

Dinâmicas espaciais de espalhamento do SARS-COV-2 no Brasil*

Resumo

O presente artigo expõe brevemente a problemática da atual pandemia de SAR-CoV-2 que assola a humanidade. Relata que autores afirmam que as pandemias podem se tornar rotineiras, considerando que a troca virais entre espécies tem ocorrido com mais frequência devido à presença regular do homem em vários ecossistemas. Apresenta uma breve caracterização do transporte interestadual de passageiros no Brasil. Descreve o conceito de PageRank, métrica que classifica páginas web. Rememora os conceitos de escalonamento e correlação. Obtém, via dados abertos, informações sobre o transporte interestadual e internacional no Brasil. Cria, na forma de um grafo, uma rede do fluxo de passageiros e nela realiza análise exploratória preliminar. Paralelizando semanticamente a rede criada com a web, calcula sua métrica PageRank e estabelece sua correlação com a data da primeira infecção em cada Unidade Federativa do país. Por fim, esboça uma previsibilidade dos dias ordinários em que cada Unidade Federativa seria infectada por meio de regressão linear, destacando que, para além dos cálculos, a metodologia possui mais relevância como resultado apresentado.

Palavras-chave: redes, grafos, SARS-CoV-2, COVID-19, recuperação da informação.

Manoel Camilo Souza Netto

Doutorando em Ciência da Informação, PGCIN/UFSC. Mestre em Ciência da Informação, PGCIN/UFSC. Especialização em Ciências Policiais, Escola Superior de Polícia Federal. Graduação em Ciência da Computação, UFC. Agente da Polícia Federal do Brasil.
manoelcamilo@gmail.com
<https://orcid.org/0000-0002-7762-7958>

Angel Freddy Godoy Viera

Doutor em Engenharia de Produção, UFSC. Mestre em Ciência da Computação, UFSC. Professor do Departamento de Ciência da Informação, PGCIN/UFSC.
recinfor@gmail.com
<https://orcid.org/0000-0001-6657-4734>

Adilson Pinto

Doutor em Documentação, Carlos III de Madrid. Mestre em Ciência da Informação, PUC-Campinas. Graduado em Biblioteconomia, PUC-Campinas. Professor do Departamento de Ciência da Informação, PGCIN/UFSC.
adilson.pinto@ufsc.br
<https://orcid.org/0000-0002-4142-2061>

Cómo citar este artículo: Sousa-Netto, Manoel Camilo; Godoy-Viera, Angel Freddy; Pinto, Adilson Luiz (2021). Dinâmicas espaciais de espalhamento do SARS-COV-2 no brasil. *Revista Interamericana de Bibliotecología*, 44(3), e345708. <https://doi.org/10.17533/udea.rib.v44n3e345708>

Recibido: 2021-03-24 / **Aceptado:** 2021-08-09

* O artigo Dinâmicas espaciais de espalhamento do SARS-COV-2 no Brasil surgiu como proposta em um contexto do projeto de doutorado em Ciência da Informação da UFSC do orientando Manoel Camilo de Sousa Netto, Msc. sob a orientação do Prof Dr. Adilson Luiz Pinto. Considerando que um dos requisitos da linha de pesquisa exige competências de análise e visualização de dados, o projeto utilizou a construção do artigo como piloto na obtenção dessas habilidades, tais como obtenção de dados em fontes abertas e seu pré-processamento, visualização da informação, data cleaning, recuperação da informação e utilização de modelos estatísticos para análise de dados. A temática possibilitou a aquisição dessas competências utilizando um tema de relevância não apenas para o Brasil, mas para toda humanidade no atual cenário da pandemia de SARS-CoV-2, cujas consequências assolam vários países.

Spatial Dynamics for Spreading Sars-COV-2 in Brazil

Abstract

This paper exposes the problem of the current SARS-CoV-2 pandemic that is plaguing humanity. It reports that authors claim that pandemics can become routine, considering that the viral exchange between species has occurred more frequently due to the regular presence of man in various ecosystems. It presents a brief characterization of interstate passenger transport in Brazil. Describes the concept of PageRank, a metric that classifies web pages. Reminds the concepts of scheduling and correlation. It obtains, via open data, information on interstate and international transport in Brazil. It creates, in the form of a graph, a passenger flow network and performs a preliminary exploratory analysis there. Semantically paralleling the network created with the web, it calculates its PageRank metric and establishes its correlation with the date of the first infection in each Federative Unit in the country. Finally, it outlines a predictability of the ordinal days when each Federative Unit would be infected by means of linear regression, highlighting that, in addition to the calculations, the methodology has more relevance as a result presented.

Keywords: Networks, graphs theory, SARS-CoV-2, COVID-19, information retrieval.

Dinámica espacial para la propagación del SARS-COV-2 en Brasil

Resumen

Este estudio expone brevemente el problema de la pandemia SARS-CoV-2 que recae sobre la humanidad. Informa que las pandemias pueden volverse rutinarias, considerando que el intercambio viral entre especies ha ocurrido con mayor frecuencia debido a la presencia regular del hombre en varios ecosistemas. Presenta una breve caracterización del transporte interestatal de pasajeros en Brasil. Describe el concepto de PageRank, una métrica que clasifica las páginas web. Recuerda los conceptos de *programación* y *correlación*. Obtiene, a través de datos abiertos, información sobre el transporte interestatal e internacional en Brasil. Crea, en forma de grafos, una red de flujo de pasajeros y

realiza un análisis exploratorio preliminar. En paralelo y semánticamente, la red creada con la web calcula su métrica PageRank y establece su correlación con la fecha de la primera infección en cada Unidad Federativa de Brasil. Finalmente, se esboza una predictibilidad de los días ordinales en los que se infectaría cada Unidad Federativa mediante regresión lineal; se destaca que, además de los cálculos, la metodología tiene más relevancia como resultado presentado.

Palabras clave: redes, teoría de grafos, SARS-CoV-2, Covid-19, recuperación de información.

1. Introdução

Desde o surto da síndrome respiratória aguda grave (SARS) há quase duas décadas, um grande número de coronavírus relacionados à SARS (SARS-CoVs) potencialmente infectantes aos humanos foi descoberto. O SARS-CoV-2, causador de uma epidemia de síndrome respiratória aguda, foi estudado a partir de sequências de genoma que foram obtidas de cinco pacientes em um estágio inicial da doença, as quais se caracterizaram por serem 96% idênticas no nível do genoma inteiro a um coronavírus detectado em morcegos (Peng et al., 2020).

