

Big data de redes sociais pode prever focos de epidemias?¹

Alexandre Gori Maia (IE/UNICAMP); Jose Daniel Morales Martinez (IE/UNICAMP); Leticia Junqueira Marteleto (UT-Austin); Luiz Gustavo Sereno (IE/UNICAMP); Cristina Guimaraes Rodrigues (FIPE/USP)

Resumo

Na ausência de tratamentos farmacológicos eficazes, o monitoramento e intervenções não farmacológicas localizadas se mostraram políticas efetivas de combate à pandemia COVID-19. Nosso estudo combina big data do Twitter com dados administrativos dos casos de COVID-19 no Brasil para demonstrar que o rastreamento digital por *tweets* é uma ferramenta que pode ser usada para ajudar a prever o comportamento local da pandemia de COVID-19. Nós coletamos *tweets* com conteúdo associado à pandemia, medidas de segurança, vacinação, saúde e política. Em seguida, utilizamos modelos de dados espaciais para relacionar a frequência das palavras-chave selecionadas com a taxa de casos e mortes por COVID-19. Nossos resultados destacam que menções associadas ao comportamento social, além daquelas associadas à própria pandemia, são preditores importantes dos casos e mortes futuros. Nossa discussão final destaca como as redes sociais oferecem uma rica fonte de informação para políticas de monitoramento e contenção da pandemia.

Palavras-chave: twitter, análise de sentimentos, modelos de dados espaciais, comportamento social, políticas de saúde

Can big data from social networks predict epidemic outbreaks?

Abstract

In the absence of effective pharmacological treatment, monitoring and targeted nonpharmacologic interventions showed to be effective policies to contain the COVID-19 pandemic. Our study combines big data from Twitter with administrative data from COVID-19 cases in Brazil to demonstrate that digital tracking by tweets can predict the local behavior of the COVID-19 pandemic. We collect tweets with content associated with the pandemic, safety measures, vaccinations, health, and politics. We then used spatial data models to relate the frequency of selected keywords to the rate of cases and deaths from COVID-19. Our results highlight that mentions associated with social behavior, in addition to those associated with the pandemic itself, are significant predictors of future cases and deaths. Our final discussion highlights how social networks can provide a rich source of information for pandemic monitoring and containment policies.

Keywords: twitter, sentiment analysis, spatial data models, social behavior, public health

JEL: I12, I18, C23

ANPEC: Área 12 - Economia Social e Demografia Econômica

¹ Artigo submetido ao 49º Encontro Nacional de Economia, 6 a 10 de Dezembro de 2021.

Introdução

Na ausência de tratamento farmacológicos eficazes, diversas intervenções não farmacológicas (INFs) foram adotadas para conter a pandemia COVID-19 pelo mundo (Iezadi et al. 2020). Essas medidas incluem, por exemplo, o isolamento de casos, higienização e uso de máscaras, distanciamento social, fechamento de estabelecimentos públicos e comerciais, proibição de aglomerações e circulação de pessoas (Aquino et al. 2020). Entre essas, o fechamento de estabelecimentos e a proibição da circulação demonstraram ser uma das medidas mais eficazes no controle da epidemia (Haug et al. 2020). O problema é que muitas INFs geram custos sociais, econômicos e políticos, além de impor mudanças radicais nos hábitos e afetar a saúde das pessoas (Tsamakis et al. 2020). Para minimizar os impactos indesejáveis das INFs, é fundamental delimitar a sua amplitude espacial, o momento ideal de implementação (*timing*) e a duração das ações. Por exemplo, estudos indicam que intervenções administradas regionalmente, ao invés de ações locais isoladas, apresentam os melhores resultados na redução da pandemia (Gori Maia et al. 2021). A implementação de medidas de restrição de mobilidade nos momentos mais oportunos também se mostraram mais eficazes na redução do pico das hospitalizações (Oraby et al. 2021).

Uma condição essencial para definir a amplitude, *timing* e duração das INFs é dispor de dados em tempo real sobre a evolução da pandemia em cada localidade. Testes clínicos em massa seriam uma solução, mas nem sempre estão facilmente disponíveis para toda a população e seus resultados podem não ser obtidos em tempo hábil para conter a epidemia em seus estágios iniciais. Quando a disponibilidade de dados em tempo real é limitada, a coleta de dados por meio da raspagem (*data scraping*) de plataformas *online* pode oferecer uma rica quantidade de dados (*big data*) para orientar políticas de controle ou mitigação de pandemias (Hammer, Kostroch and Quiros 2017). Big data de redes sociais tem sido crescentemente utilizado para avaliar problemas sociais, incluindo pandemias (Salvatore, Biffignandi and Bianchi 2020). Signorini et al. (2011) utilizaram dados públicos do Twitter (*tweets*) para mapear a evolução do sentimento das pessoas em relação à gripe H1N1 e medir a atividade real da epidemia nos Estados Unidos. Souza et al. (2019) desenvolveram uma estratégia para identificar pessoas infectadas e áreas de risco de contaminação pela Dengue no Brasil a partir de *tweets* públicos georreferenciados. Marques-Toledo et al. (2017) identificaram forte associação positiva entre *tweets* com relacionados à palavra dengue com casos de dengue registrados oficialmente no Brasil.

Nosso estudo combina big data do Twitter com dados administrativos dos casos de COVID-19 no Brasil para demonstrar que o rastreamento digital por *tweets* pode ser capaz de prever o comportamento local da pandemia de COVID-19. Nós coletamos *tweets* com conteúdo associado à pandemia, medidas de segurança, vacinação, saúde e política. Em seguida, utilizamos modelos de dados espaciais para relacionar a frequência das palavras-chave selecionadas com a taxa de casos e mortes por COVID-19. Nossos resultados destacam que os *tweets* podem prever significativamente a evolução da pandemia localmente, sugerindo que as redes sociais oferecem uma rica fonte de informação para políticas de monitoramento e contenção da pandemia.

Material e Métodos

Fonte dos dados e variável dependente

Nossas análises utilizam dados de corte transversal para 557 microrregiões do Brasil. Os dados de casos e morte por COVID-19 são provenientes do Sistema de Informação de Vigilância Epidemiológica da Gripe – SIVEP Gripe (Ministério da Saúde 2020). O SIVEP-Gripe é um sistema de vigilância epidemiológica criado no ano de 2009 para combater a pandemia de influenza H1N1 e para monitorar o comportamento de eventos de alcance endêmico e sanitário, como a COVID-19. Coletamos o total de novos casos e mortes no mês de Dezembro de 2020 por município. Em seguida, agregamos os dados para as 557 microrregiões brasileiras. A partir dos agregados das microrregiões,

calculamos duas variáveis dependentes: casos de COVID-19 por cada 1000 habitantes e mortes de COVID-19 por cada 1000 habitantes. Em dezembro de 2020, a taxa média de casos por cada 1000 habitantes nas microrregiões brasileiras foi de 6.28, enquanto que a taxa média de óbitos diários por COVID-19 foi de 0.09 por cada 1000 habitantes (Tabela 1).

<Inserir Tabela 1>

No período analisado, as maiores taxas de casos e mortes por COVID-19 localizaram-se nas microrregiões do sul e sudeste do Brasil, situação que também se observa para algumas microrregiões do norte e do nordeste associadas a regiões metropolitanas (Figura 1). Esse padrão confirma a dinâmica de difusão espacial da doença, a qual ocorreu primeiramente nos centros mais populosos e densos (Rogério Silveira et al. 2020). Os centros urbanos das regiões sul e sudeste pertencem à principal região econômica e turística do país, com maior proporção de idosos (pessoas acima de 65 anos). Essa faixa populacional é a considerada de risco para o COVID-19, por apresentar maiores comorbidades e deficiências do sistema imunológico (Nunes et al. 2020).

