

***Head/tail breaks*: uma proposta metodológica para a classificação de dados regionalizados dos fluxos do Cartão BNDES**

Ricardo Agostini Martini
Lucas Azeredo da Silva Teixeira*

Resumo

O presente artigo objetiva colaborar para a melhor compreensão quanto à regionalidade da atuação do BNDES. Mais especificamente, a contribuição do presente trabalho é o desenho de uma metodologia que cumpra dois objetivos. Em primeiro lugar, dentro do universo de informações do Cartão BNDES, segregar um subgrupo de informações relevantes. Em segundo lugar, revelar a hierarquia presente nos dados. Para isso, foi aplicada a metodologia *head/tail breaks*, desenvolvida por Jiang (2013), capaz de classificar dados para georreferenciamento levando em consideração sua assimetria.

Palavras-chave

Geoprocessamento. Cartão BNDES. Desenvolvimento regional. *Head/tail breaks*.

* Economistas do BNDES. Este artigo foi elaborado dentro do projeto corporativo Desenvolvimento Regional e Territorial. Os autores agradecem aos participantes do projeto as discussões e sugestões. Em especial, à equipe do GeoBNDES, Gumersindo Lopez e Penélope Dantas, e a Luiz Felipe Cabral Caldas, Marcelo Machado da Silva, Marcio José de Oliveira, Raffael Cardoso Maciel Attilio e Rodrigo Ferreira Madeira. Agradecem também as críticas e contribuições ao trabalho aos pareceristas anônimos. Este artigo é de exclusiva responsabilidade dos autores, não refletindo, necessariamente, a opinião do BNDES.

Abstract

This article was prepared in order to be a step toward a better understanding of the regionality of BNDES operations. More specifically, the contribution of this work is the design of a methodology that fulfills two objectives. First, within the universe of BNDES Card information, segregate a subset of relevant information. Second, reveal the hierarchy present in the data. For this, it applied the head/tail breaks methodology developed by Jiang (2013), able to sort data for georeferencing considering its asymmetry.

Keywords

Geoprocessing. BNDES Card. Regional development. Head/tail breaks.

Introdução

O presente artigo foi elaborado com o intuito de ser um passo na direção de melhor compreensão da atuação regional do BNDES. Para isso, o Cartão BNDES mostra-se apropriado para um piloto, pois conta com uma base de dados bastante completa, com informações dos usuários do cartão, isto é, as micro, pequenas e médias empresas que o usam como compradoras ou o aceitam como forma de pagamento para suas vendas.

Com as metodologias tradicionais disponíveis nos pacotes estatísticos, já era possível fazer o geoprocessamento dessas informações. Contudo, dada a extensão do banco de dados, muitas vezes os mapas gerados não permitiam distinguir os fluxos por sua importância. Portanto, a contribuição do presente trabalho é o desenho de uma metodologia de classificação de dados que cumpra dois objetivos. Em primeiro lugar, dentro do universo de informações do Cartão BNDES, segregar um subgrupo de informações relevantes para fins de geoprocessamento. Em segundo lugar, revelar a hierarquia presente nos dados.

A classificação de dados pode ser compreendida como um processo de generalização pela diferenciação de informações consideradas vitais daquelas informações consideradas triviais, de maneira a permitir que as primeiras se destaquem (JIANG, 2013). Nesse processo, existem dois parâmetros de interesse: o número de classes em que os dados serão classificados e o intervalo de dados contido em cada classe.

Para trabalhos que envolvem mapeamento estatístico, muitos métodos de classificação de dados podem ser utilizados. Entre esses métodos, os mais simples buscam a construção de classes via intervalos igualitários (*equal steps*), quantis, progressões geométricas e desvios-padrão, por exemplo. Outro método, mais sofisticado,

porém bastante comum, é o de quebras naturais. Esse método classifica os dados em diferentes classes de acordo com as quebras ou lacunas que existem naturalmente nos dados. Para isso, é realizada uma plotagem da frequência dos dados, de modo a permitir a identificação de quebras em seu conjunto.

Com base na noção de quebras naturais, foi desenvolvido o método de otimização de Jenks.¹ Essa metodologia busca computar, para certo número de classes subjetivamente escolhido, intervalos de dados para cada classe que ao mesmo tempo minimize a variância intraclasse e maximize a variância entre as classes. O método de Jenks está bastante difundido na bibliografia de estudos que fazem classificação de dados estatísticos para o mapeamento. Isso inclui trabalhos publicados pelo Instituto Brasileiro de Geografia e Estatística (IBGE), como o registro de Regiões de Influência das Cidades (Regic) e o documento Gestão do Território.² Também está presente em muitos pacotes estatísticos e de geração de mapas, inclusive na forma de comandos automáticos.

Todavia, esses métodos partem do pressuposto de que os dados que serão classificados têm uma distribuição normal. Isto é, a ênfase desses métodos é destacar como informações vitais os eventos frequentes, de intensidade próxima à média e de variância limitada. Contudo, em fenômenos geográficos e econômicos, muitas vezes são os eventos menos frequentes que têm os maiores impactos e que merecem ser destacados como informações vitais e não triviais.

Por isso, a metodologia aplicada no presente trabalho é o *head/tail breaks*, desenvolvida por Jiang (2013), a qual é capaz de classificar dados para georreferenciamento levando em consideração sua

¹ Ver Jenks (1967).

² Para detalhes sobre esses estudos, ver, respectivamente, IBGE (2007) e IBGE (2014).

assimetria. Essa metodologia já foi aplicada para o mapeamento de diversos tipos de dados, mas nunca havia sido aplicada para dados econômicos, o que é uma contribuição original deste trabalho.

O presente artigo aplica a metodologia para classificar os dados e mapear os fluxos monetários do Cartão BNDES para transações inter e intraestaduais no Brasil em 2014. Os mapas gerados por meio desse estudo de caso revelaram três aspectos principais sobre o uso do cartão no país. Primeiro, São Paulo destaca-se como o estado mais importante na qualidade de usuário do Cartão BNDES. Segundo, as transações mais importantes ocorrem entre compradores e fornecedores localizados no mesmo estado, principalmente em São Paulo, Bahia, Paraná, Minas Gerais e Rio Grande do Sul. Terceiro, as principais transações interestaduais envolvem os estados das regiões Sudeste e Sul na posição de fornecedores. Essas características estão de acordo com estudos anteriores de georreferenciamento de atividades econômicas no Brasil.