Segundo Quammen (2020), o surto em Wuhan (China) não é um infortúnio, mas faz parte de uma sequência de contingências como parte de um padrão de escolhas humanas, dentre elas apropriar-se de recursos naturais e gerar distúrbios ecológicos que aumentam as trocas virais pelo aumento da frequência de interações entre humanos e outras espécies. Ainda, segundo Quammen (2020), destruimos os ecossistemas e liberamos os vírus de seus hospedeiros naturais e, quando isso acontece, eles precisam de um novo hospedeiro. Muitas vezes, somos nós humanos.

A obsessiva necessidade da humanidade de estar onipresente pode fazê-la conviver com esse transbordamento viral entre espécies de forma recorrente. Estando o homem contaminado, estará o vírus apto a usufruir das várias redes de transporte dos seus hospedeiros viajantes, inclusive das rotas que eficientemente interligam vários países. A criatura nanométrica poderá chegar a destinos longínquos em questão de horas. Se obtiver sucesso em sua empreitada, esse vírus pode gerar uma pandemia como a ora em curso, durante a qual terá mais chances de sofrer mutações e tornar-se até mais letal ou mais transmissível.

O peregrino SARS-CoV-2 tem executado esse script darwiniano eficientemente, visto que suas variantes têm maior avides pelo receptor celular humano e podem levar a uma menor eficácia da vacina dependendo do tipo de imunizante. Algumas linhagens também escapam ao tratamento com anticorpos monoclonais (Brandão, 2021).

O SARS-CoV-2 não foi o primeiro e não será o último parasita intracelular a expandir seus espaços territoriais de contaminação se valendo do deslocamento humano. Vírus diversos executaram – e continuarão a cumprir – o roteiro: usufruir da tecnologia de transporte que o homem construiu ao longo da sua história. É possível, inclusive, que essa realidade se torne rotineira.

Considerando a possibilidade de a problemática exposta configurar-se recorrente, o deslocamento do homem no espaço territorial é uma questão de estudo relevante, pois é o portador do vírus que eventualmente ampliará a área territorial infectada, fazendo surgir novos casos inaugurais em locais outrora livres da doença. Em particular, um objeto de estudo poderia ser a existência de uma correlação formalmente descritível entre o surgimento do primeiro caso de SARS-CoV-2 em unidades territoriais (países, estados, cidades) e as quantidades de viajantes que a esses locais se destinaram.

No trabalho ora proposto, os estados (frequentemente designados como Unidades Federativas do Brasil, ou simplesmente pelo acrônimo UFs) foram convencionados como a unidade territorial padrão de destino dos viajantes. A hipótese aventada pressupõe que quanto mais passageiros viajarem à uma UF ainda livre do vírus, maiores as chances de que um deles esteja infectado pelo SARS-CoV-2. A consequência seria o aumento da probabilidade de antecipar a ocorrência da primeira infecção no destino. Assim, se a quantidade de passageiros é previamente conhecida, então talvez seja possível a predição da ordem de contaminação das UFs antes mesmo que uma futura – e indesejável – pandemia ocorra. A pesquisa também almejou analisar se o local de onde o passageiro advém pode influenciar a data do primeiro caso de SARS-CoV-2 no destino, pois locais de origem que recebem mais passageiros também teriam maiores chances de exportarem infecções aos destinos finais de outros viajantes.

2. Fundamentação teórica

2.1 Breve caracterização do transporte interestadual de passageiros no Brasil

Paiva e Müller (2014) afirmam que, no Brasil, os principais meios de transporte públicos utilizados em viagens são o avião e o ônibus, já que os transportes ferroviário e hidroviário, residuais, pouco atendem as rotas do país.

Segundo Amora e Matais (2011), o número de passageiros de avião superou, em 2010, o número de viajantes de ônibus interestaduais pela primeira vez no Brasil. A informação da reportagem é coerente com o exposto pela página Dados e Estatísticas da Agência Nacional de Transporte Terrestres (ANAC, 2019), que divulgou o gráfico de evolução das quantidades percentuais e anuais de passageiros que utilizam os modais aéreo e rodoviário em viagens interestaduais, conforme a Figura 1.

A partir do fato de que esses são os principais meios de transporte no Brasil, é possível afirmar que o SARS-CoV-2 se serviu com avides ímpar das rotas dos viajantes infectados: em um país de dimensões continentais, trafegando em seus hospedeiros pelas principais rotas interestaduais, o SARS-CoV-2 registrou casos em todas as vinte e sete UFs em apenas vinte e seis dias¹ após sua chegada ao país, conforme demonstra o a Figura 2.

Assim, para entender a ordem de ocorrência dos primeiros casos em cada UF do Brasil, uma alternativa seria estudar os deslocamentos daqueles que carregam o vírus país a fora: os viajantes nas modalidades aéreas e rodoviárias (aviões e ônibus interestaduais, em ambos os casos em rotas comerciais regulares).

Existem outras modalidades de deslocamento de viajantes não utilizadas na pesquisa, como as dos condutores e passageiros de transportes de carga rodoviária (caminhões), os que viajam por meios próprios (aviões, carros e motos particulares), os passageiros e condutores da malha hidroviária e ferroviária, e outros meios que certamente o estudo deveria considerar.