< Inserir Figura 1>

Variáveis Independentes

Utilizamos um algoritmo em linguagem de programação Python para coletar todos os *tweets* de acesso público com informação da localização espacial (*geotag*) no mês de Novembro de 2020. A partir dessa localização espacial, identificamos o município de origem de cada *tweet*. O algoritmo foi executado aleatoriamente em diversos dias do mês de Novembro de 2020. Em cada execução, coletamos todos os *tweets* emitidos no dia com palavras-chave de interesse para nossas análises. Havia um limite de 15 minutos para cada coleta estabelecido pelo Twitter. Ao final da coleta, nossa base de dados continha 510.550 *tweets* para um total de 3.599 municípios (65% do total de municípios no país). Os *tweets* foram então agregados nas 557 microrregiões brasileiras.

Há diversas estratégias que podem ser utilizadas para identificar sentimentos expressos nas redes sociais (Salvatore et al. 2020). Nós utilizamos uma estratégia simples e intuitiva de análise lexical. A ideia é medir a intensidade de palavras-chave associadas aos principais temas mais debatidos no Twitter. Nossas variáveis de interesse referem-se às frequências dos *tweets* contendo palavras-chave associadas a cinco dimensões de análise: pandemia (*tweets* com palavras covid, coronavírus, vírus ou pandemia); medidas de segurança (máscara, quarentena, isolamento, distanciamento ou *lockdown*); vacinação (vacina); presidente (Bolsonaro, presidente ou presidência); saúde (saúde). Na média, 34% dos *tweets* continham palavras associadas à pandemia, 18,4% se relacionaram com medidas de segurança, 16% dos *tweets* trataram sobre vacina, 3,3% falaram sobre o presidente e sua presidência, e 15% dos *tweets* estiveram associados à temática da saúde (Tabela 1).

Também usamos as seguintes fontes de variáveis de controle: Índice de Desenvolvimento Municipal FIRJAN, para as dimensões de educação, emprego e renda; Pesquisa Municipal de Informações Básicas Municipais – MUNIC do Instituto Brasileiro de Geografia e Estatística (IBGE), para informações relacionadas ao sistema de saúde; MapBiomas (MapBiomas 2020), para informações relacionadas ao uso do solo nos municípios. As variáveis de controle são: nível de urbanização, variáveis socioeconômicas, autonomia local em saúde, número de leitos e de estabelecimentos de atenção média da rede do SUS (em logaritmo natural), vigilância e controle de endemias, e binárias para controle de regiões geográficas. Em média, 20% da área das microrregiões é coberta por

infraestrutura urbana; o índice de emprego e renda é de 0.46, enquanto que o índice de educação é de 0.76; 94% dos municípios administram diretamente os estabelecimentos de saúde; o número médio de leitos por microrregião é de 72,1; e o número médio do SUS é de 137,3 ambulatórios, 9,9 leitos de internação, 17,6 de urgência, e 40,1 de SADT (Serviço de Apoio à Diagnose e Terapia) (Tabela 1).

Estratégia Empírica

Nosso modelo de referência assume que a taxa de casos ou mortes por COVID-19 na i -ésima microrregião no mês de Dezembro de 2020 (Y_i) possa ser explicada pela frequência de palavras-chave no Twitter (vetor I_i , contendo frequências dos grupos de palavras-chave), que expressariam os sentimentos das pessoas do mês imediatamente anterior (Novembro de 2020) aos casos reportados, além de uma série de variáveis de controle (vetor x_i):

$$Y_i = \alpha + I_i \delta + x_i' \beta + e_i \quad (1)$$

O vetor δ contém os coeficientes dos indicadores de sentimentos e o vetor β os coeficientes das variáveis de controle. O termo α representa o intercepto que é constante para as microrregiões, e o erro aleatório é representado por e_i .

Uma limitação da equação (1) é supor que a transmissão da COVID-19 é independente da taxa de transmissão, das condições socioeconômicas e sentimentos expressos em outras microrregiões. Entretanto, estudos prévios já identificaram que a dinâmica regional influencia significativamente a situação sanitária em uma determinada localidade (Gori Maia et al. 2021; Castro et al. 2021). Para controlar a existência de dependência espacial nas análises, testamos diferentes especificações de modelos espaciais (Anselin 2001):

$$Y_i = \alpha + \rho w_i' Y_i + I_i \delta + x_i' \beta + e_i \quad (2)$$

$$Y_i = \alpha + I_i \delta + x_i' \beta + \xi_i, \xi_i = \lambda w_i' \xi_i + e_i \quad (3)$$

$$Y_i = \alpha + \rho w_i' Y_i + I_i \delta + x_i' \beta + \xi_i, \xi_i = \lambda w_i' \xi_i + e_i \quad (4)$$

$$Y_i = \alpha + I_i \delta + x_i' \beta + \pi_1 w_i' I_i + \pi_2 w_i' x_i' + e_i \quad (5)$$

$$Y_i = \alpha + \rho w_i' Y_i + I_i \delta + x_i' \beta + \pi_{j1} w_i' I_i + \pi_2 w_i' x_i' + e_i \quad (6)$$

$$Y_i = \alpha + I_i \delta + x_i' \beta + \pi_1 w_i' I_i + \pi_2 w_i' x_i' + \xi_i, \xi_i = \lambda w_i' \xi_i + e_i \quad (7)$$

$$Y_i = \alpha + \rho w_i' Y_i + I_i \delta + x_i' \beta + \pi_1 w_i' I_i + \pi_2 w_i' x_i' + \xi_i, \xi_i = \lambda w_i' \xi_i + e_i \quad (8)$$

Onde a matriz de vizinhança espacial w_i' contém valores positivos para microrregiões com fronteiras comuns, incluindo vértices (critério de contiguidade rainha), e zero caso contrário. A escolha da matriz obedece a Castro et al. (2021), que utiliza contiguidade de fronteira para estabelecer as relações de proximidade espacial municipal na análise da COVID-19 no Brasil. A matriz é normalizada pelo critério de linha, com a finalidade de suavizar os dados e facilitar a interpretação dos pesos espaciais (Almeida 2012).

A equação (2) representa o modelo SAR (*Spatial Autoregressive*) e assume a existência de *spillovers* epidemiológicos, ou seja, que as taxas de casos ou mortes de COVID-19 na microrregião i (Y_i) são influenciadas pelas taxas médias das microrregiões vizinhas ($w_i' Y_i$) (Anselin 2001). O coeficiente de autocorrelação espacial ρ representa o grau de dependência espacial. Para evitar comportamento explosivo, a restrição imposta sobre este coeficiente é que ele se situe no intervalo aberto entre -1 e 1 (LeSage, LeSage and James 2008).

A equação (3) representa o modelo SEM (*Spatial Error Model*), e assume a influência de características não observáveis na propagação da COVID-19, através do coeficiente de autocorrelação

espacial λ . Assim, os erros ξ_i são uma função da média dos erros da vizinhança ($\mathbf{w}'_i \xi_i$) mais um componente de erro aleatório (e_i). Para evitar comportamento instável, é imposta a restrição de que $|\lambda| < 1$ (Fingleton 2008). A equação (4) descreve o modelo SAC (*Spatial Autocorrelation*), representando a presença combinada dos modelos SAR e SEM.

A equação (5) apresenta o modelo SLX (*Spatially Lagged X-variables*), que testa a influência dos sentimentos e das características da vizinhança ($\mathbf{w}'_i \mathbf{I}_i$ e $\mathbf{w}'_i \mathbf{x}'_i$ respectivamente) no comportamento local da COVID-19 (Halleck Vega and Elhorst 2015). Os coeficientes π_1 e π_2 expressam o grau de associação entre as variáveis independentes das microrregiões vizinhas e a taxa de transmissão da COVID-19 na microrregião i . A equação (6) corresponde ao modelo SDM (*Spatial Durbin Model*), o qual propõe uma combinação dos modelos SAR e SLX. A equação (7) descreve o modelo SDEM (*Spatial Durbin Erro Model*), que incorpora conjuntamente a estrutura do modelo SEM e do SLX.

