The correspondence between the structure of the terrestrial mobility network and the emergence of COVID-19 in Brazil

Vander L. S. Freitas[1], Jeferson Feitosa[2], Catia S. N. Sepetauskas[3], Leonardo B. L. Santos[3,4]

[1]Department of Computing, Federal University of Ouro Preto (UFOP), Ouro Preto, Brazil
[2]São Paulo State University (UNESP), São José dos Campos, Brazil
[3]National Center for Monitoring and Early Warning of Natural Disasters (Cemaden), São José dos Campos, Brazil
[4]Institute of Physics, Humboldt University of Berlin, Germany

Corresponding author: vandercomp@gmail.com

Abstract: The inter-cities mobility network is of great importance in understanding outbreaks, especially in Brazil, a continental-dimension country. Grounded on the complex networks approach, cities are here represented as nodes and the flows as weighted edges - these geographical graphs, (geo)graphs, are handled in a Geographical Information System. We adopt the IBGE database from 2016, which contains the weekly flow of people between cities in terrestrial vehicles. The present work aims to investigate the correspondences of the networks' measures, like strength, degree, and betweenness with the emergence of cities with confirmed cases of COVID-19 in Brazil, and special attention is given to the state of São Paulo. We show that the results are better when certain thresholds are applied to the networks' flows to neglect the lowest-frequency travels. The correspondences presented statistical significance for most measures up to a certain period. Until the end of April, the best matchings are with the strength measure (total flow related to a node/city) under a high flow threshold in the São Paulo state, when the most connected cities are reached. After this stage, the lower thresholds become more suitable, indicating a possible signature of the outbreak interiorization process. Surprisingly, some countryside cities such as Campina Grande (state of Paraíba), Feira de Santana (state of Bahia), and Caruaru (state of Pernambuco) have higher strengths than some states' capitals. Furthermore, some cities from the São Paulo state such as Presidente Prudente and Ribeirão Preto are captured in the top-rank positions of all the analyzed network measures under different flow thresholds. Their importance in mobility is crucial and they are potential super spreaders like the states' capitals. Our analysis offers additional tools for understanding and decision support to inter-cities mobility interventions regarding the SARS-CoV-2 and other epidemics.

Keywords: Complex Networks, Mobility Networks, Geographical Networks, COVID-19, SARS-CoV-2.

Financial support: São Paulo Research Foundation (FAPESP) Grant Number 2015/50122-0 and DFG-IRTG Grant Number 1740/2; FAPESP Grant Number 2018/06205-7; CNPq Grant Numbers 420338/2018-7 and 101720/2020-3.

INTRODUCTION

The complex network approach[1] emerges as a natural mechanism to handle mobility data computationally, taking areas as nodes (fixed) and movements between origins and destinations as connections (flows)[2,3,4]. The inter-cities mobility network is vital for understanding outbreaks, especially in Brazil, a continental-dimension country[5,6,7].

As of May 1st, 2020, the pandemic of COVID-19, caused by the SARS-CoV-2, has globally spread, with about 2,066,023 confirmed cases and 239,447 deaths. In Brazil, there are more than 92,665 confirmed cases and more than 6,439 deaths[8,9,10], with the first documented case located in the city of São Paulo on February 25th, 2020.

This paper presents an investigation on how topological properties of terrestrial mobility networks relate to the emergence of COVID-19 cases in Brazil, considering cities as nodes and flows as weighted edges. We compute three pointwise measures for each node, namely the strength, degree, and betweenness centrality to find the structurally more important cities and contrast them with the documented cases of COVID-19 until May 1st, 2020.

The most common mobility data used in studies of this nature in Brazil are the pendular travels, from the 2010 national census (IBGE)[11]. In this paper, we use the roads' IBGE data from 2016[12], which contains the flows between cities considering terrestrial vehicles in which it is possible to buy a ticket (mainly buses and vans). The information collected by that research seeks to quantify the interconnection between cities, the movement of attraction that urban centers carry out for the consumption of goods and services, and the long-distance connectivity of Brazilian cities. The North region is not included in this paper, because neither the fluvial nor the air modals are covered and their roles are key to understanding the spreading process there, especially in the Amazon region.

Our contributions are the analysis of i) the Brazilian inter-cities mobility networks under different flow thresholds to neglect the lowest-frequency travels, especially in the beginning of the outbreak, when the interiorization of the disease is still not in progress; ii) the correspondence between the networks' statistics and the emergence of COVID-19 in Brazil. The present investigation offers additional tools for understanding and decision support in the containment of the ongoing epidemiological spreading[13,14] and others in the future. From the mobility data, the authorities have a preliminary list of cities with a high likelihood of having patients to further employ preventive actions like social distancing.