O artigo seguirá a seguinte estrutura: iniciará com uma revisão bibliográfica sobre as metodologias para a classificação de dados, assim como suas aplicações para o Brasil em estudos anteriores. Depois disso, haverá: uma descrição geral das informações contidas na base de dados do Cartão BNDES; um comentário sobre outras metodologias e a característica dos dados utilizados; apresentação da metodologia; e uma aplicação para os fluxos entre os estados brasileiros.

Geoprocessamento de dados econômicos no Brasil: estudos anteriores

Os estudos voltados para o geoprocessamento de dados econômicos no Brasil costumam focar a identificação de cidades que fun-

cionam como centros de gestão do território e têm se tornado cada vez mais comuns desde a década de 1960 (CORREA, 1989b). Os estudos que focalizam esse tema percorrem duas vias principais. Em primeiro lugar, destacam-se os trabalhos que definem os centros de gestão do território em relação a decisão e controle. Nesse caso, a centralidade urbana é mensurada pelo número de sedes sociais que cada centro concentra e pelo montante de capital fixo que controlam. Em segundo lugar, existem os estudos que definem os centros de gestão em relação a atuação espacial. Nesse caso, a centralidade urbana é calculada de acordo com a localização das filiais das firmas cujas sedes localizam-se em cada centro. Outro indicador de mensuração comum a esses trabalhos é o número de empregados localizados no espaço de ação das sedes sociais das empresas e delas dependentes. Nesse caso, é muito importante o conceito de assalariados externos ao centro urbano, que consiste no conjunto de empregados que trabalham em estabelecimentos pertencentes a uma empresa cuja sede social não está localizada no mesmo lugar que o estabelecimento.

Em relação aos estudos sobre o Brasil, identificam-se quatro trabalhos principais sobre a função da gestão das atividades econômicas pelos centros urbanos. Primeiro, o estudo de Correa (1968), sobre os assalariados externos industriais das metrópoles brasileiras. Esse trabalho utilizou dados do Registro Industrial do IBGE de 1962. Segundo, o estudo de Magnanini e Lima (1971), sobre os centros de gestão segundo a intensidade e a especialização da função de direção e controle com base no número de sedes sociais ali existentes. Para isso, foram abordados dados da revista *Visão* de setembro de 1967. Terceiro, o trabalho de Cordeiro (1987), que definiu os centros de gestão das atividades econômicas como pontos de controle da economia transnacional, discriminando empresas financeiras e não financeiras. Esse trabalho observou dados da revista *Visão*, do *Guia do*

Banco do Brasil e da *Revista Bancária Brasileira* em dois pontos no tempo (1968 e 1984). Quarto, a abordagem de Correa (1989a) acerca dos efeitos da concentração dos bancos comerciais sobre o número de centros de gestão da atividade bancária, isto é, das cidades onde estão as sedes sociais dos bancos, no período de 1961 a 1985.

Mais recentemente, o documento Regic (IBGE, 2007) foi elaborado com o objetivo de investigar a rede urbana brasileira, que foi estabelecida segundo a definição da hierarquia entre os centros urbanos do país. Essa hierarquia, por sua vez, é baseada na classificação dos centros de gestão do território, na intensidade de relacionamentos e na dimensão de influência de cada centro. Nessa hierarquia, foram definidos três níveis de metrópoles, três níveis de capital regional, dois níveis de centro sub-regional, dois níveis de centro de zona e um nível de centro local (incluindo todas as cidades não classificadas anteriormente).

O documento construiu mapas para um total de 43 redes urbanas pertencentes aos centros de mais alto nível, incluindo as metrópoles e os dois primeiros níveis de capitais regionais. A região de influência de cada centro foi definida pelo conjunto de conexões entre as cidades, tanto para o centro principal como para os centros secundários de cada rede urbana. As conexões refletem ligações, isto é, a busca por bens ou serviços, sendo ranqueadas de acordo com sua intensidade.

A metodologia para a construção das redes urbanas brasileiras consistiu em três passos. Em primeiro lugar, procurou-se identificar a centralidade dos centros. A principal noção de centralidade urbana adotada é a gestão do território, isto é, a presença de órgãos do Estado e de sedes de empresas que tomam decisões capazes de afetar outras cidades. Para qualificar melhor a centralidade dos núcleos identificados, assim como para incluir outros núcleos não identificados pelo critério anterior, foi considerado o acesso a equipamentos e

serviços por parte do município. Isso inclui atividades de comércio e serviços, atividade financeira, ensino superior, serviços de saúde, acesso à internet, redes de TV aberta e transporte aéreo.

Em segundo lugar, buscou-se investigar as ligações entre as cidades, de modo a delinear as áreas de influência dos centros e esclarecer a articulação das redes no território. Observa-se que, no caso dos centros urbanos menores, esse ponto foi investigado por meio de uma pesquisa direta via questionários. A definição das regiões de influência dos centros urbanos foi feita pela combinação das redes de ligação dos centros de gestão e dos centros menores, nesse caso segundo as informações respondidas nos questionários.

Terceiro, buscou-se hierarquizar os centros urbanos. A hierarquização foi feita pela classificação dos centros de gestão, a intensidade dos relacionamentos e a dimensão da região de influência de cada centro.

O documento Divisão Urbano Regional (IBGE, 2013) teve o objetivo de contribuir à análise da dinâmica territorial brasileira. Para isso, o trabalho buscou delimitar regiões, com base nas informações contidas na publicação Regic, de 2007, de modo a obter mapas de agregados coesos de municípios, resolvendo possíveis múltiplas vinculações. Foram construídos mapas em três escalas regionais: as regiões imediatas, intermediárias e ampliadas de articulação urbana. Tais mapas permitiram análises estatísticas regionais nas quais cada município pertencia somente a uma região, de modo a oferecer uma visão regional do Brasil com base nos fluxos articulados por sua rede urbana.

O documento Gestão do Território (IBGE, 2014) teve o objetivo de abordar o tema das redes e fluxos do território. Mais especificamente, o estudo focou na dimensão dos fluxos de gestão entre os núcleos urbanos, isto é, o conjunto de fluxos de ordens, hierarquias, informações, poder e dinheiro entre os agentes que se organi-

zam de forma multilocalizada no território. Com isso, pretendeu-se compreender como as diferentes regiões e cidades conectam-se por meio da gestão do território e quais são os centros que concentram a capacidade de comando e de controle no país.