Finalmente, a equação (8) refere-se ao modelo GSM (*General Spatial Model*), o qual, segundo Almeida (2012), representa um processo envolvendo componentes de transbordamento de alcance global e local, ao combinar os três principais modelos da econometria espacial: SAR, SEM e SLX.

Estimamos a equação (1) pelo método *Ordinary Least Squares* (OLS) com estimador robusto à heterocedasticidade (White 1980). As equações (2) a (8) são estimadas usando *Quasi-Maximum Likelihood Estimate* (QMLE). A função log verossimilhança na estimação QMLE de modelos espaciais incorpora a matriz jacobiana, que representa as relações entre a variável dependente que é observada e o termo de erro aleatório que não é observado (Bell and Bockstael 2000). Essa matriz é necessária por causa do problema de reflexão de Manski (1993), ou seja, pela simultaneidade da interação entre as observações. Portanto, a estimação espacial por QMLE fornece estimativas consistentes e eficientes na presença de defasagens espaciais (Lee 2004).

Resultados

Padrões de associação

Os clusters LISA (*Local Indicator of Spatial Association*) (Anselin 2010) identificam grupos de microrregiões com indicadores semelhantes de transmissão da COVID-19 (Figura 1). Para a taxa de casos, a concentração espacial mais importante é de *clusters* tipo Alto-Alto (grupo de microrregiões com taxas acima da média). Esses *clusters* estão concentrados nas regiões sul, norte e nordeste. Conforme Sousa et al. (2020), a incidência de casos foi maior nas regiões mais pobres, norte e nordeste, e nos grandes centros urbanos do Brasil, região sul. Os *clusters* do tipo Baixo-Baixo (agrupamento de microrregiões com taxas abaixo da média) estão localizados principalmente nos estados de Minas Gerais e Maranhão. No caso da taxa de óbitos por 1000 habitantes, os agrupamentos Alto-Alto se encontram espalhados nas regiões sul, centro-oeste e norte, enquanto que as concentrações do tipo Baixo-Baixo estão localizadas majoritariamente na parte norte da região sudeste.

< Inserir Figura 2 >

As Tabelas 2.1 e 2.2 apresentam as estimativas para as equações (1) a (8). Os modelos para mortes por COVID-19 apresentam melhor qualidade de ajuste, com coeficiente de determinação variando entre 28.1% e 40.7% (entre 21.7% e 40.9% para taxa de casos). O teste de correlação espacial de Moran sugere a existência de dependência espacial entre as microrregiões. A defasagem espacial ($\mathbf{w}'_i \mathbf{Y}_i$) é estatisticamente significativa em todos os modelos, e a defasagem dos erros ($\mathbf{w}'_i \xi_i$) não é significativa apenas no modelo GSM. As defasagens espaciais das variáveis de interesse ($\mathbf{w}'_i \mathbf{I}_i$) e de controle ($\mathbf{w}'_i \mathbf{x}'_i$)

são também significativas na maioria dos modelos. Os testes de Wald e de razão de verossimilhança corroboram a melhor qualidade do ajuste dos modelos com inclusão da dependência espacial.

< Inserir Tabela 2.1 >

< Inserir Tabela 2.2 >

Nossos resultados indicam que os sentimentos das pessoas reportados nos *tweets* são bons preditores da dinâmica da COVID-19 no Brasil. As estimativas para a intensidade de *tweets* sobre a pandemia são significativas em todos os modelos espaciais: para cada aumento de 1 ponto percentual (p.p.) na proporção desses *tweets*, as taxas de casos (ou óbitos) por cada 1000 habitantes no mês seguinte crescem entre 1.46 e 4.46 (ou entre 0.34 e 0.53, respectivamente). Resultado semelhante é observado para as opiniões sobre medidas de segurança: para cada aumento de 1 p.p. na proporção desses *tweets*, as taxas de casos (ou óbitos) no mês seguinte crescem entre 1.64 e 2.65 (ou entre 0.40 e 0.61, respectivamente). As estimativas para os *tweets* associadas à vacina têm sinal negativo e significância estatística somente para a taxa de mortes. Esse resultado indica que, quanto maior a intensidade de sentimentos relacionados à vacinação, menor seria a taxa de óbitos no mês seguinte (entre -0.036 e -0.062).

As estimativas para os *tweets* associados à presidência são significativos somente nos modelos que incorporam defasagens exógenas da vizinhança, ou seja, os modelos SLX, SDM, SDEM e GSM. As estimativas sugerem que um aumento das menções sobre o presidente nos *tweets* estaria correlacionado com aumento na taxa de óbitos por 1000 habitantes no mês seguinte (estimativas positivas entre 0.10 e 0.13). Por sua vez, as estimativas relacionadas às opiniões sobre saúde são significativas na maioria dos modelos que abordam a autocorrelação espacial. Essas estimativas indicam que incrementos dessas opiniões estariam associados a menores taxas de casos confirmados de COVID-19 por cada 1000 habitantes (resultados negativos entre -2.69 e -7.60). Não encontramos evidências de que as menções sobre vacina e presidência afetem significativamente a taxa de casos, assim como nos impactos das opiniões sobre saúde na taxa de mortes pela COVID-19.

As estimativas para as defasagens espaciais $\mathbf{w}'_i \mathbf{Y}_{ij}$ e $\mathbf{w}'_i \boldsymbol{\xi}_{ij}$ indicam que: (i) existem *spillovers* epidemiológicos, ou seja, os indicadores locais dependem do comportamento da pandemia nas microrregiões vizinhas; (ii) fatores não observáveis ou de difícil mensuração da vizinhança influenciariam na dinâmica local da COVID-19. Os resultados para as defasagens das variáveis de interesse $\mathbf{w}'_i \mathbf{I}_i$ sugerem ainda transbordamentos espaciais das opiniões sobre a pandemia, ou seja, opiniões na vizinhança sobre a pandemia também estão positivamente associadas às taxas locais de casos e óbitos (entre 5.45 e 16.12, e 0.08 e 0.19, respectivamente).

As estimativas para as defasagens das variáveis de controle $\mathbf{w}'_i \mathbf{x}'_i$ apontam que: (i) a gestão municipal dos estabelecimentos de saúde nas microrregiões vizinhas estaria negativamente associada a taxa local de óbitos pela COVID-19 (entre -0.08 e -0.13); (ii) o nível de emprego e renda das microrregiões mais próximas influenciaria negativamente as taxas locais de casos confirmados da doença (entre -4.01 e -10.60). Por fim, de acordo com o critério de informação do BIC, os melhores modelos para as taxas de casos e óbitos são os SDM. De agora em diante, restringiremos nossas análises a esses modelos.

A Tabela A1 do Anexo A apresenta as estimativas completas incluindo as variáveis de controle. Os resultados sugerem que a cobertura de infraestrutura urbana e a autonomia administrativa local dos estabelecimentos de saúde estariam negativamente associados às taxa de casos de COVID-19. Por outro lado, o nível de emprego e renda, a educação, o número de leitos e de estabelecimentos ambulatoriais do SUS estariam positivamente correlacionados com a taxa de casos COVID-19. Em relação às mortes pela COVID-19, a autonomia administrativa local dos estabelecimentos de saúde e

o número de estabelecimentos de internação do SUS estariam negativamente associados à taxa de óbitos, enquanto que o número de leitos, de estabelecimentos ambulatoriais e do SAT do SUS estariam positivamente associados.

Finalmente, testamos a robustez das estimativas do modelo SDM utilizando o método *generalized spatial two-stage least-squares* (gs2sls) proposto por Kelejian e Prucha (1999) e Drukker, Egger and Prucha (2013). Nessas estimativas usamos uma matriz de ponderação espacial *Rook*² de segunda ordem³. As pequenas diferenças na magnitude das estimativas (Tabela A1 do Anexo A) sugerem a robustez de nossos resultados tanto a diferentes métodos de estimação quanto a diversos regimes espaciais.