This paper is organized as follows: the Method section presents the data and the techniques we employ, such as the complex networks' measures, and the geographical visualization tools. Following, the analysis results are exhibited with the discussion and final remarks.

METHOD

The above-cited IBGE data[12] contains the weekly travel frequency (flow) between pairs of Brazilian cities/districts. The frequencies are aggregated within the round trip, which means that the number of travels from city A to city B is the same as from B to A. We produce two types of undirected networks with a different number N of nodes to capture actions in two scales (country and state):

1. $N = 4987$ - Brazil without the North region (BRWN): nodes are cities and edges are the flow of direct travels between them.

2. $N = 620$ - São Paulo state (SP): a subset of the previous network, containing only cities within the São Paulo state.

We focus on two versions of each network for certain flow thresholds η, the η_0 ($\eta = 0$) that is the original network from the IBGE data and η_d ($\eta = d$), to neglect travels with lower-level frequencies. The d corresponds to the higher flow threshold that produces the network with the largest diameter. The motivation behind η_d is to get a threshold high enough to not consider the least frequent connections and to not disregard the most frequent ones[4].

Complex network measures

The topological degree k of a node is the number of links it has to other nodes. As here the networks are undirected, there is no distinction between incoming and outgoing edges.

In a connected graph, there is at least one shortest path σ_{vw} between any pair of nodes v and w. The betweenness[2] centrality b of a node i is the rate of those shortest paths that pass through i:

$$b_i = \sum_{v \neq w \neq i} \frac{\sigma_{vw}(i)}{\sigma_{vw}} \quad . \tag{1}$$

Although it is a pointwise measure, it takes into account non-local information related to all shortest paths on the network. It is worth highlighting that in the present context this centrality index is not a transportation (physical) measure but a mobility (process) one. Besides, both degree and betweenness do not account for the network flows here, but the binary (weightless) networks. The diameter of a network is the distance between the farthest nodes, given by the maximum shortest path.

The strength of a node on the other hand is the accumulated flow from incident edges:

$$s_i = \sum_{j=1}^{N} F_{ij} \tag{2}$$

in which F_{ij} is the flow between nodes i and j.

In our context, the degree gives the number of cities that a city is connected to, showing the number of possible destinations for the SARS-CoV-2. The strength captures the total number of people that travel to (or come from) such places in a week. From a probability perspective, the cities that receive more people are more vulnerable to SARS-CoV-2. The betweenness centrality, on the other hand, considers the entire network to depict the topological importance of a city in the routes that are more likely to be used.

Geographical visualization

A geographical approach for complex systems analysis is especially important for mobility phenomena[14]. Santos *et al.* (2017)[15] proposed a graph where the nodes have a known geographical location, and the edges have spatial dependence, the (geo)graph. It provides a simple tool to manage, represent, and analyze geographical complex networks in different domains[4,16] and it is used in the present work. The geographical manipulation is performed with the PostgreSQL Database Management System and its spatial extension PostGIS. Lastly, the maps are produced using the Geographical Information System ArcGIS.

RESULTS AND DISCUSSION

This section presents the results of the topological analysis for the previously mentioned networks. Table 1 showcases the size N of each network, number of edges $|E|$, average strength $\langle s \rangle$, average degree $\langle k \rangle$, and average betweenness $\langle b \rangle$.

TABLE 1

Statistics for the Brazilian (BRWN) and São Paulo state (SP) networks, with two flow thresholds: η_0 (original flows) and η_d (the higher threshold with maximum diameter). For each network and threshold, the number of edges $|E|$, number of nodes N, average degree $\langle k \rangle$, average strength $\langle s \rangle$, and average betweenness $\langle b \rangle$ are displayed.

		η_0				η_d			
Network	N	$\|E\|$	$\langle s \rangle$	$\langle k \rangle$	$\langle b \rangle$	$\|E\|$	$\langle s \rangle$	$\langle k \rangle$	$\langle b \rangle$
BRWN	*4987*	*59453*	*1169.4*	*23.8*	*5219.4*	*2482*	*414.0*	*1.0*	*1385.6*
SP	*620*	*4796*	*1132.4*	*15.5*	*504.2*	*486*	*535.0*	*1.6*	*169.9*

The |E| decreases for increasing η, due to the removal of edges with lower flows. The resulting networks are undirected. Throughout the paper, both the degree and the betweenness measures do not account for the flows, but weightless edges instead. Two nodes are connected when between them there is a nonzero flow, which means that the number of connections decreases for increasing threshold (η). We compute the diameter of the networks for varying η.