A análise da gestão territorial incluiu a observação de instituições públicas e privadas. Em relação às instituições públicas, o estudo considerou as relações entre as agências, as unidades de atendimento, as gerências regionais, as superintendências e as sedes do Instituto Nacional de Seguridade Social (INSS), do Ministério do Trabalho e Emprego (MTE), da Secretaria da Receita Federal (SRF), da Justiça Federal, dos Tribunais Regionais Eleitorais (TRE), dos Tribunais Regionais do Trabalho (TRT) e do próprio IBGE. Em relação às instituições privadas, foi pesquisada a gestão empresarial, caracterizada principalmente pelas relações entre as sedes e as filiais das empresas quando localizadas em municípios diferentes. Para isso, são observados dados do Cadastro Central de Empresas (Cempre) do IBGE, levantados em 2011. Para os órgãos públicos federais, foram levantadas informações sobre sua estrutura organizacional interna referente a 2013.

Entre os resultados principais obtidos pela análise, ficou constatado que os centros de gestão pública são aquelas cidades que detêm entidades das instituições públicas, com poder de comando sobre o conjunto de suas redes. Geograficamente, os centros seguem a distribuição demográfica e econômica do país, mas nota-se que sua estrutura tem um caráter político, isto é, busca a cobertura territorial da atuação dos organismos.

As conexões entre as cidades são, majoritariamente, de curta distância. Há uma relativa concentração no Centro-Sul do país, mas também estão presentes nas demais regiões como consequência do caráter político dessas ligações, representando uma vontade política da presença do Estado em todas as regiões. As redes institucionais

são hierárquicas, sem a presença de relações entre centros de mesmo nível.

Em relação aos centros de gestão empresarial, a rede urbana, entendida como o espaço de diálogo entre as sedes e suas filiais, também tem forma hierárquica. Além disso, novamente se verifica uma concentração no Centro-Sul do país, com destaque para as regiões metropolitanas de São Paulo, Rio de Janeiro e Brasília. O grande destaque das ligações entre os centros é a conexão Rio-São Paulo. A análise dos assalariados externos revelou uma grande dependência das regiões Norte, Nordeste e Centro-Oeste em relação ao Sudeste do Brasil. Por fim, a oferta de serviços avançados de apoio às empresas reforça o padrão de concentração espacial no país.

O documento Arranjos Populacionais e Concentrações Urbanas do Brasil (IBGE, 2015) teve o objetivo de buscar a identificação dos arranjos populacionais empregando duas noções: a de integração, mensurada pelos fluxos de movimentos pendulares populacionais para trabalho e estudo, e a de contiguidade urbana. Em relação aos resultados encontrados, a pesquisa verificou uma série de aspectos dos arranjos populacionais e concentrações urbanas do Brasil. Foram identificados no país 294 arranjos populacionais, formados por 938 municípios. Em 2010, eles totalizavam 106.782.428 habitantes, ou 55,9% da população residente no Brasil naquele ano. De forma geral, os arranjos populacionais seguem o padrão da urbanização brasileira, ou seja, acompanham as metrópoles e capitais regionais de acordo com os critérios estabelecidos pelo Regic de 2007.

Ressalta-se, no entanto, que todos os trabalhos relatados classificaram os dados levantados para o georreferenciamento seguindo metodologias que partiram do pressuposto de normalidade de sua distribuição. Em muitos casos, os estudos optaram por não utilizar indicadores monetários para esse fim, em que a não normalidade

da distribuição é mais evidente. Ou seja, a não normalidade da distribuição dos dados econômicos, em especial aqueles mensurados monetariamente, é um tema que não foi apropriadamente tratado na literatura relatada sobre georreferenciamento, apesar de estar diretamente relacionada com um importante aspecto da economia regional do país, que é a desigualdade. Por isso, o presente trabalho busca contribuir com o debate ao propor uma metodologia voltada para essa característica dos dados.

Metodologia empírica³

Os métodos mais comuns de classificação de dados, como os que buscam a construção de classes via intervalos iguais, quantis, progressões geométricas, desvios-padrão e quebras naturais, partem do pressuposto de que os dados que serão classificados têm uma distribuição normal. Isto é, a ênfase desses métodos é destacar os eventos frequentes, de intensidade próxima à média e de variância limitada. Contudo, em fenômenos geográficos e econômicos, muitas vezes são os eventos menos frequentes que têm os maiores efeitos, e que merecem ser destacados como informações vitais, e não triviais. Ou seja, os dados revelam a existência de desigualdade nos eventos. Por exemplo, na geografia, é comum que a população se distribua desigualmente no território, o que caracteriza o fenômeno da concentração populacional; na economia, a renda se distribui de maneira desigual entre os indivíduos de uma determinada população, caracterizando a concentração de renda.

Nesses casos, a distribuição de dados apresenta um pico (*head*) e uma cauda (*tail*), separados pela média aritmética dos valores dos

³ Toda a presente seção foi baseada no trabalho de Jiang (2013).

dados. Mais especificamente, a série apresenta uma longa cauda à direita, ao passo que, na proximidade do eixo vertical, há uma concentração de valores de baixa intensidade e elevada frequência. Estatisticamente, assim se comporta uma distribuição de cauda pesada,⁴ que se caracteriza por ser assimétrica e leptocúrtica (ou de pico acentuado). Tecnicamente, uma distribuição de cauda pesada pode ser definida como:

given a variable X, if its values x follow a heavy tailed distribution, then the mean (m) of the values can divide all the values into two parts: a high percentage in the tail, and a low percentage in the head. For the sake of convenience, we call this regularity the head/tail division rule (JIANG; LIU, 2012, p. 4).

O exemplo matemático mais típico de uma distribuição de cauda pesada é a lei de potência, um padrão característico de sistemas complexos que contêm relações não lineares entre variáveis, de modo que pequenas causas podem acarretar grandes consequências de maneira desproporcional.⁵ Seja x uma variável de interesse e y a probabilidade de ocorrência de cada valor de x , isto é, $p(x)$, a lei de potência é definida pela seguinte equação:

$$y = cx^{-m} \quad (1)$$

Nessa equação, m é o exponencial da lei de potência e c é a constante de proporcionalidade.⁶ A equação pode ser linearizada quanto a logaritmos, obtendo-se:

$$\ln y = \ln c - m \ln x \quad (2)$$

⁴ Em inglês, *heavy tailed distribution*.