1 Dados calculados pela pesquisa a partir das datas dos primeiros casos em cada Unidade da Federação.

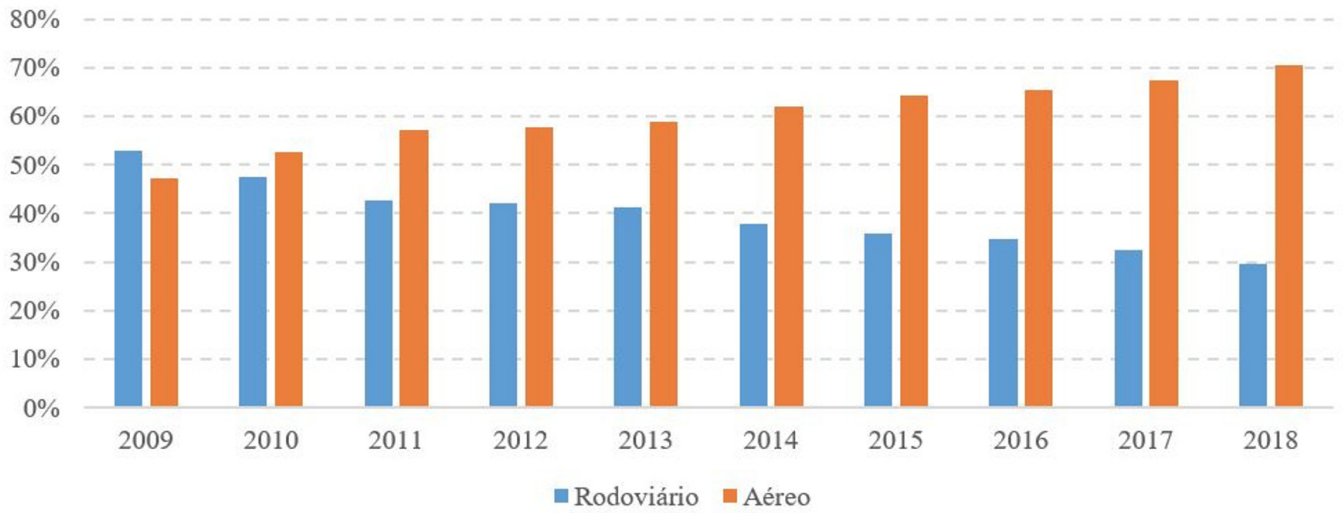


Figura 1. Participação no Transporte Interestadual Regular de Passageiros.

Fuente: extraído de ANAC (2019).

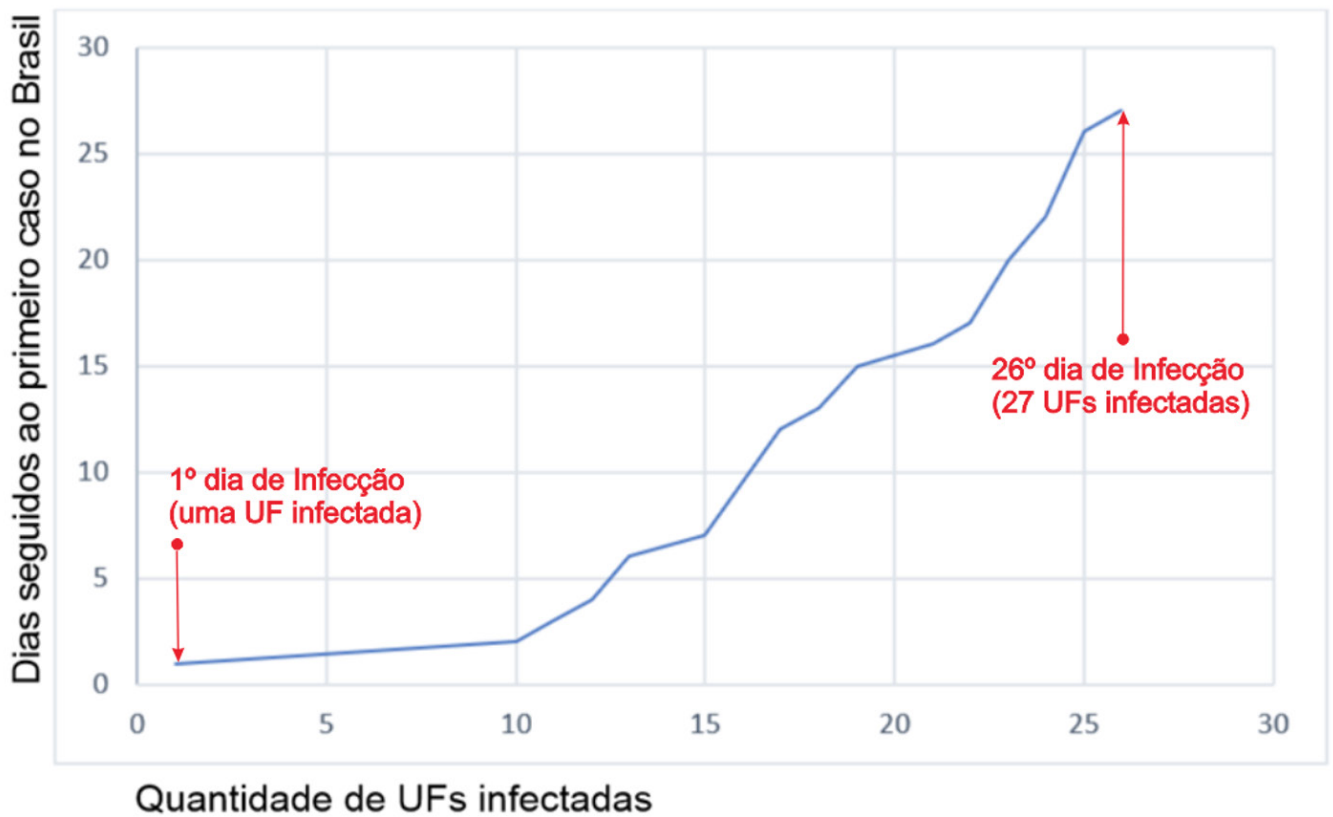


Figura 2. Dias seguidos ao primeiro caso de SARS-CoV-2 no Brasil e quantidade de Estados que registraram os primeiros casos.

Fuente: elaborado pelos autores a partir dos dados coletados.

Entretanto, diferentemente dos transportes aéreos e por ônibus interestaduais, não há dados abertos governamentais dessas outras modalidades. Esse vácuo não inviabiliza a pesquisa, visto que estão disponíveis, em sites de dados abertos do governo federal, os fluxos de passageiros dos dois principais tipos de transporte interestaduais² alhures descritos. Além disso, o trabalho ora desenvolvido não busca, como insumo de pesquisa, um retrato exato – e até mesmo impossível – de todos os fluxos de passageiros por inúmeras vias.

O que a pesquisa procurou foi uma aproximação razoável construída com dados disponíveis e que retratam os casos mais comuns dos fluxos interestaduais de pessoas. Assim, a ausência dos fluxos de viajantes excepcionais foi tratada como ruído e os dados de passageiros usados no trabalho foram aqueles permitidos pela realidade concreta.

Considerando serem estes passageiros candidatos a hospedeiros virais, as rotas por eles viajadas podem ser, em um percentual razoável, o principal meio pelo qual o vírus ocasionou os primeiros casos de SARS-CoV-2 em algumas UFs do Brasil.