Figure 1 shows the exact point (dashed line) where the higher threshold with maximum diameter is found for both networks: $\eta_d = 207.55$ for BRWN and $\eta_d = 161.01$ for SP.

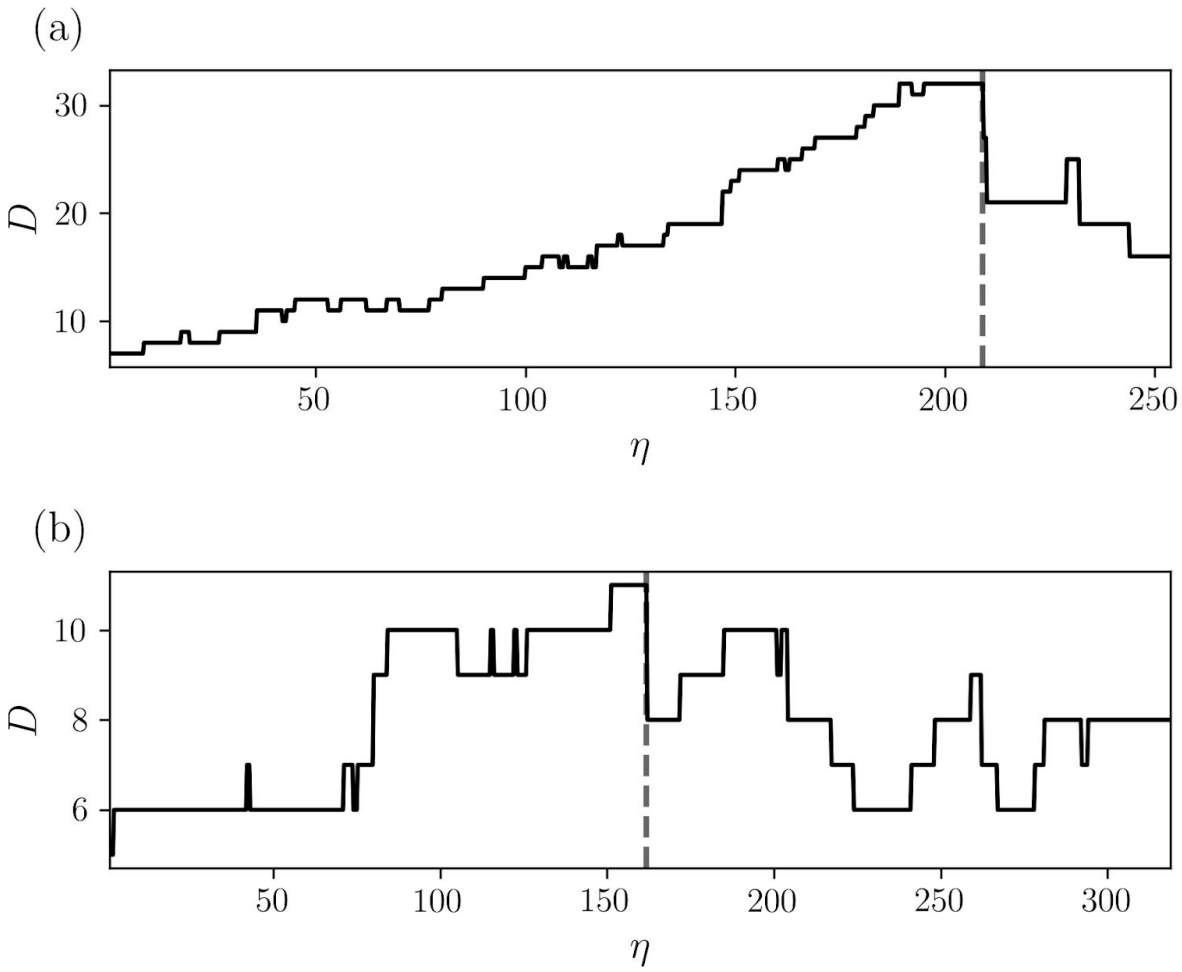

Figure 1: Diameter of the networks as a function of flow threshold η . The dashed line is the maximum diameter with higher η , with $\eta_d = 207.55$ for BRWN (a) and $\eta_d = 161.01$ for SP (b).

Following the (geo)graphs approach, it is possible to visualize nodes and edges of the Brazilian mobility network in the geographical space for η_d in Figure 2. The edges for η_0 are not plotted, because there are more than 59000 and the visualization was not clear. It is important to highlight some key cities like Belo Horizonte, Rio de Janeiro, São Paulo and Salvador, and the high number of connections between them. Figure 3 depicts the geographical graph regarding the state of São Paulo.

Figure 4 shows the map of the topological degree related to each node/city, considering all original flows (η_0), and in Figure 5 there is the equivalent for $\eta_d = 207.55$. Key cities are labeled in the maps.