⁵ Outros exemplos de distribuições de cauda pesada são a distribuição log-normal e a distribuição exponencial.

⁶ Mais detalhes sobre a distribuição de lei de potência podem ser encontrados em Guerriero (2012) e Jiang (2015).

Segundo Jiang (2013), as distribuições de dados do tipo cauda pesada são mais comuns em sistemas geográficos longos em relação a espaço e tempo.

O método *head/tail breaks* é uma proposta metodológica para a classificação de dados que se distribuem com cauda pesada, isto é, que refletem o comportamento de eventos que apresentam assimetria. Basicamente, essa metodologia foca a identificação de eventos de baixa frequência, mas com maior intensidade, o que os faz serem considerados mais importantes que os eventos de elevada frequência. Dessa maneira, os eventos de elevada frequência e baixa intensidade constituem o *background* por meio do qual se distinguem os eventos de baixa frequência e elevada intensidade. Em muitos casos, essa assimetria também ocorre entre os próprios eventos considerados primeiramente de baixa frequência e elevada intensidade: é verificada uma nova minoria de eventos de frequência muito baixa e intensidade muito elevada. Assim esses eventos também podem ser divididos ao redor de sua média aritmética entre muitos pequenos valores e poucos grandes valores. Em outras palavras, os valores dos eventos têm uma estrutura hierárquica inerente ao conjunto de dados que constitui a fundação do método de classificação desenvolvido neste artigo. Operacionalmente, o método pode ser explicado em quatro passos:

1. Calcular a média aritmética da variável de interesse.
2. Agrupar os valores dos dados em duas partes acerca da média aritmética. Denominam-se *head* (pico) os valores acima da média, e *tail* (cauda) os valores abaixo da média.
3. Verificar se há desigualdade na proporção de observações agrupadas como *head* e como *tail*, segundo critério escolhido pelo pesquisador e de acordo com as características dos dados classificados. Por exemplo, uma proporção acima de 40%/60%.

4. Verificada a desigualdade, continuar o processo de particionamento para valores acima da média (*head*). O processo segue iterativamente até que os valores do pico não estejam mais distribuídos de maneira desigual em relação à cauda, ou até que o último *head* calculado contenha uma única observação.

Assim, a primeira classe construída vai se estender desde o valor mínimo verificado para a variável até a primeira média, e constituirá o *background* para a próxima classe. A segunda classe vai se estender da primeira até a segunda média, e constituirá o *background* para a terceira classe. Esse processo segue iterativamente até uma classe final que vai se estender desde a última média calculada até o valor máximo verificado para a variável, o que constitui o primeiro plano da distribuição. Dessa maneira, todas as classes calculadas no processo seguirão o mesmo padrão de desigualdade verificado nos dados originais, contando com um pequeno primeiro plano (*head*) e um extenso *background* (*tail*).

Outra preocupação de suma importância em relação ao geoprocessamento de dados distribuídos com cauda pesada refere-se ao cálculo do número de classes que, partindo da menor para a maior, formam uma estrutura hierárquica. Para isso, o número de classes determinado pelo método *head/tail breaks* é definido como Índice ht (JIANG; YIN, 2014). Conceitualmente, os autores definem esse índice como um mais o número de vezes em que há mais unidades pequenas do que grandes dentro do universo de dados analisados. Mais especificamente, um sistema geográfico tem um Índice ht igual a h se o padrão de haver mais unidades pequenas do que grandes se repete $h-1$ vezes nesse sistema em diferentes escalas. Segundo os autores, quanto maior o Índice ht, mais complexo, heterogêneo, maduro e diverso é o sistema geográfico em que os dados classificados estão inseridos, servindo, portanto, como uma importante medida de heterogeneidade espacial.

Esse método apresenta três vantagens principais em relação aos métodos citados para a classificação de dados, sobretudo em relação ao método de Jenks:

1. O método *head/tail breaks* revela o padrão hierárquico verificado na distribuição dos dados. Esse é um aspecto fundamental para dados sobre eventos desiguais, e deve estar refletido na geração de mapas.
2. Tanto o número de classes como os intervalos contidos em cada classe são naturalmente determinados pelos dados nesse processo. O método *head/tail breaks* reduz a subjetividade presente em métodos baseados em quebras naturais.
3. O número de classes construídas segundo o processo raramente é elevado. Mesmo nesses casos, as classes mais elevadas, com poucas observações contidas nelas, podem ser incorporadas umas nas outras sem que o conjunto perca sua identidade hierárquica.

As aplicações do método *head/tail breaks* para a classificação de dados vêm se destacando em estudos recentes, que abordam bases de dados extensas. A metodologia já foi adotada para classificar dados do tamanho das quadras das cidades da França, Alemanha e do Reino Unido (JIANG; LIU, 2012), do trânsito nas estradas da Suécia (JIANG, 2013), da iluminação noturna no mapa dos Estados Unidos (JIANG; YIN, 2014), da distribuição espacial dos usuários da rede social Brightkite nos Estados Unidos (JIANG; MIAO, 2014) e para a análise dos dados da plataforma aberta OpenStreetMap em relação a usuários, elementos geográficos e contribuições (MA; SANDBERG; JIANG, 2015). Destaca-se que o método ainda não foi utilizado em estudos econômicos, ou mesmo para a classificação de dados monetários, sendo esta uma contribuição original do presente trabalho.

Base de dados do Cartão BNDES

O Cartão BNDES, criado pelo BNDES, é um instrumento de crédito para as micro, pequenas e médias empresas (MPME) financiarem bens de capital, insumos e serviços selecionados. Mais especificamente, o cartão é uma linha de crédito rotativa e pré-aprovada destinada à aquisição de itens necessários às atividades das MPMEs que tenham fabricação no país e que estejam cadastrados no *site* do Cartão BNDES, por fornecedores devidamente credenciados.⁷

Podem obter o Cartão BNDES para a aquisição de itens as empresas com faturamento bruto anual de até R\$ 90 milhões, de controle nacional, que exerçam atividade econômica compatível com as Políticas Operacionais e de Crédito do BNDES e que estejam em dia com o INSS, o Fundo de Garantia do Tempo de Serviço (FGTS), os tributos federais e a Relação Anual de Informações Sociais (Rais), além de não ter restrições na Serasa, Serviço de Proteção ao Crédito (SPC) ou Banco Central. Quando o solicitante fizer parte de um grupo econômico, o faturamento bruto anual do grupo também não poderá exceder o limite de R\$ 90 milhões.