Decomposição de efeitos espaciais

Fizemos ainda a decomposição das estimativas dos modelos SDM em efeitos espaciais diretos, indiretos e totais (Tabela 3). Essa estratégia considera que mudanças nas variáveis explicativas do modelo podem estar associadas ao comportamento da COVID-19 tanto na microrregião *i* quanto nas microrregiões vizinhas (LeSage and Pace 2009). O efeito direto representa o efeito das variáveis independentes localmente, ou seja, na microrregião *i*. O efeito indireto representa a influência das variáveis explicativas da vizinhança, através do *spillover* epidemiológico. O efeito total é simplesmente a somatória dos efeitos diretos e indiretos. Estamos particularmente interessados em analisar os efeitos indiretos, que representam a influência de ocorrências na vizinhança.

< Inserir Tabela 3 >

Em relação à taxa de casos de COVID-19, os efeitos indiretos representam 83% do efeito total das opiniões sobre a pandemia (21.66) e aproximadamente 63% do efeito total das opiniões sobre medidas de segurança (3.6) e saúde (-5.26). Percentuais semelhantes foram encontrados para o efeito indireto das variáveis de controle: urbanização (-8.99), educação (11.86), estabelecimentos de saúde de gestão local (-5.56), e número de estabelecimentos ambulatoriais do SUS (2.26). O efeito indireto da variável emprego e renda apresentou sinal negativo diferente ao efeito direto, indicando que a influência dessa variável da vizinhança ocorre em sentido contrário ao efeito causado na microrregião.

Nos modelos para taxa de óbitos de COVID-19, os maiores efeitos indiretos referem-se aos *tweets* sobre a pandemia (0.22, ou 79% do efeito total) e da variável de controle estabelecimentos de saúde com gestão local (-0.22, ou 76% do efeito total). Os efeitos indiretos representam aproximadamente 48% do efeito total das variáveis: medidas de segurança (0.05), vacina (-0.05), presidente (0.10), número de leitos (0.01), estabelecimentos ambulatoriais do SUS (0.02), de internação do SUS (-0.01), e de SADT do SUS (0.01).

² O nome faz menção ao movimento da peça torre no jogo de xadrez. A matriz torre considera vizinhos apenas às regiões com extensão de fronteira diferente de zero, sem levar em consideração vértices.

³ É composta pelos vizinhos dos vizinhos da primeira ordem (vizinhos diretos). Ou seja, o vizinho *k* de *j* será vizinho de segunda ordem de *i*, se *i* e *j* forem vizinhos.

Discussão

Redes sociais e plataformas digitais, como Twitter, Facebook, e Google, oferecem uma rica base de informações sobre fenômenos sociais. Milhões de pessoas pesquisam, compartilham e trocam opiniões, sentimentos e expressam comportamentos sobre problemas diversos, incluindo preocupações com as condições de saúde. Nosso estudo demonstrou que o acompanhamento de dados digitais da rede social Twitter oferece uma ferramenta valiosa para monitoramento e orientação de políticas localizadas para contenção de pandemias.

Os dados da rede social Twitter já demonstraram a capacidade de prever casos de Dengue no Brasil (Castro et al. 2021; Marques-Toledo et al. 2017). Uma contribuição importante do nosso trabalho é demonstrar que menções associadas a comportamentos sociais, além das palavras associadas à própria epidemia, são também preditores importantes dos casos e mortes futuros. As relações mais significativas são aquelas com as menções à pandemia. Entretanto, menções às medidas de segurança, vacinação, saúde e até ao posicionamento político também apresentam relação significativa com casos e mortes futuras. A relação positiva dos casos e mortes com as menções às medidas de segurança, como uso de máscaras e distanciamento social, podem expressar tanto a preocupação da população com o avanço da epidemia quanto a falta de confiança nas instituições e nos governantes (Bavel et al. 2020). As relações negativas para as menções à vacinação e saúde podem revelar tanto a percepção de risco à saúde quanto a compreensão dos benefícios sociais da vacinação em massa (Trueblood, Sussman and O’Leary 2021). Finalmente, as relações positivas entre as menções à presidência e os casos ou mortes futuros podem revelar o comportamento de risco de seus seguidores políticos em relação às ameaças da pandemia (Cabral, Pongeluppe and Ito 2021).

Outra contribuição importante deste artigo é aplicar uma estratégia empírica capaz de identificar e controlar o efeito de transbordamento espacial da pandemia e do comportamento social. Reforçando evidências de estudos anteriores (Gori Maia et al. 2021), os resultados de nossos modelos de dependência espacial demonstram que a dinâmica local da epidemia está fortemente associada à dinâmica na vizinhança. Ou seja, esforços para combater a epidemia localmente terão pouco resultado se as ações não forem conjugadas regionalmente. Nossos resultados também destacam que sentimentos expressos na vizinhança em relação à pandemia também estão fortemente associados à dinâmica da pandemia localmente. A mobilidade das pessoas e dos tweets pode explicar parcialmente esses resultados. Mas as redes sociais podem ainda expressar sentimentos com o que acontece na vizinhança ou prever a transmissão espacial da pandemia.

O monitoramento de epidemias é um problema recorrente das políticas de saúde no Brasil. Na última década, o país já foi severamente afetado pelas epidemias de Dengue, Zika e COVID-19. Enquanto testes clínicos não estão facilmente disponíveis para aplicação em massa na população, dados de redes sociais e plataformas digitais podem ajudar a orientar os esforços de contenção da pandemia em regiões endêmicas. Nossos resultados demonstraram que o uso de palavras-chave simples foi capaz de prever significativamente focos futuros de casos e mortes por COVID-19. Um possível desdobramento desse estudo seria aprimorar a estratégia de análise, por exemplo, com algoritmos de aprendizado de máquina para identificar sentimentos positivos ou negativos em relação a cada tema. Mas, além da complexidade operacional e analítica, essas estratégias exigem, muitas vezes, decisões subjetivas sobre os padrões de associação. Nosso estudo aplicou uma estratégia simples, intuitiva, de baixo custo e, acima de tudo, efetiva, para análise de dados públicos e gratuitos coletados na rede social Twitter.

Agradecimentos

Pesquisa financiada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, processos números 401935/2020-5 e 303397/2018-7). Os autores agradecem o suporte de apoio técnico de Chloe Brown, bolsista do *MIT Brazil Remote Internship initiative*.