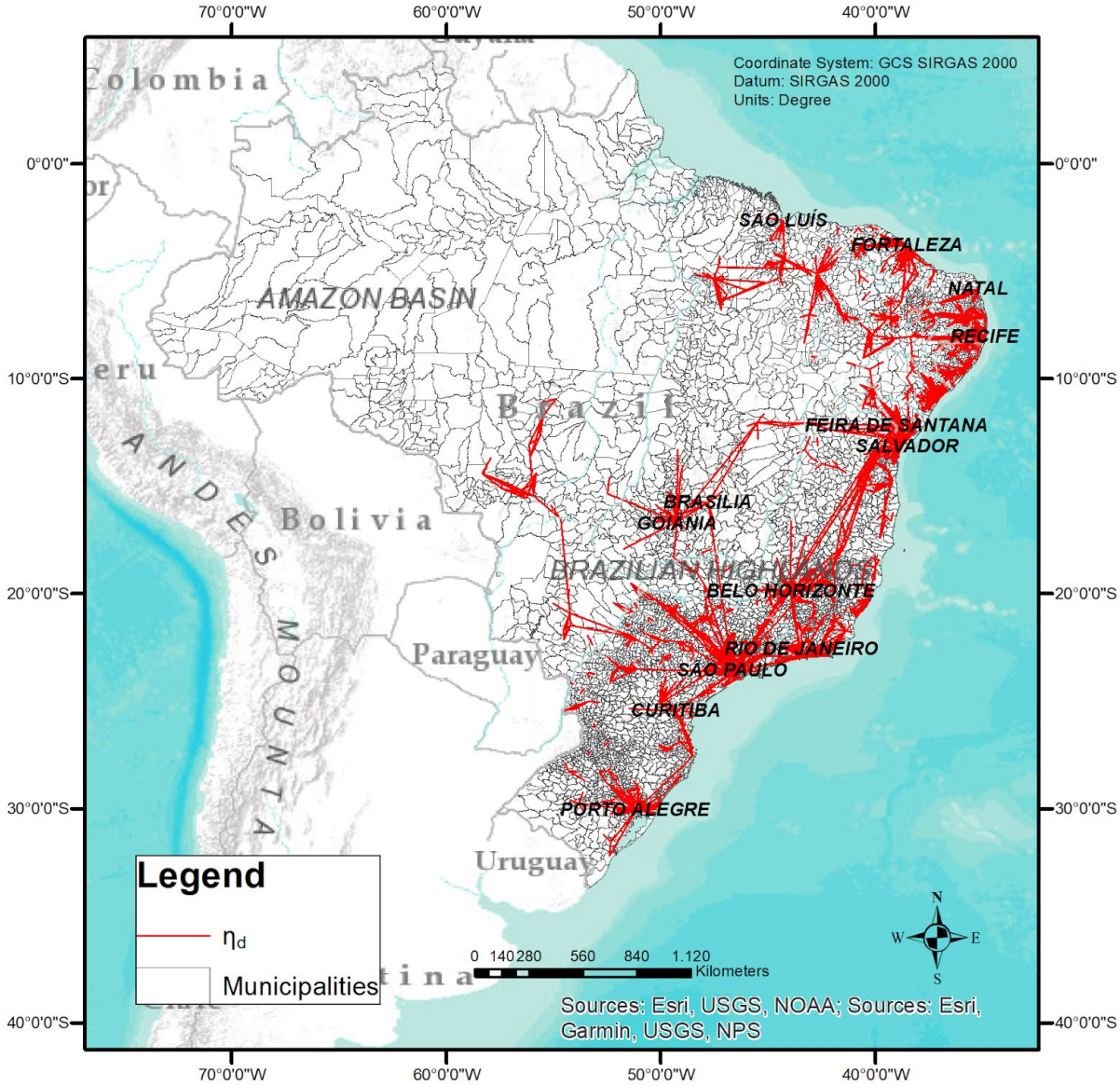

Figure 2: Map of the Brazilian municipalities with the sets of edges for η_d. The edges for η_0 are not plotted, because there are more than 59000 and the visualization was not clear.