O vírus acompanhou seu hospedeiro e percorreu, portanto, uma rede formada por Estados e as suas interligações que pode ser representada por um grafo de uma rede social no qual essas UFs são os vértices. Uma única aresta direcionada (ligação) representaria todas as rotas aéreas e rodoviárias entre uma UF de origem e outra de destino. Cada uma dessas ligações teriam um peso diferenciado, pois elas seriam ponderadas pelo total de passageiros que percorrem mensalmente essa rota em ambos os modais (aéreo e rodoviário). Nesse cenário, o grafo descrito pode ser analisado por suas métricas oriundas dos estudos de análise de redes sociais.

Segundo Gündüz-Ögüdücü e Etaner-Uyar (2014), a análise de redes sociais envolve a medição de métricas estruturais particulares, a fim de compreender os conceitos fundamentais dos grafos sociais. As métricas são usadas para caracterizar e analisar conexões dentro de uma determinada rede social. Algumas dessas métricas representam as características de vértices individuais,

2 O quantitativo de passageiros das rotas interestaduais regulares está disponível nos sites de dados abertos da ANAC e da ANTT (nos modais aéreo e rodoviário).

enquanto outras inferem um padrão que pertence à rede como um todo.

2.2 Métrica Pagerank³

Os motores de busca da web visam destilar os resultados por sua relevância e autoridade, e alguns deles consideram que uma boa autoridade é uma página que é apontada por muitos bons *hubs*, enquanto um bom *hub* é uma página que aponta para muitas boas autoridades (Elmasri & Navathe, 2011). Usando esse princípio, a empresa Google criou o algoritmo denominado PageRank para classificar páginas web por sua relevância e autoridade. O algoritmo dessa métrica calcula uma classificação dos vértices no grafo G com base na estrutura dos *links* de entrada. Ele foi originalmente projetado para determinar a posição relativa de uma página em relação a outra, classificando-as numa ordem de importância.

A métrica de rede Pagerank é descrita como:

A literatura de citações acadêmicas foi aplicada à web, em grande parte contando citações ou *backlinks* para uma determinada página. Isso dá uma estimativa da importância ou qualidade de uma página. O PageRank estende essa ideia por não contar os *links* de todas as páginas de forma igualitária, e por normalizar pelo número de *links* em uma página (Brin & Page, 1998, p. 4).

Segundo Elmasri e Navathe (2011), supondo $P(X)$ o PageRank de qualquer página X , $C(X)$ o número de links de saída da página X e sendo d um fator de amortecimento no intervalo $0 < d < 1$ (normalmente d é definido a 0,85); então, o PageRank para uma página A pode ser calculado como:

$$P(A) = (1 - d) + d \left(\frac{P(T_1)}{C(T_1)} + \dots + \frac{P(T_n)}{C(T_n)} \right)$$

Na expressão, T_1, T_2, \dots, T_n são as páginas que apontam para a página A . O PageRank forma uma distribuição de probabilidade nas páginas da web, de modo que a soma dos Pageranks de todas as páginas web é 1 (um) (Brin & Page, 1998). Para ilustrar o algoritmo, considere uma pequena web que consiste em três páginas A, B e C e, entre elas, algumas ligações (links), conforme ilustra a Figura 3.

3 PageRank e Google são marcas comerciais da Google Inc., Mountain View CA, EUA.

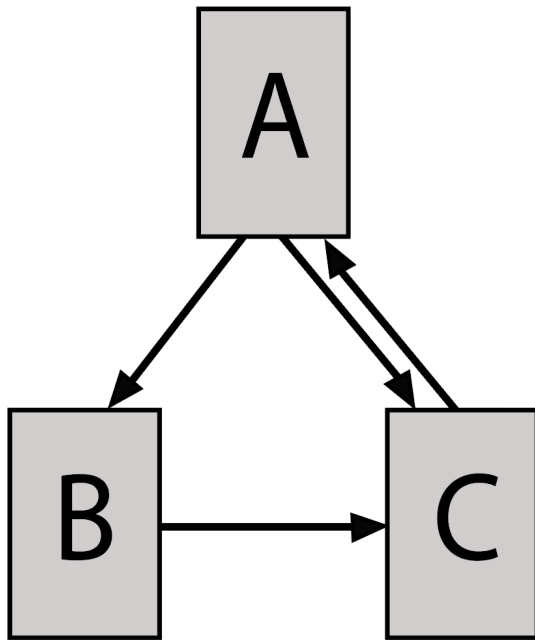


Figura 3. Pequena rede web ilustrativa do PageRank.

Fuente: extraído de Sobek (2002).

O fator de amortecimento d geralmente é definido como 0,85, mas para simplificar o cálculo, foi adotado o valor 0,5. Abaixo estão montadas as equações para o cálculo do PageRank (Sobek, 2002):

$$PR(A) = 0,5 + 0,5 * PR(C)$$

$$PR(B) = 0,5 + 0,5 * \frac{PR(A)}{2}$$

$$PR(C) = 0,5 + 0,5 * \left(\frac{PR(A)}{2} + PR(B) \right)$$

A solução das equações anterior nos dá como resultados: $PR(A) = 1,07692308$; $PR(B) = 0,76923077$ e $PR(C) = 1,15384615$. A soma dos três PageRanks obtidos corresponde a 3 (três), ou seja, 100% das páginas. Calculadas as métricas, as páginas agora podem ser classificadas.

O PageRank pode ser adaptado para redes ponderadas como uma extensão do algoritmo convencional baseado no mesmo conceito, mas realizando a multiplicação pelos pesos (Kulshreshtha, 2020).

2.3 Escalonamento de valores numéricos via normalização

A normalização é uma técnica de escalonamento de valores numéricos utilizada em uma etapa de pré-processamento dos dados. Sua ideia central consiste em realocar variáveis de ordens de grandeza diferentes em intervalos iguais.

Segundo Patro e Sahu (2015), a técnica que fornece transformação linear incidente sobre a faixa original de dados é chamada de Normalização Min-Max, a qual pode ajustar as variáveis dentro de um limite pré-definido e pode ser obtida pela aplicação da seguinte expressão:

$$A' = \left(\frac{A - \text{valor mínimo de } A}{\text{Valor máximo de } A - \text{Valor mínimo de } A} \right) * (D - C) + C$$

Onde:

- A' contém dos dados normalizados por Normalização Min-Max;
- Se há limites pré-definidos, estes são C e D.

