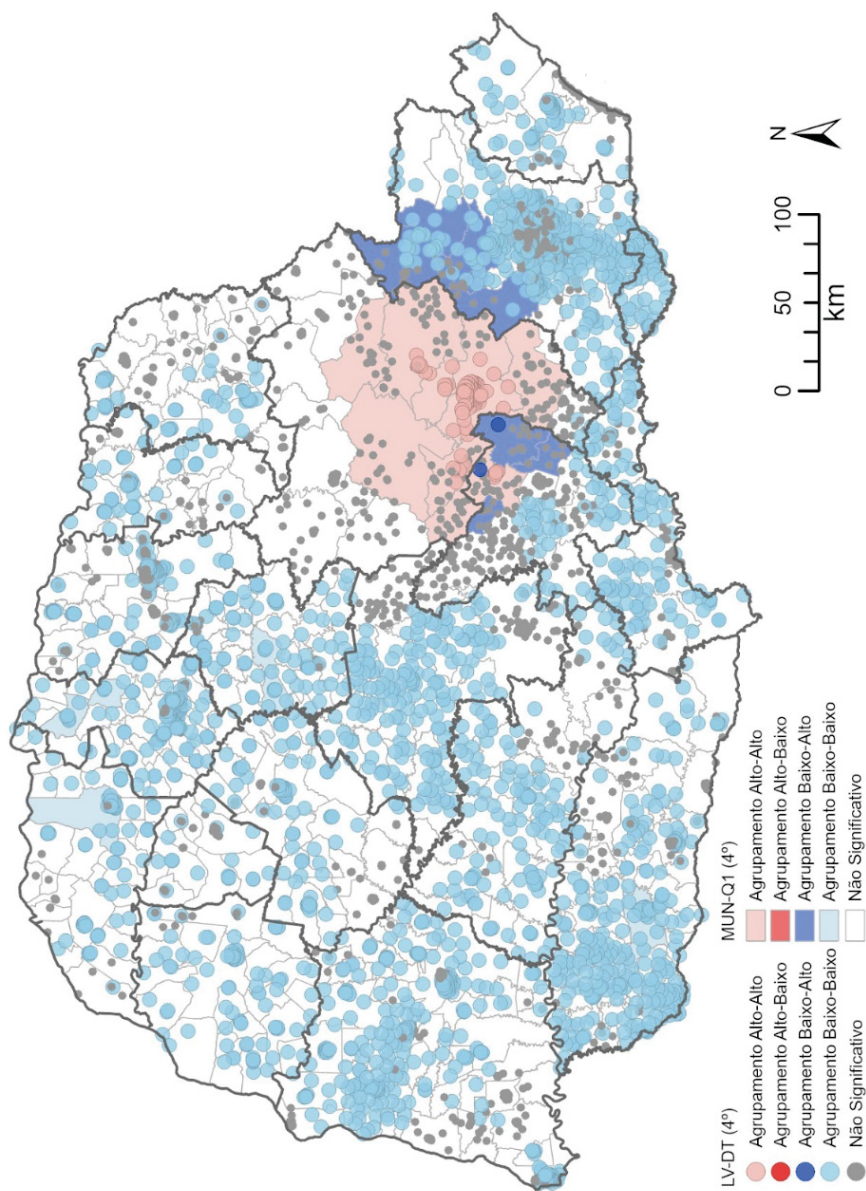


Figura B.9: Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato MABEL CORA CANTO, 2018, Paraná, Brasil.

Fonte: O autor (2021)

Nota: Candidato 20100. Diferença de I de Moran:  $-0,739$  (A-Q1 =  $0,044$ ; P-DL =  $0,783$ )