Referências

- Almeida, E. 2012. *Econometria Espacial Aplicada*.
- Anselin, L. 2010. “Local Indicators of Spatial Association-LISA.” *Geographical Analysis* 27(2):93–115. Available at: <http://doi.wiley.com/10.1111/j.1538-4632.1995.tb00338.x>.
- Anselin, L. 2001. “Spatial econometrics.” In B. H. Baltagi, ed. *A companion to theoretical econometrics*. New York: Blackwell Publishing Ltd, pp. 310–330.
- Aquino, E.M.L., I.H. Silveira, J.M. Pescarini, R. Aquino, J.A. de Souza-Filho, A. dos S. Rocha, A. Ferreira, A. Victor, C. Teixeira, D.B. Machado, E. Paixão, F.J.O. Alves, F. Pilecco, G. Menezes, L. Gabrielli, L. Leite, M. da C.C. de Almeida, N. Ortelan, Q.H.R.F. Fernandes, R.J.F. Ortiz, R.N. Palmeira, E.P. Pinto Junior, E. Aragão, L.E.P.F. de Souza, M. Barral Netto, M.G. Teixeira, M.L. Barreto, M.Y. Ichihara, and R.T. dos R.S. Lima. 2020. “Medidas de distanciamento social no controle da pandemia de COVID-19: potenciais impactos e desafios no Brasil.” *Ciência & Saúde Coletiva* 25(suppl 1):2423–2446. Available at: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-81232020006702423&tlng=pt.
- Bavel, J.J. Van, K. Baicker, P.S. Boggio, V. Capraro, A. Cichocka, M. Cikara, M.J. Crockett, A.J. Crum, K.M. Douglas, J.N. Druckman, J. Drury, O. Dube, N. Ellemers, E.J. Finkel, J.H. Fowler, M. Gelfand, S. Han, S.A. Haslam, J. Jetten, S. Kitayama, D. Mobbs, L.E. Napper, D.J. Packer, G. Pennycook, E. Peters, R.E. Petty, D.G. Rand, S.D. Reicher, S. Schnall, A. Shariff, L.J. Skitka, S.S. Smith, C.R. Sunstein, N. Tabri, J.A. Tucker, S. van der Linden, P. van Lange, K.A. Weeden, M.J.A. Wohl, J. Zaki, S.R. Zion, and R. Willer. 2020. “Using social and behavioural science to support COVID-19 pandemic response.” *Nature Human Behaviour* 4(5):460–471. Available at: <http://www.nature.com/articles/s41562-020-0884-z>.
- Bell, K.P., and N.E. Bockstael. 2000. “Applying the generalized-moments estimation approach to spatial problems involving microlevel data.” *Review of Economics and Statistics* 82(1):72–82.
- Cabral, S., L. Pongeluppe, and N. Ito. 2021. “The Disastrous Effects of Leaders in Denial: Evidence from the COVID-19 Crisis in Brazil.” *SSRN Electronic Journal*. Available at: <https://www.ssrn.com/abstract=3836147>.
- Castro, R.R., R.S.C. Santos, G.J.B. Sousa, Y.T. Pinheiro, I.M. Martins, M.L.D. Pereira, and R.A.R. Silva. 2021. “Spatial dynamics of the COVID-19 pandemic in Brazil.” *Epidemiology and Infection* 149.
- Drukker, D.M., P. Egger, and I.R. Prucha. 2013. “On Two-Step Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances and Endogenous Regressors.” *Econometric Reviews* 32(5–6):686–733.
- Fingleton, B. 2008. “A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors.” *Spatial Economic Analysis* 3(1):27–44.
- Gori Maia, A., L. Marteleto, C.G. Rodrigues, and L.G. Sereno. 2021. “The short-term impacts of coronavirus quarantine in São Paulo: The health-economy trade-offs” C. A. Zúniga-González, ed. *PLOS ONE* 16(2):e0245011. Available at: <https://dx.plos.org/10.1371/journal.pone.0245011>.
- Halleck Vega, S., and J.P. Elhorst. 2015. “The SLX Model.” *Journal of Regional Science* 55(3):339–363.
- Hammer, C., D. Kostroch, and G. Quiros. 2017. “Big Data: Potential, Challenges and Statistical Implications.” *Staff Discussion Notes* 17(06):1.

- Haug, N., L. Geyrhofer, A. Londei, E. Dervic, A. Desvars-Larrive, V. Loreto, B. Piniór, S. Thurner, and P. Klimek. 2020. "Ranking the effectiveness of worldwide COVID-19 government interventions." *Nature Human Behaviour* 4(12):1303–1312. Available at: <http://www.nature.com/articles/s41562-020-01009-0>.
- Iezadi, S., S. Azami-Aghdash, A. Ghiasi, A. Rezapour, H. Pourasghari, F. Pashazadeh, and K. Gholipour. 2020. "Effectiveness of the non-pharmaceutical public health interventions against COVID-19; a protocol of a systematic review and realist review" W.-J. Tu, ed. *PLOS ONE* 15(9):e0239554. Available at: <https://dx.plos.org/10.1371/journal.pone.0239554>.
- Kelejian, H.H., and I.R. Prucha. 1999. "A generalized moments estimator for the autoregressive parameter in a spatial model." *International Economic Review* 40(2):509–533.
- Lee, L.-F. 2004. "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models." *Econometrica* 72(6):1899–1925.
- LeSage, J.P., LeSage, and James. 2008. "An Introduction to Spatial Econometrics." *Revue d'économie industrielle* n° 123(3):19–44.
- LeSage, J.P., and R.K. Pace. 2009. *Introduction to spatial econometrics*. CRC Press.
- Manski, C.F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies* 60(3):531.
- MapBiomas. 2020. "Coleção 5 da Série Anual de Mapas de Cobertura e Uso de Solo do Brasil."
- Marques-Toledo, C. de A., C.M. Degener, L. Vinhal, G. Coelho, W. Meira, C.T. Codeço, and M.M. Teixeira. 2017. "Dengue prediction by the web: Tweets are a useful tool for estimating and forecasting Dengue at country and city level" B. Althouse, ed. *PLOS Neglected Tropical Diseases* 11(7):e0005729. Available at: <https://dx.plos.org/10.1371/journal.pntd.0005729>.
- Ministério da Saúde. 2020. "Sistema de Informação de Vigilância Epidemiológica da Gripe - Sivep Gripe."
- Nunes, B.P., A.S.S. De Souza, J. Nogueira, F.B. De Andrade, E. Thumé, D.S. Da Cruz Teixeira, M.F. Lima-Costa, L.A. Facchini, and S.R. Batista. 2020. "Multimorbidity and population at risk for severe COVID-19 in the Brazilian Longitudinal Study of Aging." *Cadernos de Saude Publica* 36(12):e00129620.
- Oraby, T., M.G. Tyshenko, J.C. Maldonado, K. Vatcheva, S. Elsaadany, W.Q. Alali, J.C. Longenecker, and M. Al-Zoughool. 2021. "Modeling the effect of lockdown timing as a COVID-19 control measure in countries with differing social contacts." *Scientific Reports* 11(1):3354. Available at: <http://www.nature.com/articles/s41598-021-82873-2>.
- Rogério Silveira, M., N.F. Felipe Junior, R. Giraldo Cocco, R. Matos Felácio, and L. Azeredo Rodrigues. 2020. "Novo coronavírus (Sars-CoV-2): difusão espacial e outro patamar para a socialização dos investimentos no Brasil." *Revista Brasileira de Estudos Urbanos e Regionais* 22.
- Salvatore, C., S. Biffignandi, and A. Bianchi. 2020. "Social Media and Twitter Data Quality for New Social Indicators." *Social Indicators Research*. Available at: <http://link.springer.com/10.1007/s11205-020-02296-w>.
- Signorini, A., A.M. Segre, and P.M. Polgreen. 2011. "The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic" A. P. Galvani, ed. *PLoS ONE* 6(5):e19467. Available at: <https://dx.plos.org/10.1371/journal.pone.0019467>.
- Sousa, G.J.B., T.S. Garces, V.R.F. Cestari, T.M.M. Moreira, R.S. Florêncio, and M.L.D. Pereira. 2020. "Estimation and prediction of COVID-19 cases in brazilian metropolises." *Revista Latino-*

Americana de Enfermagem 28:1–8.

- Souza, R., D.B. Neill, R.M. Assuncao, and W. Meira, Jr. 2019. “Identifying High-Risk Areas for Dengue Infection Using Mobility Patterns on Twitter.” *Online Journal of Public Health Informatics* 11(1). Available at: <https://journals.uic.edu/ojs/index.php/ojphi/article/view/9754>.
- Trueblood, J.S., A.B. Sussman, and D. O’Leary. 2021. “The Role of Risk Preferences in Responses to Messaging About COVID-19 Vaccine Take-Up.” *Social Psychological and Personality Science*:194855062199962. Available at: <http://journals.sagepub.com/doi/10.1177/1948550621999622>.
- Tsamakis, K., M. Gavriatopoulou, D. Schizas, A. Stravodimou, A. Mougkou, D. Tsiptsios, V. Sioulas, E. Spartalis, A. Sioulas, C. Tsamakis, N. Charalampakis, C. Mueller, D. Arya, P. Zarogoulidis, D. Spandidos, M. Dimopoulos, C. Papageorgiou, and E. Rizo. 2020. “Oncology during the COVID-19 pandemic: challenges, dilemmas and the psychosocial impact on cancer patients (Review).” *Oncology Letters* 20(1):441–447. Available at: <http://www.spandidos-publications.com/10.3892/ol.2020.11599>.
- White, H. 1980. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica* 48(4):817.