2.4 Correlação entre variáveis

Segundo Devore (2005):

Dados os n pares de observações $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, é natural que se fale de x e y como tendo uma relação positiva, se x_s grandes estiverem pareados com y_s grandes e x_s pequenos com y_s pequenos. De modo semelhante, se x_s grandes estiverem pareados com y_s pequenos e x_s pequenos com y_s grandes, então está implícita uma relação negativa entre as variáveis. (p. 468)

As relações (negativa e positiva) acima descritas podem ser expressas pelo coeficiente de correlação, cujo cálculo é obtido pela seguinte expressão:

$$r = \frac{n \sum xy - (\sum x) (\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}$$

Devore (2005) ilustra com os seguintes gráficos (Figura 4) algumas possibilidades visuais para um gráfico gerado entre duas variáveis e a caracterização do coeficiente de correlação r entre ambas:

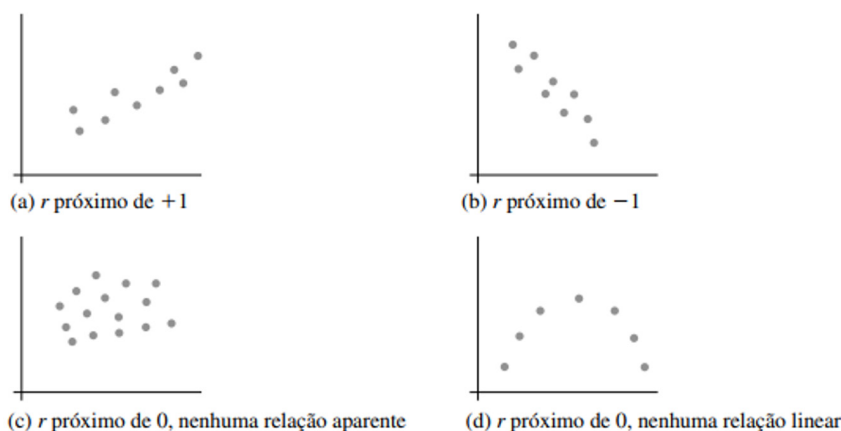


Figura 4. Exemplos de correlação.

Fuente: extraído de Devore (2005, p. 470).

2.5 Erro Médio Absoluto

Considere algumas observações estatísticas, cada uma delas nomeadas como ($O_i; i=1,2,3...n$) e considere algumas predições ($P_i; i=1,2,3,4...n$) da mesma variável. O erro e_i entre o que foi observado e o que foi predito e o erro médio absoluto (*Mean Absolute Error* ou MAE) podem ser calculados (Willmott & Matsuura, 2005), como:

$$e_i = P_i - O_i \quad MAE = [n^{-1} \sum |e_i|]$$

3. Metodologia

3.1 Obtenção dos dados

Os dados da pesquisa foram obtidos em diversas fontes de dados abertos, todas listadas na Tabela 1 abaixo:

Tabela 1. Fontes dos dados utilizados na pesquisa.

Tipo de Informação	Fonte	Local de download
Dados dos Estados (UFs)	IBGE	http://geoftp.ibge.gov.br/
Dados estatísticos do Transporte aéreo	ANAC	https://www.anac.gov.br/aceso-a-informacao/dados-abertos/areas-de-atuacao/voos-e-operacoes-aereas/dados-estatisticos-do-transporte-aereo
Dados de bilhete de passagem coletados pelo Sistema de Monitoramento do Transporte Rodoviário Interestadual e Internacional Coletivo de Passageiros (MONTRIIP).	ANTT	https://dados.antt.gov.br/dataset/montriip-bilhetes-de-passagem
Dados do Sistema de Controle dos Serviços de Transporte Rodoviário de Passageiros SISDAP	ANTT	https://portal.antt.gov.br/estatisticas-e-estudos-de-servicos-de-transporte-de-passageiros
Dados dos Casos de Covid - Brasil	Site Brasil.IO4 *	https://brasil.io/dataset/covid19/files/

Fuente: elaborados pelos autores.

* O site Brasil.IO tem facilita o acesso a dados públicos brasileiros. O projeto é colaborativo e todo o código está disponível como software livre, cujos custos são pagos através de uma campanha de financiamento coletivo.

3.2 Estruturação de uma analogia entre a web e a rede de passageiro interestaduais

É possível fazer uma analogia entre páginas web e o fluxo de passageiros em um país. Sob esse aspecto, enquanto as ligações entre páginas web são hiperlinks de entrada e de saída, nas Unidades Federativas de um país elas são os fluxos de partida e chegada de passageiros entre dois Estados distintos.

Na web, as páginas podem ser classificadas quanto ao prestígio pela métrica PageRank. Uma página com alto PageRank não é só aquela com muito links de entrada, mas aquela que é bem apontada por páginas também prestigiadas. Por isso, o PageRank de um vértice não depende apenas dos *links* nele incidentes, mas também da qualidade dos vértices donde as ligações se originam.

No contexto semanticamente adaptado para a pesquisa, o Pagerank de uma rede de passageiros não seria apenas a quantidade de pessoas que desembarcam em um local de destino da viagem; para além disso, a métrica também depende do prestígio dos locais de origem donde esses passageiros partiram.

3.3 Cálculo do PageRank das UFs

De posse da data do primeiro caso de SARS-CoV-2 em cada UF, mas considerando que datas são menos manipuláveis em termos quantitativos, foi realizada uma conversão dessa variável para números ordinais, da seguinte forma: a

data do primeiro caso ocorrido no país, no Estado de São Paulo, em 25/02/2020, foi convencionada como $d_{SP} = 1$. O Rio de Janeiro, por sua vez, teve seu primeiro registro de infecção em dia 05/03/2020, logo $d_{RJ} = 10$ (décimo dia desde que o primeiro caso foi registrado).

Para cada Unidade Federativa do Brasil, a pesquisa também realizou o cálculo do respectivo Pagerank PR.

A Tabela 2 relaciona as UFs brasileiras, seus PageRanks ponderados e a data ordinal da ocorrência do 1º caso registrado. Apresenta também essas variáveis normalizadas pelo método Min-Max para um intervalo entre [0,1].