Do ponto de vista dos fornecedores, podem se credenciar no *site* do Cartão BNDES as empresas fabricantes de bens e insumos de setores autorizados pelo BNDES, com índice de nacionalização mínimo de 60% e necessários às atividades das MPMEs. Além disso, estão aptas ao credenciamento as entidades acreditadas pelo Instituto Nacional de Metrologia, Qualidade e Tecnologia (Inmetro) ou credenciadas por outras instituições federais que prestam serviços de metrologia, normalização e avaliação de con-

⁷ Mais detalhes sobre os aspectos institucionais do Cartão BNDES podem ser obtidos no *site* oficial desse instrumento (<<https://www.cartaobndes.gov.br/cartaobndes/>>) e em CNI (2009).

formidade. Por fim, os fabricantes também poderão indicar seus distribuidores e revendedores para realizar vendas financiadas com o Cartão BNDES.

O Cartão BNDES pode ser utilizado para adquirir mais de 265 mil itens de variados setores, expostos no *site* institucional pelos fornecedores credenciados. A lista de itens inclui, por exemplo, computadores, móveis e utensílios, motos e caminhões, componentes, partes e peças de bens de capital, equipamentos de informática e automação industrial e até serviços de pesquisa, desenvolvimento e inovação relacionados à pesquisa aplicada para a criação de produtos e processos.

Em 2014, o total de desembolsos via Cartão BNDES superou os R\$ 11,5 bilhões, o que representa 6,1% do total de desembolsos do Banco, conforme mostra a Tabela 1. São Paulo destaca-se no total desembolsado via cartão (R\$ 3,1 bilhões), ao passo que proporcionalmente ao total de desembolsos do Banco em um estado destaca-se o Amazonas (16,5%). Nesse mesmo ano, o número de total de operações via Cartão BNDES foi de 795.242, 70,3% do total de operações do Banco. O destaque em termos absolutos foi São Paulo (quase duzentas mil operações), e, em termos relativos, o Amapá (92,1%).

O universo de informações do Cartão BNDES é bastante vasto. Contudo, algumas informações essenciais devem constar na base de dados que dará origem a mapas com os fluxos hierarquizados. Essa base deve conter dados referentes a, no mínimo, três categorias distintas: informação georreferenciável do comprador, informação georreferenciável do fornecedor e valor da operação.

Tabela 1

Operações do Cartão BNDES nas unidades da Federação do Brasil, em 2014

Unidade da Federação	Desembolsos do Cartão BNDES (R\$ milhões)	Participação do Cartão BNDES no total de desembolsos (%)	Número de operações do Cartão BNDES	Participação do Cartão BNDES no número total de operações (%)
Acre	33,7	9,6	3.377	85,1
Alagoas	76,5	13,8	4.502	78,2
Amapá	49,8	8,7	3.648	92,1
Amazonas	121,1	16,5	7.143	83,8
Bahia	601,7	10,3	41.053	81,3
Ceará	282,9	8,9	20.528	82,5
Distrito Federal	152,0	2,8	10.270	85,6
Espírito Santo	239,0	7,9	18.016	67,4
Goiás	467,9	8,6	34.400	75,0
Maranhão	222,9	4,9	12.928	76,3
Mato Grosso	352,1	5,5	25.368	68,8
Mato Grosso do Sul	186,0	4,4	14.121	71,7
Minas Gerais	1.145,5	7,5	84.781	70,7
Pará	262,0	2,8	15.805	78,3
Paraíba	127,8	11,0	7.830	82,0
Paraná	1.086,4	7,1	76.542	49,6
Pernambuco	277,7	4,7	15.522	77,5
Piauí	127,8	7,7	8.056	82,7
Rio de Janeiro	555,0	2,1	36.067	79,9
Rio Grande do Norte	118,6	10,1	8.074	85,3
Rio Grande do Sul	950,8	7,1	74.654	59,8
Rondônia	96,4	5,8	7.110	63,7
Roraima	11,6	8,2	906	72,8
Santa Catarina	824,0	8,5	56.635	63,8
São Paulo	3.051,2	6,8	199.231	80,4
Sergipe	44,5	10,9	2.931	64,9
Tocantins	82,7	6,8	5.744	63,4
Brasil	11.547,6	6,1	795.242	70,3

Fonte: Elaboração própria, com base em dados do Boletim de Desempenho do BNDES.

O banco de dados do Cartão BNDES trabalha com duas figuras básicas: o comprador e o fornecedor. O comprador é a empresa possuidora e usuária de um Cartão BNDES. O fornecedor é uma empresa cadastrada à rede de fornecedores do Portal de Operações do Cartão BNDES e que oferta produtos feitos por empresas credenciadas, que atendem às normas do BNDES. O fornecedor pode ser um fabricante ou um distribuidor, que, nesse caso, apenas revende produtos fabricados por outros.

Há diversas informações passíveis de georreferenciamento para compradores e fornecedores: município, estado, região, logradouro, bairro, Código de Endereçamento Postal (CEP). A informação a ser utilizada depende dos objetivos do pesquisador e das limitações do *software* de geoprocessamento a ser utilizado. Para a metodologia de hierarquização de informação é indiferente qual informação será usada.

Cabe salientar uma limitação da base de dados do Cartão BNDES. O fornecedor tem um único Cadastro Nacional de Pessoas Jurídicas (CNPJ) e um único endereço, porém pode dispor de uma rede de estabelecimentos em diferentes localidades. Além disso, no cadastramento no BNDES, não há distinção entre os estabelecimentos de uma mesma empresa (definida pelos primeiros oito dígitos do CNPJ). Isso abre espaço para que a representação cartográfica de algumas transações não seja fidedigna. Por exemplo, suponha-se uma operação entre uma empresa compradora e um fornecedor no estado de Minas Gerais. Ainda que o fornecedor seja uma filial localizada no mesmo estado, no cadastro do Cartão BNDES, o endereço que consta é de sua matriz (por exemplo, em São Paulo).