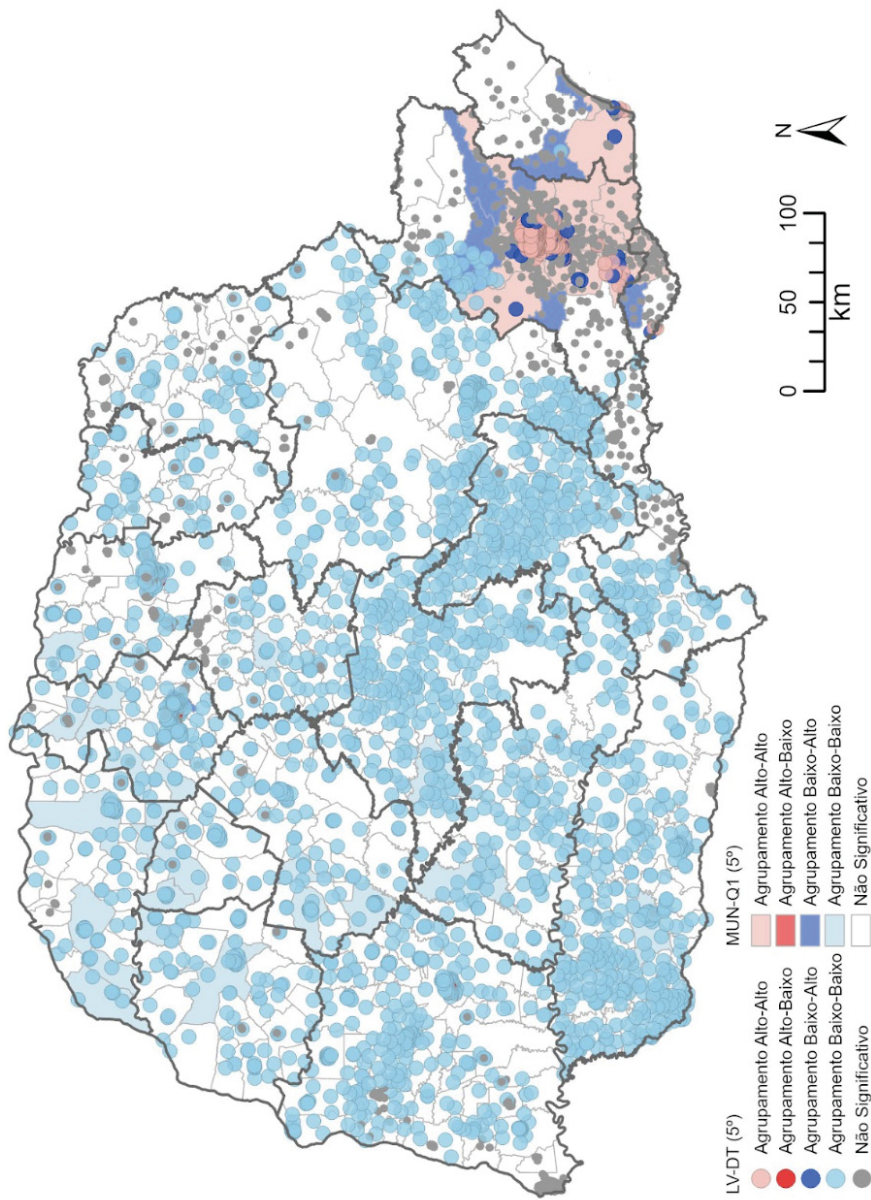


Figura B.10: Mapa do Indicador Local de Associação Espacial nos métodos A-Q1 e P-DL, % de votos do candidato GUSTAVO BONATO FRUET, 2018, Paraná, Brasil.

Fonte: O autor (2021)

Nota: Candidato 1234. Diferença de I de Moran: -0,736 (A-Q1 = 0,050; P-DL = 0,787)

## APÊNDICE C – ETNIAS POR TERRAS INDÍGENAS DO PARANÁ

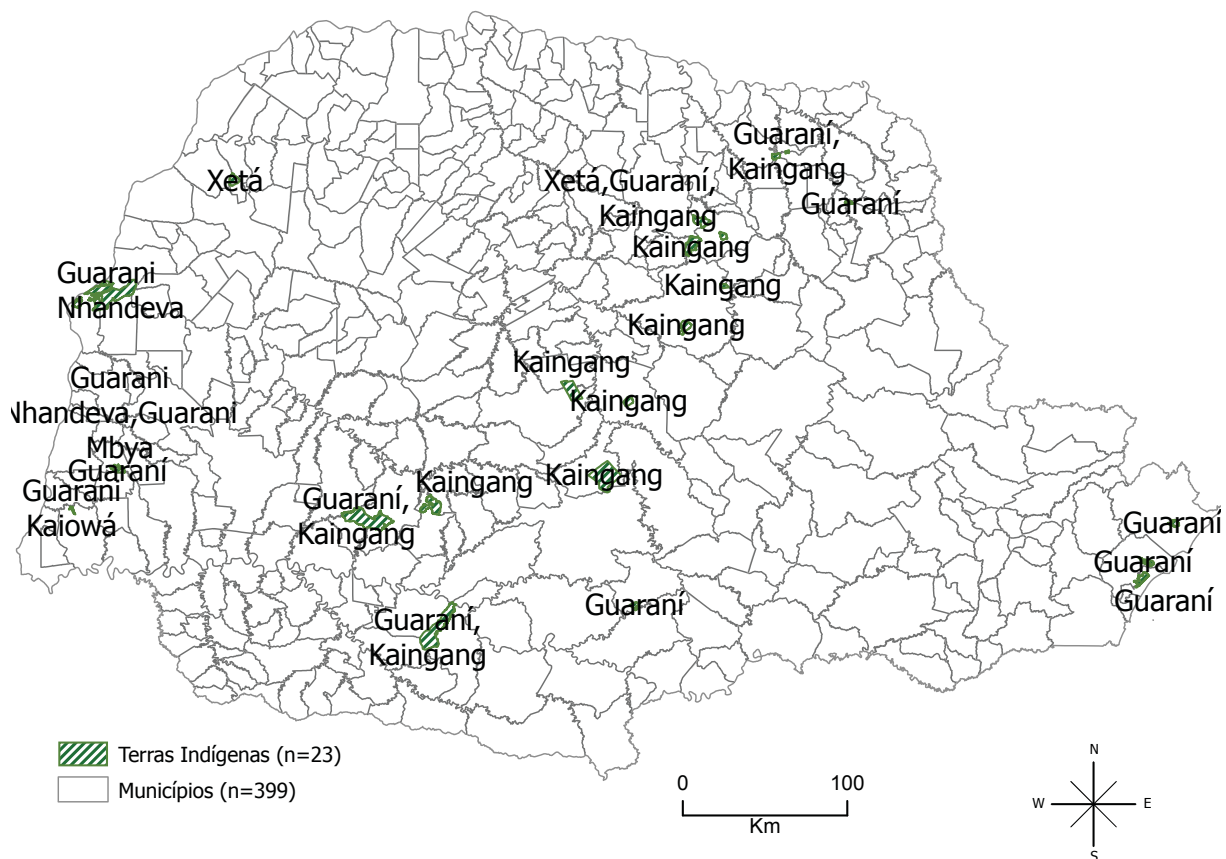


Figura C.1: Terras Indígenas com Legenda de Etnias, Paraná

Fonte: O autor (2021)