Tabela 2. 1º registro de SARS-CoV-2 nas UFs e cálculo da métrica de rede PageRank

UF	Nome da UF	Data do	Variáveis		Variáveis normalizadas pelo método Min-Max para o intervalo [0,1]	
		1º caso	Dia ordinal	Pagerank	Dia Normalizado	Pagerank Normalizado
SP	São Paulo	25/02/2020	Normalizado	Pagerank	intervalo [0,1]	1
RJ	Rio de Janeiro	05/03/2020	Normalizado	0,068760292	0,36	0,250921804
BA	Bahia	06/03/2020	11	0,05829879	0,4	0,2106896
DF	Distrito Federal	07/03/2020	12	0,140583265	0,44	0,527134201
AL	Alagoas	08/03/2020	13	0,008433079	0,48	0,018919096
MG	Minas Gerais	08/03/2020	13	0,075893451	0,48	0,278354068
RS	Rio Grande do Sul	10/03/2020	15	0,026717254	0,56	0,089235259
GO	Goiás	12/03/2020	17	0,110807948	0,64	0,412626107
PE	Pernambuco	12/03/2020	17	0,02799265	0,64	0,0941401
PR	Paraná	12/03/2020	17	0,036400133	0,64	0,126473084
RN	Rio Grande do Norte	12/03/2020	17	0,008456317	0,64	0,019008463
SC	Santa Catarina	12/03/2020	17	0,032421007	0,64	0,111170401
AM	Amazonas	13/03/2020	18	0,011288054	0,68	0,029898585
MS	Mato Grosso do Sul	14/03/2020	19	0,009948375	0,72	0,024746531
SE	Sergipe	14/03/2020	19	0,006806576	0,72	0,012663989
CE	Ceará	16/03/2020	21	0,022496306	0,8	0,073002596
AC	Acre	17/03/2020	22	0,003630093	0,84	0,000448067
TO	Tocantins	18/03/2020	23	0,005511989	0,88	0,007685349
PA	Pará	18/03/2020	23	0,014256179	0,88	0,041313217
PB	Paraíba	18/03/2020	23	0,008066788	0,88	0,017510437
ES	Espírito Santo	19/03/2020	24	0,015829626	0,92	0,047364286
PI	Piauí	19/03/2020	24	0,008295436	0,92	0,018389758
AP	Amapá	20/03/2020	25	0,003513583	0,96	0
MA	Maranhão	20/03/2020	25	0,008768114	0,96	0,020207555
MT	Mato Grosso	20/03/2020	25	0,012341023	0,96	0,03394803
RO	Rondônia	20/03/2020	25	0,005512056	0,96	0,007685604
RR	Roraima	21/03/2020	26	0,003644261	1	0,000502553

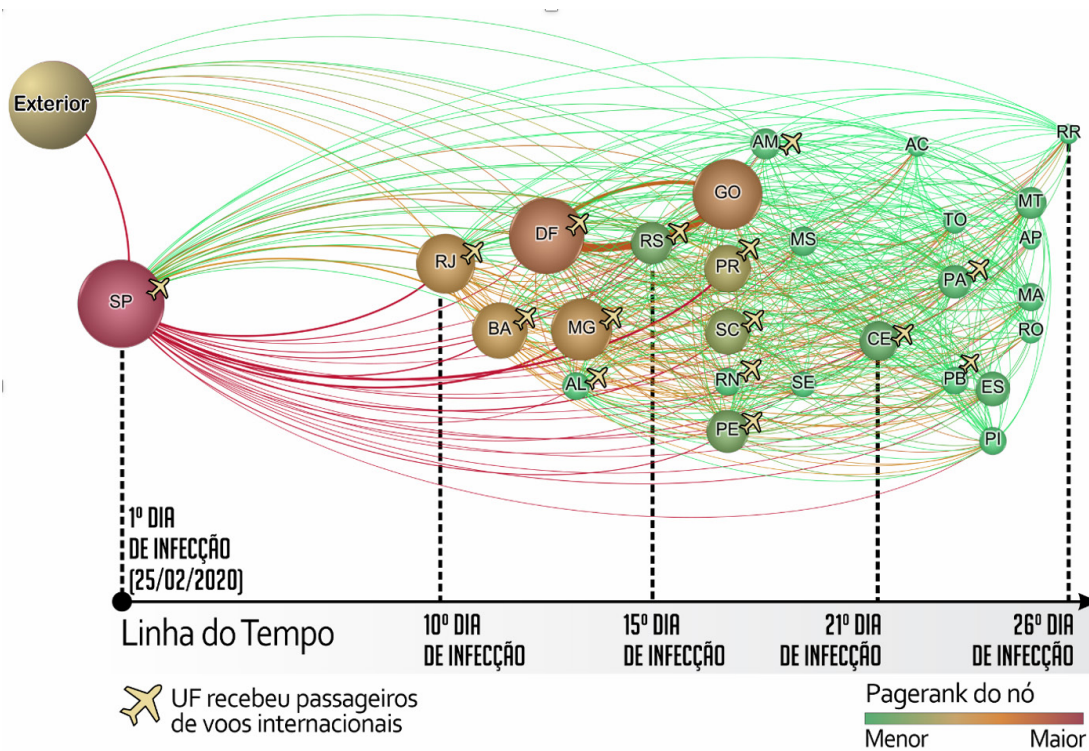
Fuente: elaborados pelos autores.

3.4 Análise exploratória da rede de fluxos de viajantes interestaduais

A partir do cálculo do PageRank e de posse do dia ordinal do primeiro caso registrado de SARS-CoV-2 em cada UF do Brasil, foi possível realizar uma análise exploratória preliminar e visual da respectiva rede. Para tanto, um grafo representativo foi gerado para verificar alguma característica que gere intuições acerca de uma eventual relação entre a métrica PageRank e o dia de primeiro caso em cada UF, conforme ilustrado na Figura 5.

Na Figura 5, cada vértice representa uma UF. Presente em algumas delas, o ícone de um avião presente indica as UFs que receberam voos internacionais durante o período de vinte seis dias transcorridos até que todas registraram seus primeiros casos e, a ausência desse ícone representa que apenas voos domésticos (nacionais) incidiram naquele período. Cada ligação (ou aresta) entre um par de UFs representa a totalidade de passageiros que trafegaram tanto no modal rodoviário quanto no modal aéreo. Por sua vez, uma ligação entre

o vértice nominado como “exterior” e uma UF representa o fluxo aéreo entre qualquer país estrangeiro e uma UF brasileira apenas na modalidade aéreo. O eixo temporal da abscissa se inicia na ocorrência do caso inaugural, em 25/02/2020, em São Paulo. A análise exploratória permite perceber que os vértices das UFs com maiores PageRanks se situaram, em sua maioria, no início da linha do tempo, ou seja, nos primeiros dias da pandemia. Esse posicionamento de vértices é evidenciado pela presença predominante daqueles investidos em tons em vermelho e laranja situados - ao que parece, não por acaso - à esquerda do grafo, enquanto os tons mais verdes se acomodaram à direita, espaço reservado aos dias posteriores. Há, também, uma tendência da acomodação das UFs que receberam passageiros de voos internacionais nos primeiros dias da linha do tempo, o que é coerente, visto que a pandemia não teve origem em território nacional, mas sim em países estrangeiros. Adiante, na discussão dos resultados, a pesquisa relatará ser esse um aspecto negativo aos números obtidos no estudo de caso, mas não ao método.