A outra categoria indispensável diz respeito ao valor da operação. Na base de dados do Cartão BNDES, encontram-se três tipos distintos de informação: valor total, valor autorizado e valor finan-

ciado. O valor financiado é a parcela do valor da operação efetivamente financiado pelo BNDES. O valor autorizado é a parcela do valor da operação que o BNDES autorizou que seja financiado. Discrepâncias entre o valor autorizado e o valor financiado dizem respeito a situações nas quais o intermediário financeiro não pede o recurso ao BNDES. O valor total, por sua vez, inclui tanto o valor financiado quanto o montante da operação feita com recursos próprios do comprador.

Além disso, é possível fazer recortes dos dados a serem selecionados: por produto, por atividade do fornecedor e/ou do comprador, por faturamento do comprador, entre outras possibilidades. A seleção de dados a ser feita – e seus recortes – depende dos objetivos específicos de cada pesquisa.

Estudo de caso: fluxos do Cartão BNDES para transações inter e intraestaduais, no Brasil, em 2014

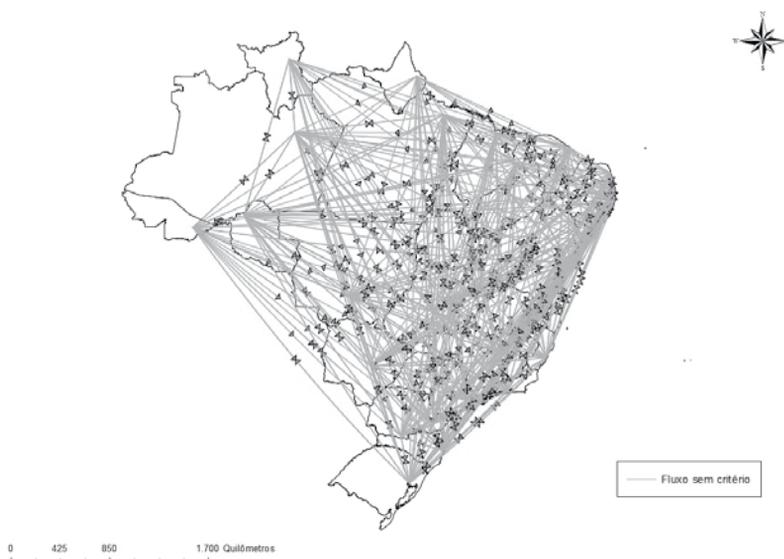
Como um primeiro exercício, a metodologia foi aplicada para o Brasil em geral. A base de dados que foi extraída do universo do Cartão BNDES adotou os seguintes critérios: foram selecionadas todas as operações do ano de 2014, agregando os compradores e os fornecedores de acordo com sua unidade federativa. As operações, por sua vez, foram mensuradas por seu valor total, considerado neste trabalho a informação mais completa sobre as operações. Todavia, é importante destacar que a escolha do tipo de valor de operação não afeta seu geoprocessamento, pois há assimetria na distribuição dos dados para qualquer categoria de valor da operação. Todas as operações realizadas para o mesmo par “estado comprador-estado fornecedor” foram agregadas, pois o interesse é na relação que se

estabelece entre diferentes localidades geográficas. Essa medida agregada aplicada pelo trabalho é denominada de “transações”, representando o conjunto de operações reais entre empresas de estados diferentes ou não.

Caso houvesse transações entre todas as unidades federativas, o número máximo de transações a serem mapeadas seria de 729. Em 2014, ocorreram 626 transações entre e intraestados. Caso todas fossem mapeadas sem nenhuma metodologia de classificação de dados, o mapa seria de impossível leitura, conforme se pode ver no Mapa 1.

Mapa 1

Fluxos do Cartão BNDES unidade da Federação x unidade da Federação, em 2014, sem classificação

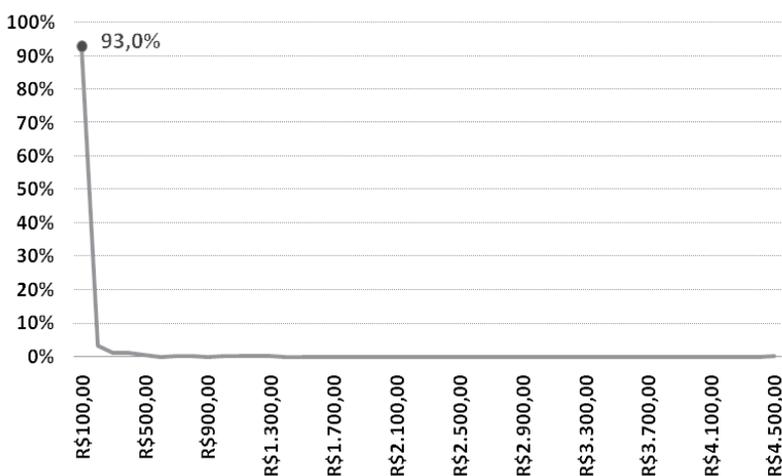


Fonte: Elaborado pelo GeoBNDES, com base em dados do Cartão BNDES.

Nesse universo de 626 transações, o menor fluxo é de R\$ 463,68, tendo o Pará como estado comprador e o Acre como estado fornecedor. O maior fluxo foi de R\$ 4,5 bilhões, tendo o estado de São Paulo nas duas pontas da transação. Nesse universo, 93% das transações são menores que R\$ 100 milhões, como mostra o Gráfico 1. Esse gráfico é uma aproximação contínua para o histograma da base de dados deste artigo, agrupando-os em classes com intervalos de R\$ 100 milhões. Como pode ser visto, a distribuição apresenta acentuada assimetria, característica de uma distribuição de cauda pesada.

Gráfico 1

Distribuição dos dados por classes de R\$ 100 milhões



Fonte: Elaboração própria, com base em dados do Cartão BNDES.

De acordo com o que foi exposto na seção anterior, a primeira etapa do processo de georreferenciamento consistiu em hierarquizar os dados segundo a metodologia *head/tail breaks*. Os níveis hierárquicos se revelaram conforme a Tabela 2.