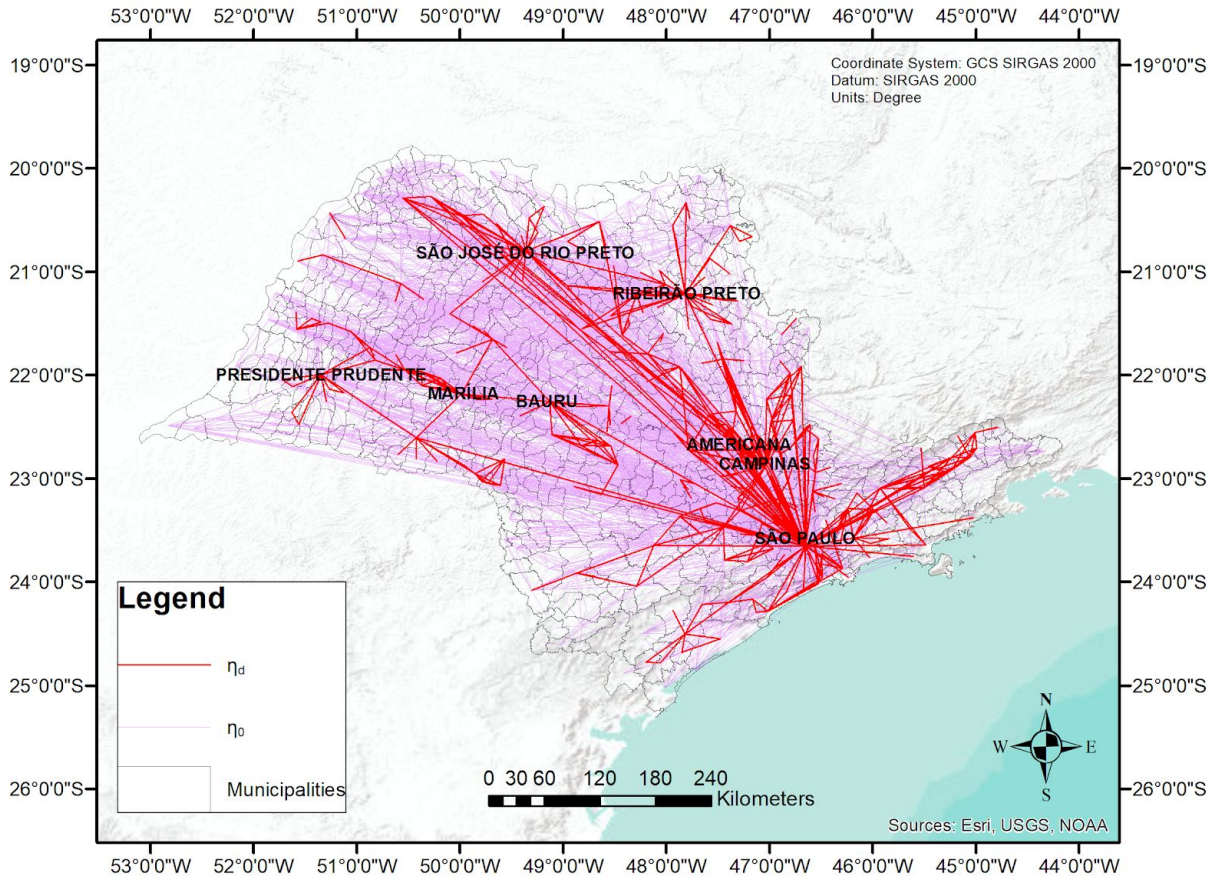

Figure 3: Map of the municipalities of the São Paulo state with the sets of edges for η_0 and η_d.