[Dinâmicas espaciais de espalhamento do SARS-CoV-2 no Brasil]

Figura 5. Grafo da rede de fluxos de passageiros interestaduais.

Fuente: elaborado pelos autores.

Em apenas 26 (vinte e seis) dias todos os Estados do Brasil já haviam registrado casos de contaminação por SARS-CoV-2. As percepções do grafo apresentado indicam que pode haver uma correlação entre a métrica PageRank e o registro do 1º caso de SARS-CoV-2 em boa parte das UFs do Brasil.

3.5 Correlação entre variáveis da pesquisa

Para confirmar a existência de uma relação entre o dia do primeiro caso de SARS-CoV-2 e a métrica PageRank, é necessário calcular o coeficiente de correlação. O valor obtido do coeficiente de correlação foi -0,79. Há, portanto, uma tendência de que as UFs com maiores PageRanks sejam infectadas antes daquelas que apresentam valores menores para essa métrica. Segundo Mukaka (2012), uma correlação no intervalo entre -70 a -90 é considerada negativa forte. O gráfico da Figura 6 permite mostrar a tendência por meio de uma reta de regressão linear.

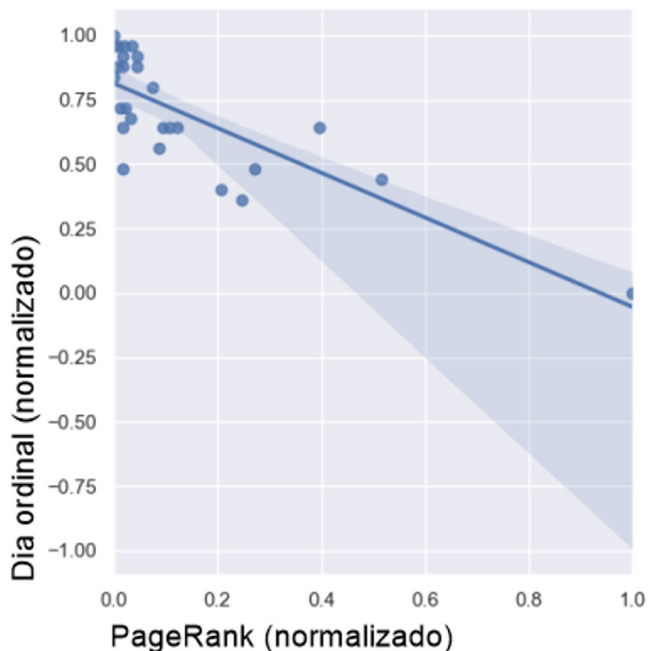


Figura 6. Gráfico de Correlação entre o dia do 1º caso e o PageRank da UF.

Fuente: elaborado pelos autores.

4. Resultados e discussão

4.1 Previsibilidade da ordem de UFs a serem infectadas

Visto que a regressão linear foi determinada, podemos obter a reta que a descreve. A reta da equação é: $y = -82,77 * x + 21,65$.

Assim, de posse da equação, podemos calcular qual seria o y previsto para cada uma das UFs (Tabela 3).

Como resultado, o erro médio absoluto (MAE) calculado com base na tabela acima foi de 3,19 dias.

Tabela 3. Data ocorrida e data prevista de primeira Infecção em cada UF

UF	PageRank	Dia ocorrido	Dia previsto
SP	0,26354164	1	1
RJ	0,06876029	10	16
BA	0,05829879	11	17
DF	0,14058326	12	10
AL	0,00843308	13	22
MG	0,07589345	13	16
RS	0,02671725	15	20
Go	0,11080795	17	13
PE	0,02799265	17	20
PR	0,03640013	17	19
RN	0,00845632	17	22
SC	0,03242101	17	19
AM	0,01128805	18	21
MS	0,00994838	19	21
SE	0,00680658	19	22
CE	0,02249631	21	20
AC	0,00363009	22	22
TO	0,00551199	23	22
PA	0,01425618	23	21
PB	0,00806679	23	22
ES	0,01582963	24	21
PI	0,00829544	24	22
AP	0,00351358	25	22
MA	0,00876811	25	22
MT	0,01234102	25	21
RO	0,00551206	25	22
RR	0,00364426	26	22

Fuente: elaborados pelos autores.

4.2 Discussão dos resultados

Seria desejável que o erro médio absoluto (3,19 dias) estivesse em uma faixa menor, considerando que o vírus infectou todas as UFs dentro de apenas 26 dias, o que faz com que 3,19 dias sejam um erro de magnitude razoável nesse intervalo de tempo. Os valores previstos, entretanto, perdem importância diante do método apresentado o qual, essencialmente, propõe a utilização do PageRank por paralelismo semântico com os deslocamentos humanos geradores do aumento da área infectada.

A priori, o estudo seria mais útil em unidades administrativas menores, como as cidades. No Brasil, 15 dos 27 Estados (ou 55,56% deles) receberam passageiros internacionais durante os 26 dias em que o vírus levou para marcar presença em todas as unidades federativas. Esse fato pode alterar a ordem dos primeiros casos da doença, tendo em vista que os primeiros contaminados advêm do estrangeiro para o interior do país, fato que independe dos fluxos internos de passageiros. Caso a pesquisa tivesse as cidades como unidade territorial, cada resultado seria menos influenciado pelos voos internacionais, a quantidade de previsões seria maior e o intervalo de tempo muito além dos 26 dias que o vírus levou para atingir todas as UFs. Isso possibilitaria uma melhor avaliação dos valores preditos, desta feita menos influenciados pelos voos internacionais e mais dependente do fluxo interno, pois apenas 27 das 5570 cidades receberam passageiros internacionais, ou 0,48% delas, o que revela uma característica excepcional. Isso certamente implica que os voos internacionais, como é de se pressupor, são responsáveis pela introdução dos casos no país e, os fluxos entre cidades pela interiorização no cenário de transmissão comunitária,⁴ que ocorre quando as transmissões virais se tornam irrastráveis entre os habitantes. Esse estágio foi oficialmente decretado no Ministério da Saúde do Brasil (Brasil, 2020)

4 Na prática, o conceito de transmissão comunitária diz respeito à incapacidade de rastrear de onde veio o vírus que contaminou alguém, o que indica que ele já está espalhado em um determinado lugar e pode se disseminar com velocidade. Quando um indivíduo chega ao sistema de saúde e testa positivo para o Coronavírus, as autoridades de saúde o questionam com relação ao contato com outros doentes, para rastrear de onde surgiu o vírus. Quando ocorre transmissão comunitária, esse rastreamento já não é mais possível, o que indica que se espalhou.

em 20/03/2020 em todo o território nacional, menos de um mês após o registro do primeiro caso de Covid-19 no Brasil.