Tabela 2

**Transações do Cartão BNDES entre unidades da Federação:
distribuição *head* e *tail breaks***

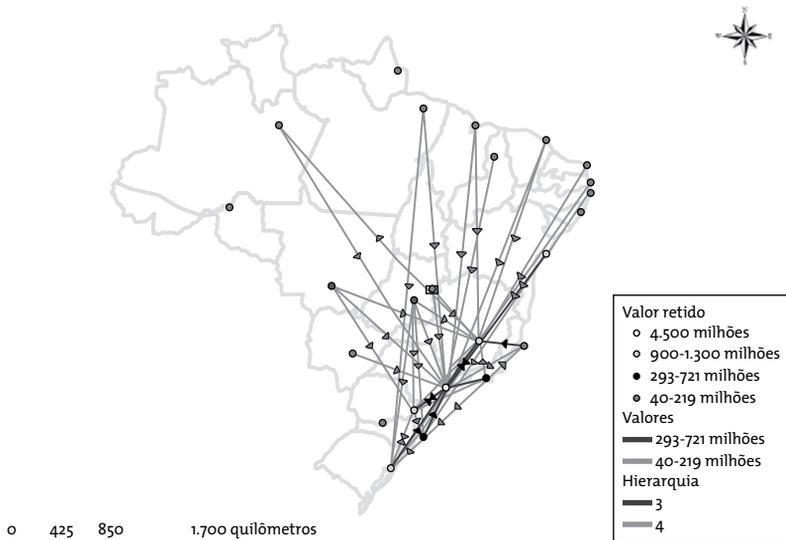
Intervalo de valores (R\$)	Número de transações	Média	Número de <i>head</i>	% de <i>head</i>	Número de <i>tail</i>	% de <i>tail</i>
464-4.498.614.860	626	38.211.051,62	82	13,1	544	86,9
38.211.052-4.498.614.860	82	256.106.156,87	19	23,2	63	76,8
256.106.157-4.498.614.860	19	773.082.051,05	5	26,3	14	73,7
773.082.051-4.498.614.860	5	1.803.280.766,63	1	20,0	4	80,0
1.803.280.767-4.498.614.860	1	4.498.614.860,01	-	-	-	-

Fonte: Elaboração própria, com base em dados do Cartão BNDES.

Como se vê na Tabela 2, foi necessário criar cinco grupos para lidar com a assimetria da base de dados. Ou seja, o Índice ht obtido pela classificação dos dados do Cartão BNDES é igual a 5, refletindo uma considerável heterogeneidade no território brasileiro em relação ao uso desse produto. Para gerar um mapa de leitura mais fácil sem perder informações vitais para a representação do quadro geral, o segundo passo consistiu na eliminação visual de todo o grupo de menor nível hierárquico, que agrega as transações de menor valor. Além disso, as transações foram demarcadas conforme seu nível hierárquico. O resultado é o Mapa 2, em que as setas indicam a direção do fluxo monetário, isto é, do comprador para o fornecedor. Já os pontos presentes em alguns estados representam as transações intraestaduais, isto é, os valores retidos no estado.

Mapa 2

Fluxos do Cartão BNDES unidade da Federação x unidade da Federação, em 2014, metodologia *head/tail breaks*



Fonte: Elaborado pelo GeoBNDES, com base em dados do Cartão BNDES.

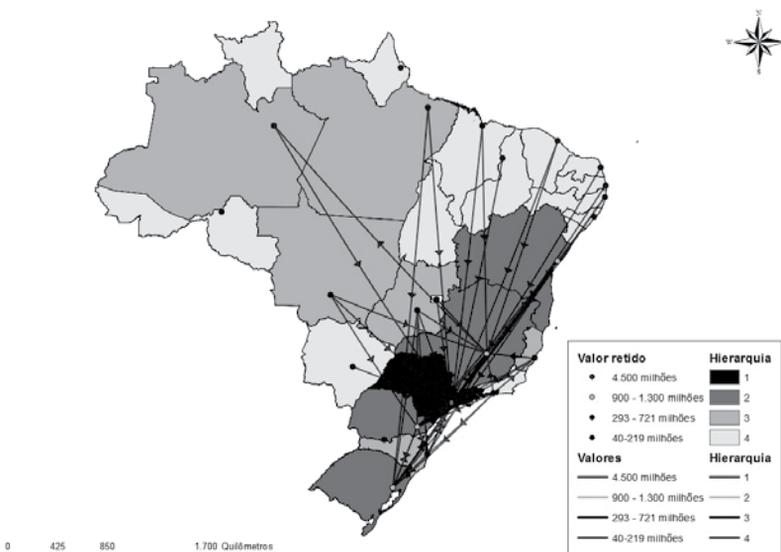
Observa-se que os intervalos de dados contidos em cada nível hierárquico no mapa não coincidem exatamente com os intervalos calculados na Tabela 1. Isso acontece porque os intervalos na tabela foram estabelecidos pelo cálculo de médias dos valores das transações. No mapa, os valores que limitam os níveis hierárquicos são os das transações reais que estão contidas nos intervalos calculados.

Além disso, foi necessário hierarquizar não apenas os fluxos, mas também a importância relativa dos estados como geradores de gastos via Cartão BNDES. Para isso, procurou-se uma medida que priorizasse os estados não só com maior uso do cartão, mas também com uma baixa assimetria em seu uso, ou seja, que o estado compra-

Os dois últimos mapas podem ser combinados em um único, agrupando todas as informações georreferenciáveis relativas ao uso do Cartão BNDES por estados em 2014.

Mapa 4

Fluxos do Cartão BNDES unidade da Federação x unidade da Federação, em 2014, classificação completa



Fonte: Elaborado pelo GeoBNDES, com base em dados do Cartão BNDES.