Tabelas

Tabela 1. Estimativas completas e teste de robustez de modelos SDM para taxa de casos e óbitos de COVID-19 por cada 1000 habitantes. Microrregiões – Brasil.

Variáveis	Descrição	Fonte	Média	Desvio Padrão
<i>Dependentes</i>				
Casos de COVID-19	Taxa por 1000 habitantes	Ministério da Saúde	6.282	4.496
Mortes por COVID-19	Taxa por 1000 habitantes		0.086	0.068
<i>Independentes de Análise (Tweets)</i>				
Pandemia	% de Tweets sobre Covid, pandemia e vírus	Twitter	0.339	0.165
Medidas de segurança	% de Tweets sobre máscara, isolamento e distanciamento	Twitter	0.184	0.124
Vacina	% de Tweets sobre vacinas	Twitter	0.158	0.124
Presidente	% de Tweets sobre o presidente e sua presidência	Twitter	0.033	0.065
Saúde	% de Tweets sobre saúde	Twitter	0.151	0.084
<i>Independentes de Controle</i>				
Urbanização	% de área da microrregião com infraestrutura urbana	MapBiomias	0.020	0.057
Emprego e renda	Índice de 0 a 1	FIRJAN	0.464	0.081
Educação	Índice de 0 a 1	FIRJAN	0.759	0.104
Estabelecimentos de saúde de gestão local	% de municípios da microrregião que administram diretamente os estabelecimentos de saúde	MUNIC	0.936	0.095
Leitos	Número total de leitos na microrregião	Data SUS	72.106	238.624
Estabelecimentos ambulatoriais SUS	Número de estabelecimentos ambulatoriais do SUS	Data SUS	137.325	132.855
Estabelecimentos internação SUS	Número de estabelecimentos de internação do SUS	Data SUS	9.946	14.291
Estabelecimentos urgência SUS	Número de estabelecimento de urgência do SUS	Data SUS	17.646	26.848
Estabelecimentos SADT SUS	Número de estabelecimentos para SADT do SUS	Data SUS	40.135	63.674
Vigilância de endemias	% de municípios da microrregião que realizam vigilância de endemias	MUNIC	0.960	0.080
Norte e Nordeste	Binária: 1 se a microrregião está nas regiões norte ou nordeste, e 0 em outro caso.	IBGE	0.451	0.498
Sul e Sudeste	Binária: 1 se a microrregião está nas regiões sul ou sudeste, e 0 em outro caso.	IBGE	0.456	0.499

Fonte: Twitter, Ministério da Saúde, MapBiomias, FIRJAN, MUNIC, Data SUS e IBGE.

Tabela 2-1. Estimativas de modelos OLS, SAR, SEM e SAC para as variáveis de interesse relacionadas às taxas de casos e óbitos de COVID-19 por cada 100 habitantes. Microrregiões – Brasil.

Variáveis	OLS		SAR		SEM		SAC		
	Casos	Óbitos	Casos	Óbitos	Casos	Óbitos	Casos	Óbitos	
Opiniões e sentimentos	Pandemia	5.426*** (1.016)	0.072*** (0.015)	2.258** (0.833)	0.042** (0.015)	1.650* (0.792)	0.034* (0.014)	1.485* (0.713)	0.041** (0.014)
	Medidas de segurança	2.978* (1.456)	0.059** (0.022)	1.796+ (1.036)	0.042* (0.018)	2.278* (1.018)	0.043* (0.018)	2.138* (0.919)	0.040* (0.017)
	Vacina	-2.007 (1.360)	-0.072** (0.022)	-1.056 (0.960)	-0.037* (0.017)	-0.282 (0.977)	-0.026 (0.018)	-0.002 (0.882)	-0.036* (0.016)
	Presidência	2.584 (2.169)	0.135*** (0.040)	-0.507 (0.905)	0.007 (0.016)	0.416 (0.892)	0.022 (0.016)	0.488 (0.808)	0.003 (0.015)
	Saúde	-2.559 (1.585)	-0.024 (0.028)	-6.026* (2.481)	0.041 (0.044)	-7.598* (2.960)	0.042 (0.052)	-6.399* (2.734)	0.040 (0.038)
Defasagens Espaciais	w'y [ρ]			0.707*** (0.033)	0.581*** (0.041)			-0.508*** (0.093)	0.713*** (0.057)
	w' ξ [λ]					0.755*** (0.031)	0.620*** (0.041)	0.896*** (0.024)	-0.294* (0.126)
	w'x [pandemia]								
	w'x [Estab de saúde de gestão municipal]								
	w'x [Emprego e renda]								
Testes de Diagnóstico	Moran	347.81***	147.77***						
	Wald [ρ]			487.95***	194.69***	624.30***	222.66***	2265.36***	333.68***
	BIC	3177.993	-1512.611	2893.94	-1642.818	2876.858	-1639.445	2864.569	-1639.49
	LR			296.70***	142.85***	313.78***	139.48***	332.39***	145.85***
	Pseudo – R ²	0.289	0.324	0.294	0.328	0.222	0.281	0.217	0.345
Número de Observações	557	557	557	557	557	557	557	557	

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Fonte: Twitter, Ministério da Saúde, MapBiomias, FIRJAN, MUNIC, Data SUS e IBGE.

Tabela 2-2. Estimativas de modelos OLS, SAR, SEM e SAC para as variáveis de interesse relacionadas às taxas de casos e óbitos de COVID-19 por cada 100 habitantes. Microrregiões – Brasil.

Variáveis	SLX		SDM		SDEM		GSM		
	Casos	Óbitos	Casos	Óbitos	Casos	Óbitos	Casos	Óbitos	
Opiniões e sentimentos	Covid e pandemia	4.457*** (0.958)	0.061*** (0.014)	2.484*** (0.727)	0.042** (0.013)	3.185*** (0.838)	0.053*** (0.014)	2.588** (0.850)	0.037** (0.013)
	Medidas de segurança	2.650* (1.256)	0.061** (0.019)	1.978* (0.945)	0.049** (0.017)	1.846* (0.941)	0.042* (0.017)	1.641* (0.837)	0.050** (0.016)
	Vacina	-1.628 (1.300)	-0.062** (0.020)	-0.501 (0.979)	-0.046** (0.017)	0.116 (1.025)	-0.045* (0.018)	0.359 (0.917)	-0.044** (0.017)
	Presidência da República	1.708 (2.490)	0.125*** (0.038)	0.627 (1.874)	0.102** (0.033)	0.114 (1.944)	0.110** (0.035)	0.037 (1.745)	0.097** (0.032)
	Saúde	-1.814 (1.848)	-0.016 (0.028)	-2.693+ (1.392)	-0.017 (0.025)	-4.060** (1.375)	-0.008 (0.025)	-3.678** (1.226)	-0.020 (0.024)
Defasagens Espaciais	w'y [ρ]			0.678*** (0.035)	0.518*** (0.044)			-0.533*** (0.095)	0.610*** (0.077)
	w' ξ [λ]					0.726*** (0.033)	0.547*** (0.045)	0.895*** (0.024)	-0.170 (0.139)
	w'x [Covid e pandemia]	16.122*** (1.820)	0.194*** (0.028)	5.843*** (1.427)	0.092*** (0.025)	7.339*** (2.117)	0.116*** (0.033)	5.450* (2.252)	0.081** (0.026)
	w'x [Estab de saúde de gestão municipal]		-0.133*** (0.033)		-0.086** (0.029)		-0.084* (0.034)		-0.080** (0.029)
	w' x [Emprego e renda]	-9.448** (3.353)		-10.596*** (2.525)		-4.041 (3.635)		4.898 (4.391)	
Testes de Diagnóstico	Moran								
	Wald [ρ]	84.02***	60.45***	494.00***	200.20***	488.77***	160.35***	2145.77***	271.93***
	BIC	3118.71	-1551.04	2874.68	-1648.87	2878.95	-1639.21	2868.45	-1643.7
	LR	78.26***	57.39***	328.61***	161.55***	324.33***	151.89***	341.15***	162.70***
	Pseudo – R ²	0.382	0.390	0.409	0.405	0.321	0.373	0.263	0.407
Número de Observações	557	557	557	557	557	557	557	557	

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Fonte: Twitter, Ministério da Saúde, MapBiomias, FIRJAN, MUNIC, Data SUS e IBGE.