Assim, seria mais adequado usar as cidades como a unidade administrativa padrão. Entretanto, isso apenas seria possível se houvesse disponibilidade de dados abertos dos passageiros intermunicipais. Entretanto, a gestão dos dados dos transportes intermunicipais não cabe ao Governo Federal,⁵ mas sim aos estados, cujas disponibilidades de dados abertos ainda estão muito aquém do mínimo necessário ao estudo.

Nesse caso a metodologia seria a mesma, o que permite concluir não serem os valores preditos na pesquisa, mas a metodologia, o principal resultado apresentado.

5. Conclusão

A pesquisa sabe da provável existência de diversas variáveis ocultas e indeterminadas que concorrem para o registro do caso inaugural do vírus em cada Unidade da Federação. É sabido também que seria impossível medir todas as causas desconhecidas e não quantificáveis que concorrem para antecipar a data do primeiro caso registrado.

Em contrapartida, para atingir o objetivo pretendido, o modelo da pesquisa incluiu as principais e mais comuns causas do efeito, não todas elas: as duas maiores formas de deslocamento dos viajantes interestaduais, inclusive dos porventura infectados, bem como os viajantes que do exterior chegam por via aérea.

O aumento dos fluxos de passageiros cujas quantidades são conhecidas não é um fator determinante, mas sim um amplificador das chances de antecipação da data do primeiro caso de SARS-CoV-2 nos Estados. Por outra forma, quanto maior a quantidade de passageiros que numa UF chegam, é razoável considerar maior ser a possibilidade que o registro de infecção ocorra mais cedo naquele Estado, pois os viajantes são os hospedeiros do vírus que irá gerar o primeiro caso da doença no lugar de destino.

5 No Brasil, as agências federais só publicam dados do transporte de passageiros interestaduais.

A proposta adotou o PageRank porque se trata de uma métrica que classifica páginas da internet considerando as ligações incidentes, bem como as origens e os pesos desses links, as quais são características que, respectivamente, são símeis às rotas, às quantidades de pessoas e às origens. Esse fato paraleliza semanticamente a web aos fluxos de passageiros entre unidades territoriais. A relevância da pesquisa se deve ao fato de que, ao saber da ordem de espalhamento da pandemia pelas unidades territoriais pode oportunizar ações preliminares e profiláticas, tais como o aumento de leitos, compra de equipamentos, criação de hospitais de campanha e a alocação de equipes médicos.

A pesquisa também alerta sobre a carência de dados abertos, visto que o nível mínimo de granularidade obtido foi aquele correspondente às UFs por indisponibilidade de microdados em uma temática tão relevante em tempos de pandemia.

6. Referências

- Amora, Dimmi; Matais, Andreza (2011). *Viagens de avião dobram em oito anos*. <https://www1.folha.uol.com.br/fsp/mercado/me2203201103.htm>
- Agência Nacional de Aviação Civil (2019). *Transporte Interestadual Regular de Passageiros: aéreo e rodoviário. Aéreo e Rodoviário*. ANAC. <https://www.anac.gov.br/assuntos/dados-e-estatisticas/mercado-de-transporte-aereo/transporte-interestadual-de-passageiros-aereo-e-rodoviario>
- Brandão, Paulo Eduardo (2021). *As variantes do coronavírus: qual é o tamanho do problema?* <https://saude.abril.com.br/blog/virosfera/as-variantes-do-coronavirus-qual-e-o-tamanho-do-problema/>
- Brasil (2020). Portaria nº 454, de 20 de março de 2020. Declara, em todo o território nacional, o estado de transmissão comunitária do coronavírus (covid-19). Portaria Nº 454. Brasília, DF, 20 mar. 2020. Seção I, 1-1. <https://www.in.gov.br/en/web/dou/-/portaria-n-454-de-20-de-marco-de-2020-249091587>
- Brin, Sergey; Page, Lawrence (1998). The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the Seventh International World Wide Web Conference*, 30, 107-117.
- Devore, Jay (2005). *Probabilidade e Estatística: para engenharia e ciências*. São Paulo: Cengage Learning.
- Elmasri, Ramez; Navathe, Shamkant B. (2011). *Fundamentals of Database Systems*. Boston: Addison-Wesley.
- Gündüz-Ögüdücü, Sule; Etaner-Uyar, Sima (2014). *Social Networks: analysis and case studies*. Nova Torque: Springer.
- Kulshreshtha, Sakshi (2020). Weighted PageRank Algorithm. <https://www.geeksforgeeks.org/weighted-pagerank-algorithm/>
- Mukaka, Mm (2012). A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal*. 24(3), 69-71. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3576830/>
- Paiva, Iara Cristina Pereira Lima; Müller, Carlos (2014). Competição entre o ônibus e o avião no transporte interestadual de passageiros na Região Metropolitana de Belo Horizonte. *Journal Of Transport Literature* 8(1), 109-124. <https://doi.org/10.1590/S2238-10312014000100006>
- Patro, S. Gopal Krishna; Sahu, Kishore Kumar (2015). Normalization: a preprocessing stage. *International Advanced Research Journal in Science, Engineering and Technology*, 2(3), 20-22. <https://doi.org/10.17148/IARJSET.2015.2305>
- Peng, Zhou; Yang, Xing-Lou; Wang, Xian-Guang; Lei Zhang, Ben Hu,; Hao-Rui Si, Wei Zhang; ... Shi, Zheng-Li (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270-273. <http://dx.doi.org/10.1038/s41586-020-2012-7>
- Quammen, David (2020). *Contágio: infecções de origem animal e a evolução das pandemias*. São Paulo: Companhia das Letras.
- Sobek, Markus (2002). O Algoritmo PageRank. <http://pr.efactory.de/e-pagerank-algorithm.shtml>
- Willmott, Cort; Matsuura, Kenji (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1): 79-82. <https://www.jstor.org/stable/24869236>