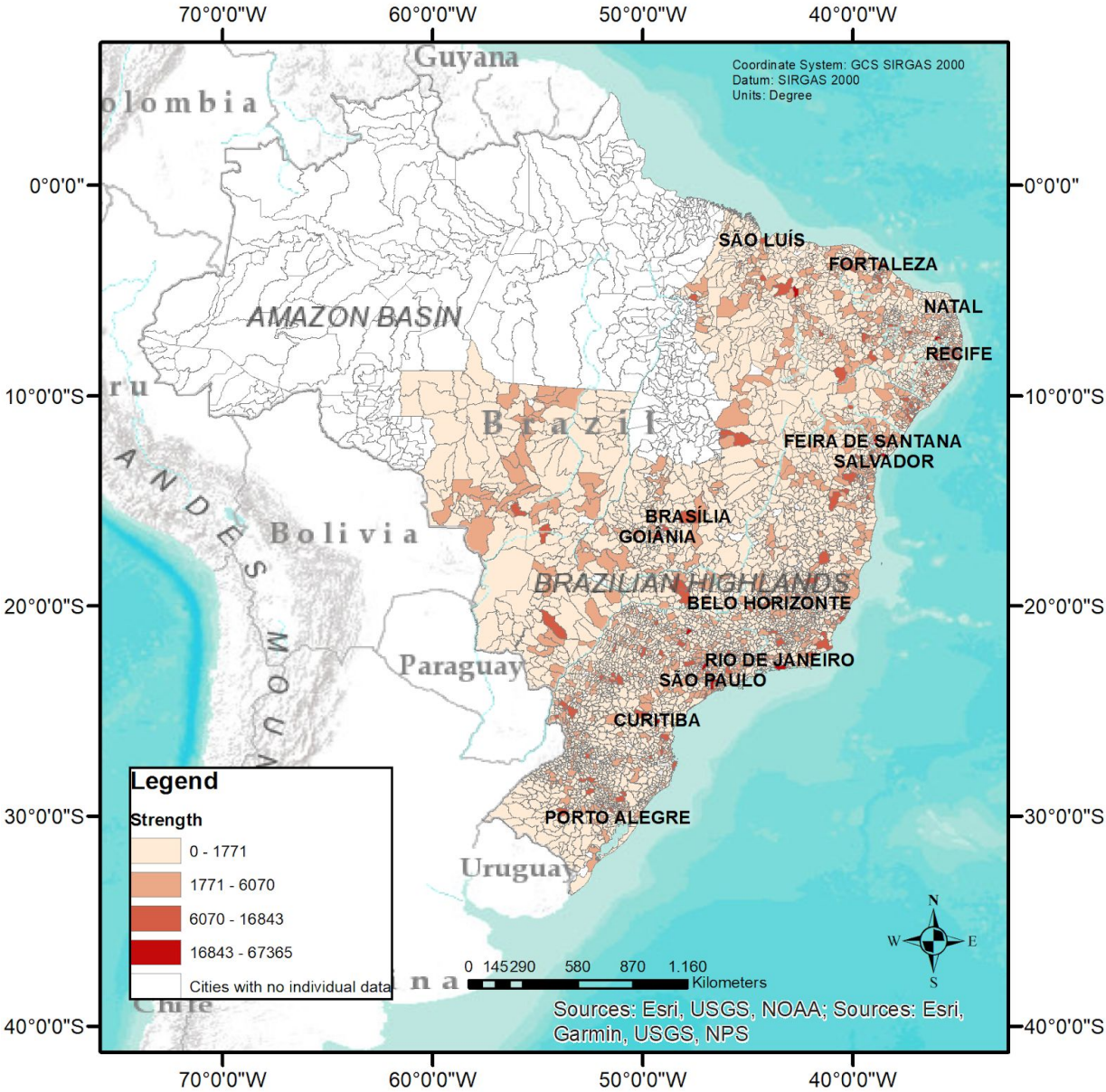

Figure 4: Map of the topological strength related to each node/city of the BRWN network, for η_0.