Como se pode ver, São Paulo é o estado mais importante como usuário do Cartão BNDES. Além disso, as transações mais importantes são intraestaduais, que concentram todas as transações dos dois níveis hierárquicos mais elevados conforme a classificação realizada pelo método *head/tail breaks*, destacando-se São Paulo (R\$ 4,5 bilhões), Bahia (R\$ 1,3 bilhão), Paraná (R\$ 1,1 bilhão), Minas Gerais (R\$ 1,1 bilhão) e Rio Grande do Sul (R\$ 996 milhões). Entre os estados, de forma geral, as principais transações envolvem

os estados do Sudeste e do Sul do país, sobretudo na posição de fornecedores de insumos. O destaque refere-se a transações de empresas de Minas Gerais como compradoras de insumos fornecidos por empresas de São Paulo (R\$ 637 milhões). O quadro geral dos fluxos monetários do Cartão BNDES é visível mesmo com a não representação de todos os fluxos classificados no nível hierárquico mais baixo, ainda que esse grupo contenha 86,9% das transações. Ou seja, o procedimento efetuado foi capaz de simplificar o geoprocessamento dos fluxos do Cartão BNDES sem acarretar perda de informações vitais. Por fim, cabe destacar que o padrão revelado pelos fluxos monetários do Cartão BNDES é semelhante ao que foi verificado no documento Gestão do Território (IBGE, 2014), principalmente em relação às características das redes dos centros de gestão empresarial, o que corrobora os resultados alcançados neste trabalho.

Considerações finais

Este artigo teve como objetivo apresentar uma metodologia de classificação para geoprocessamento dos dados do Cartão BNDES, segregando as informações mais proeminentes e revelando sua hierarquia intrínseca. Uma característica fundamental observada nos dados é sua assimetria: coexistem transações de baixo valor e elevada frequência e transações de elevado valor e baixa frequência. Esse problema ainda é pouco abordado pela literatura nacional sobre o georreferenciamento de dados de natureza econômica, apesar de estar relacionado a uma das principais características da estrutura econômica regional brasileira, que é a desigualdade. Com o uso da metodologia *head/tail breaks*, foi possível gerar mapas que evidenciem as transações de acordo com sua ordem de importância.

Os mapas gerados revelaram três características principais dos fluxos monetários do Cartão BNDES entre os estados brasileiros. Primeiro, São Paulo é o estado mais importante como usuário do Cartão BNDES. Segundo, as transações mais importantes são intraestaduais, com destaque para São Paulo, Bahia, Paraná, Minas Gerais e Rio Grande do Sul. Terceiro, as principais transações interestaduais incluem os estados das regiões Sudeste e Sul, principalmente na posição de fornecedores. De forma geral, os resultados verificados seguem o padrão observado no estudo Gestão do Território, de 2014, principalmente em relação aos centros de gestão empresarial, que apresentam uma rede de formato hierárquico, com concentração do Centro-Sul do país e destaque para as atividades das empresas paulistas. Uma diferença entre os resultados verificados pelo presente estudo e o de IBGE (2014) é que, neste, o principal parceiro de São Paulo é Minas Gerais, e não o Rio de Janeiro. Possíveis motivos para a sub-representação fluminense na hierarquia dos principais usuários do Cartão BNDES transcendem o escopo do presente artigo, ainda que inspirem futuras pesquisas. Pode-se inferir, segundo os mapas aqui construídos, que a concentração dos fluxos estaduais do Cartão BNDES reflete a concentração econômica das atividades privadas no Brasil.

O próximo passo da pesquisa é aplicar essa metodologia para diferentes extrações de dados do universo do Cartão BNDES, analisando em mais detalhes outros recortes do território brasileiro. Será dada ênfase na análise das redes municipais de fluxos monetários, como forma de propiciar um maior detalhamento das conexões entre fornecedores e compradores no território nacional, assim como sua dinâmica temporal e sua relação com os padrões verificados em outros estudos, como o Regic (IBGE, 2007).

Referências

- CNI – CONFEDERAÇÃO NACIONAL DA INDÚSTRIA. *Cartão BNDES: como as micro, pequenas e médias empresas podem se beneficiar*. 2. ed. Brasília, 2009.
- CORDEIRO, H. K. Os principais pontos de controle da economia transnacional no espaço brasileiro. *Boletim de Geografia Teorética*, Rio Claro, v. 16-17, n. 31-34, p. 153-196, 1987.
- CORREA, R. L. Contribuição ao estudo do papel dirigente das metrópoles brasileiras. *Revista Brasileira de Geografia*, v. 30, n. 2, p. 56-86, 1968.
- _____. Concentração bancária e Centros de Gestão do Território. *Revista Brasileira de Geografia*, v. 51, n. 2, p. 17-32, 1989a.
- _____. Os centros de gestão e seu estudo. *Revista Brasileira de Geografia*, v. 51, n. 4, p. 109-119, 1989b.
- GUERRIERO, V. Power law distribution: method of multi-scale inferential statistics. *Journal of Modern Mathematics Frontier*, v. 1, n. 1, p. 21-28, 2012.
- IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Regiões de influência das cidades*. Rio de Janeiro, 2007.
- _____. *Divisão urbano regional*. Rio de Janeiro, 2013.
- _____. *Gestão do território*. Rio de Janeiro, 2014.
- _____. *Arranjos populacionais e concentrações urbanas do Brasil*. Rio de Janeiro, 2015.
- JENKS, G. F. The data model concept in statistical mapping. *International Yearbook of Cartography*, v. 7, p. 186-190, 1967.

JIANG, B. Head/tail breaks: a new classification scheme for data with a heavy-tailed distribution. *The Professional Geographer*, v. 65, n. 3, p. 482-494, 2013.

_____. Geospatial analysis requires a different way of thinking: the problem of spatial heterogeneity. *GeoJournal*, v. 80, n. 1, p. 1-13, 2015.

JIANG, B.; LIU, X. Scaling of geographic space from the perspective of city and field blocks and using volunteered geographic information. *International Journal of Geographical Information Science*, v. 26, n. 2, p. 215-229, 2012.

JIANG, B.; MIAO, Y. The evolution of natural cities from the perspective of location-based social media. *The Professional Geographer*, v. 67, n. 2, p. 295-306, 2014.

JIANG, B.; YIN, J. Ht-index for quantifying the fractal or scaling structure of geographic features. *Annals of the Association of American Geographers*, v. 104, p. 530-541, 2014.

MA, D.; SANDBERG, M.; JIANG, B. Characterizing the heterogeneity of the OpenStreetMap data and community. *ISPRS International Journal of Geo-Information*, v. 4, n. 2, p. 535-550, 2015.

MAGNANINI, R.; LIMA, O. M. B. Uma medida da função da direção das cidades brasileiras. *Boletim Carioca de Geografia*, v. 22, p. 129-147, 1971.