Tabela 3. Decomposição de efeitos espaciais dos modelos SDM relacionadas às taxas de casos e óbitos de COVID-19 por cada 100 habitantes. Microrregiões – Brasil.

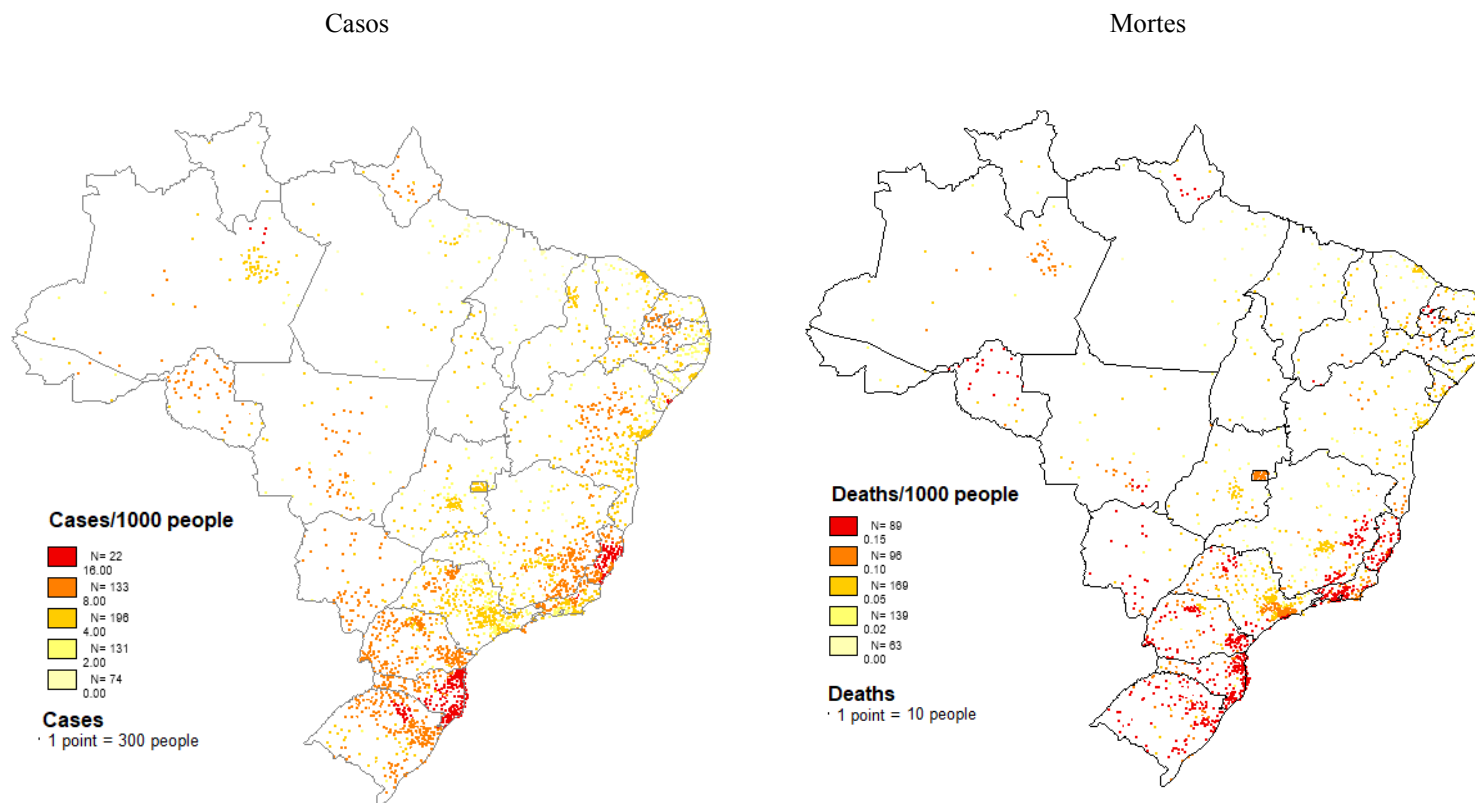
Variáveis	Decomposição de Efeitos Espaciais - Modelos SDM						
	Casos			Óbitos			
	Direto	Indireto	Total	Direto	Indireto	Total	
Opiniões e sentimentos	Pandemia	4.149***	21.662***	25.811***	0.057***	0.219***	0.277***
		(0.842)	(4.187)	(4.684)	(0.014)	(0.048)	(0.054)
	Medidas de segurança	2.275*	3.862*	6.138*	0.053**	0.049**	0.102**
		(1.088)	(1.939)	(3)	(0.018)	(0.019)	(0.036)
	Vacina	-0.576	-0.979	-1.555	-0.05**	-0.046*	-0.096**
	(1.125)	(1.913)	(3.036)	(0.019)	(0.019)	(0.037)	
Presidente	0.721	1.225	1.946	0.11**	0.102**	0.212**	
	(2.156)	(3.665)	(5.819)	(0.036)	(0.037)	(0.071)	
Saúde	-3.097+	-5.258+	-8.356+	-0.018	-0.017	-0.035	
	(1.6)	(2.811)	(4.376)	(0.026)	(0.025)	(0.051)	
Urbanização	-5.294+	-8.987+	-14.281+	0.066	0.062	0.128	
	(2.742)	(4.808)	(7.491)	(0.045)	(0.043)	(0.088)	
Emprego e renda	8.011**	-12.913+	-4.902	0.0002	0.0002	0.0003	
	(2.318)	(7)	(8.285)	(0.038)	(0.036)	(0.074)	
Educação	6.988**	11.862**	18.85**	0.008	0.007	0.015	
	(2.262)	(4.098)	(6.238)	(0.037)	(0.035)	(0.072)	
Estab de saúde de gestão local	-3.274*	-5.558*	-8.832*	-0.07**	-0.221***	-0.291***	
	(1.404)	(2.508)	(3.869)	(0.023)	(0.062)	(0.076)	
Variáveis de Controle	ln Leitos	0.283*	0.48*	0.762*	0.006**	0.005**	0.011**
		(0.116)	(0.21)	(0.322)	(0.002)	(0.002)	(0.004)
	ln Estab ambulatório SUS	1.333**	2.263**	3.596**	0.016*	0.015*	0.03*
		(0.467)	(0.84)	(1.285)	(0.008)	(0.007)	(0.015)
	ln Estab internação SUS	-0.513	-0.871	-1.384	-0.015*	-0.014*	-0.03*
		(0.368)	(0.636)	(1)	(0.006)	(0.006)	(0.012)
	ln Estab urgência SUS	-0.457	-0.776	-1.233	-0.007	-0.006	-0.013
	(0.362)	(0.619)	(0.977)	(0.006)	(0.006)	(0.012)	
ln Estab SADT SUS	0.195	0.331	0.526	0.01+	0.009+	0.018+	
	(0.34)	(0.58)	(0.92)	(0.006)	(0.005)	(0.011)	
Vigilância de endemias	-0.904	-1.535	-2.44	-0.011	-0.01	-0.021	
	(1.649)	(2.811)	(4.457)	(0.027)	(0.026)	(0.053)	
Norte e Nordeste	0.724	1.23	1.954	0.003	0.002	0.005	
	(0.623)	(1.073)	(1.691)	(0.01)	(0.009)	(0.019)	
Sul e Sudeste	0.365	0.619	0.984	0.01	0.01	0.02	
	(0.538)	(0.9)	(1.437)	(0.009)	(0.008)	(0.017)	

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Fonte: Twitter, Ministério da Saúde, MapBiomias, FIRJAN, MUNIC, Data SUS e IBGE.