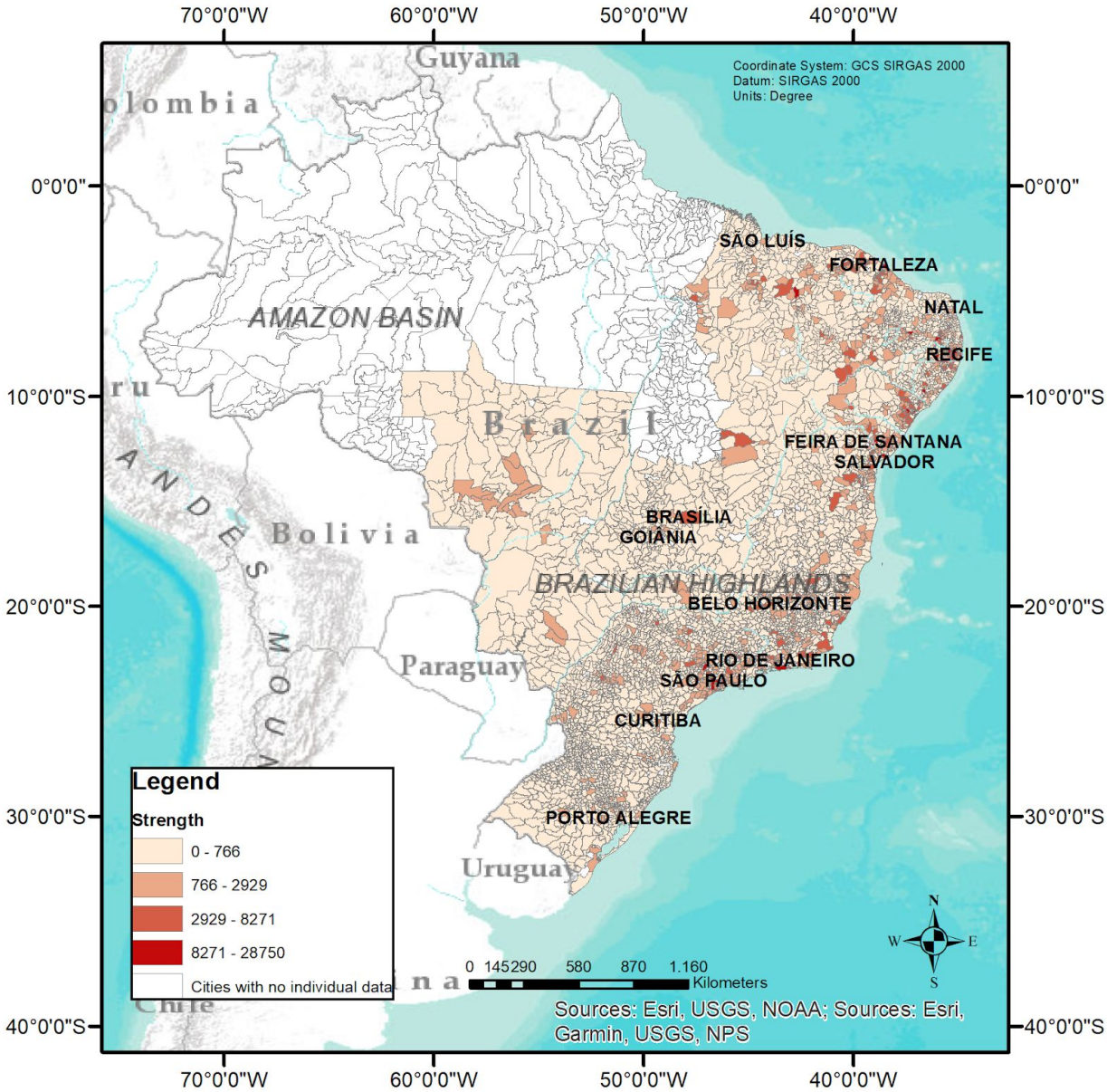

Figure 5: Map of the topological strength related to each node/city of the BRWN network, for η_d.

Figures 6 and 7 present the strength for the SP network, with η_0 and η_d, respectively. Some cities with high strength also appear in a report[17] of most vulnerable cities to COVID-19 due to their intense traffic of people, namely São Paulo, Campinas, São José do Rio Preto, São José dos Campos, Ribeirão Preto, Santos, Sorocaba, Jaboticabal, Bragança Paulista, Presidente Prudente, Bauru, and many others. Currently, they all have a significant number of confirmed cases.

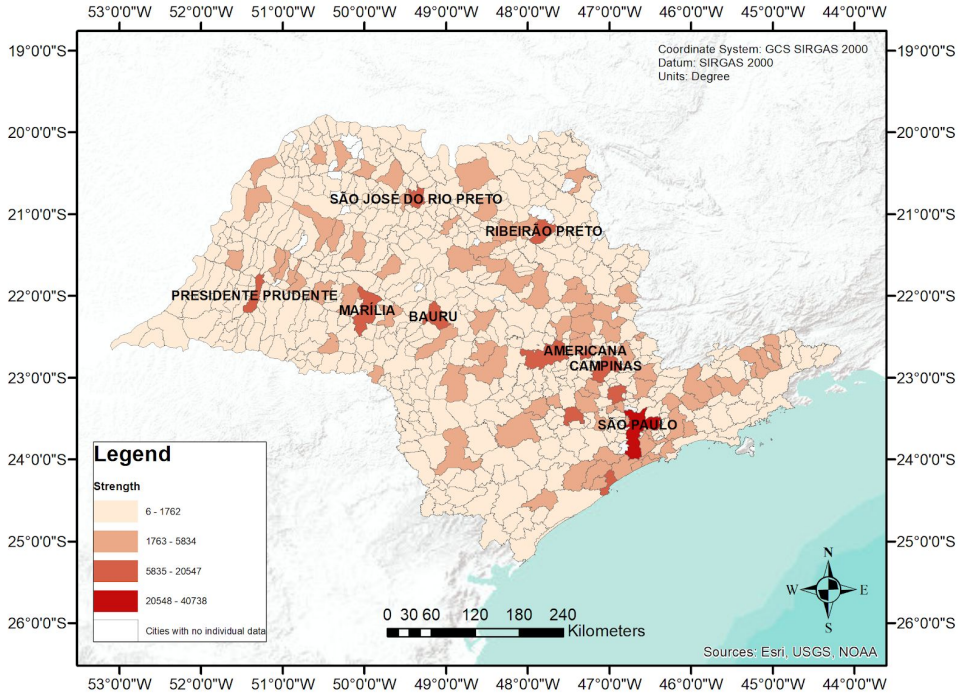

Figure 6: Map of the topological strength related to each node/city of the SP network, for η_0.