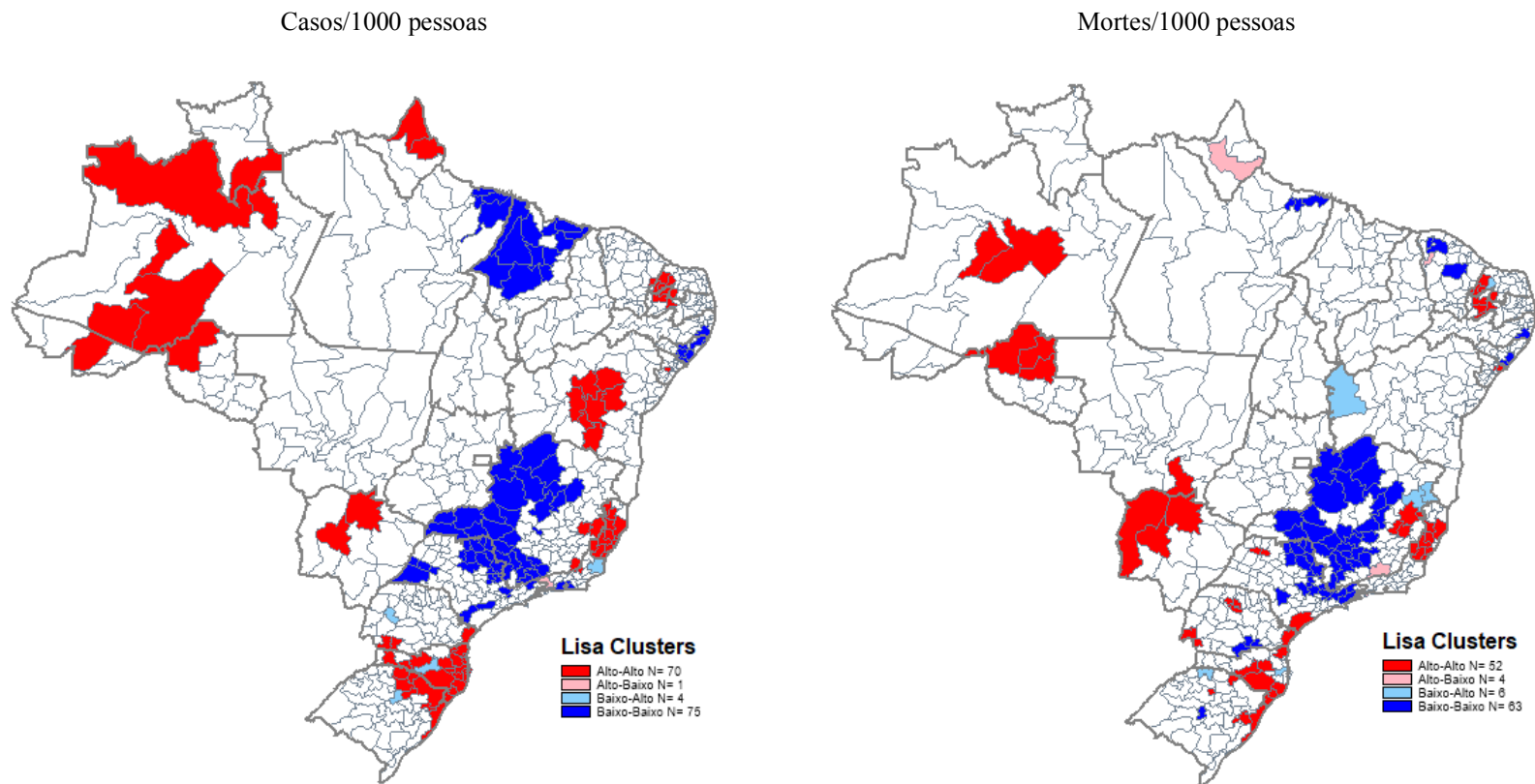
Figuras

Figura 1. Número total (pontos) e taxas de casos e óbitos COVID-19 por 1000 habitantes (cores). Microrregiões – Brasil, Dezembro de 2020.



Fonte: SIVEP-Gripe, Ministério da Saúde.

Figura 2. LISA clusters para Taxa de casos e óbitos por cada 1000 habitantes de COVID-19. Microrregiões – Brasil.



Fonte: SIVEP-Gripe, Ministério da Saúde.

Anexo A

Tabela A1. Estimativas completas e teste de robustez de modelos SDM para taxa de casos e óbitos de COVID-19 por cada 1000 habitantes. Microrregiões – Brasil.

Variáveis	SDM \emptyset		SDM Ω			
	Casos	Óbitos	Casos	Óbitos		
Opiniões e sentimentos	Pandemia	2.484*** (0.727)	0.042** (0.013)	2.353** (0.775)	0.039** (0.013)	
	Medidas de segurança	1.978* (0.945)	0.049** (0.017)	2.092* (0.984)	0.046** (0.017)	
	Vacina	-0.501 (0.979)	-0.046** (0.017)	-0.571 (1.022)	-0.051** (0.018)	
	Presidente	0.627 (1.874)	0.102** (0.033)	1.404 (1.950)	0.114*** (0.034)	
	Saúde	-2.693+ (1.392)	-0.017 (0.025)	-2.441+ (1.452)	-0.011 (0.025)	
Defasagens Espaciais	w'y [ρ]	0.678*** (0.035)	0.518*** (0.044)	0.869*** (0.077)	0.702*** (0.098)	
	w'x [Pandemia]	5.843*** (1.427)	0.092*** (0.025)	3.444+ (1.961)	0.076* (0.036)	
	w'x [Estab de saúde de gestão municipal]		-0.086** (0.029)		-0.112*** (0.014)	
	w'x [Emprego e renda]	-10.596*** (2.525)		-15.408*** (1.480)		
Variáveis de Controle	Urbanização	-4.603+ (2.387)	0.062 (0.042)	-5.209* (2.534)	0.021 (0.044)	
	Emprego e renda	9.004*** (2.137)	0.000 (0.036)	10.578*** (2.080)	-0.001 (0.036)	
	Educação	6.076** (1.977)	0.007 (0.035)	5.606** (2.042)	0.010 (0.035)	
	Estab de saúde de gestão local	-2.847* (1.222)	-0.055* (0.022)	-3.005* (1.282)	-0.051* (0.022)	
	ln Leitos	0.246* (0.101)	0.005** (0.002)	0.178+ (0.103)	0.005** (0.002)	
	ln Estab ambulatório SUS	1.159** (0.408)	0.015* (0.007)	1.050* (0.430)	0.016* (0.007)	
	ln Estab internação SUS	-0.446 (0.320)	-0.014* (0.006)	-0.319 (0.331)	-0.012* (0.006)	
	ln Estab urgência SUS	-0.397 (0.315)	-0.006 (0.006)	-0.326 (0.332)	-0.005 (0.006)	
	ln Estab SADT SUS	0.170 (0.296)	0.009+ (0.005)	0.266 (0.302)	0.009 (0.005)	
	Vigilância de endemias	-0.786 (1.434)	-0.010 (0.025)	-0.324 (1.492)	-0.005 (0.026)	
	Norte e Nordeste	0.630 (0.542)	0.002 (0.009)	0.460 (0.547)	0.001 (0.009)	
	Sul e Sudeste	0.317 (0.469)	0.010 (0.008)	0.190 (0.499)	0.006 (0.009)	
	Constant	-6.244* (3.178)	0.055 (0.060)	-5.000 (3.158)	0.052 (0.054)	
	Pseudo - R ²		0.409	0.405	0.369	0.386
	Número de Observações		557	557	557	557

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

\emptyset Usa o estimador Maximum Likelihood (ML), e matriz Queen de primeira ordem

Ω Usa o estimador Generalized Spatial Two-Stage Least-Squares (gs2sls), e matriz Rook de segunda ordem.

Fonte: Twitter, Ministério da Saúde, MapBiomas, FIRJAN, MUNIC, Data SUS e IBGE.