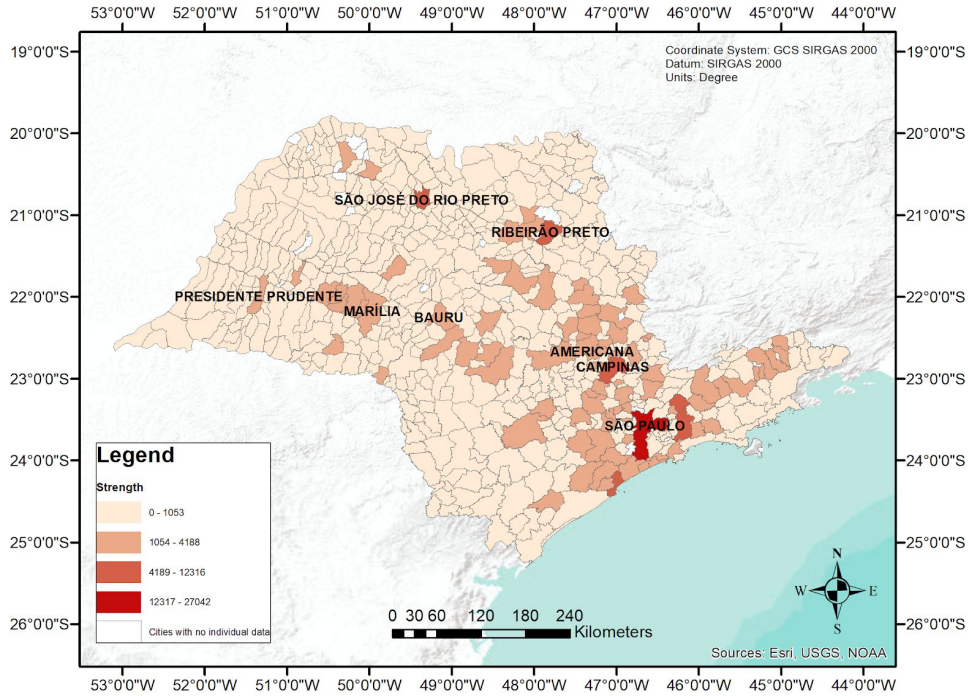

Figure 7: Map of the topological strength related to each node/city of the SP network, for η_d.

We now assess which of the computed measures (s, k, and b) better approximates the emergence of COVID-19 in Brazil. We compare the top-ranked $n \in [1, X]$ cities of each measure with the n cities that contain confirmed cases. According to the available data of the notified cases from daily state bulletins of the Brazilian Health Ministry[10], until May 1st, 2020, the number of cities with at least one confirmed patient with COVID-19 is $X = 1902$ in the BRWN network, which corresponds to 38% of the nodes, and $X = 323$ in SP (52% of the nodes). This provides a way of tracking the response of each measure in detecting vulnerable cities according to the evolution of the virus spreading process.

Some cities from the aforementioned data are not present in our network, due to a simplification that the IBGE does: it groups small neighboring cities with almost no flow into single nodes. For simplicity, and considering that such cities do not contain cases in the first days of the outbreak, they are not accounted for in our analysis.

In order to verify whether the rate of correspondence between the top-ranked cities from the networks' measures and the cities with COVID-19 cases has statistical significance, we verify what are the results of picking cities at random instead of under the measures' guidance. We perform 10^5 simulations for each $n \in [1, X]$, choosing n nodes by sort and monitoring what is the rate p of positive cases. Figure 8 presents the correspondence of the first n cities with COVID-19 documented cases and both the simulated data and the top-ranked nodes under s, k, and b. The gray region represents 95% of the rates' occurrences in the simulations for each n, and the maximum observed value is the dashed line.

In our analysis, on May 1st, about 95% of the simulations have matching rates within 0.38 ± 0.01 for the BRWN network, and the same volume is within 0.52 ± 0.03 for the SP. The results for node selection during the first days via the network indexes all lie above the dashed line, which means that all indexes are a better heuristic than picking nodes at random in the beginning. However, immediately after April 21st, k with η_0 and b for both thresholds start to cross the dashed line in SP, having results compared to the simulations. Those three curves become to have the worst results for BRWN as well, after a transient.

Oscillations are perceived in Figure 8 a) for small n, but they stabilize afterward and follow a tendency. The matching p is at maximum in the beginning, because the first documented case was in the city of São Paulo, which is the first ranked city in all measures. The curve then decreases until reaching a region where the oscillations take place.

The network quantifiers pose good correspondences already in the beginning of the spreading process as the dashed line is not touched until n approaches X. The high-frequency oscillations of Figure 8 a) are pronounced up to March 24th ($n \approx 150$). That is probably the transient needed for the spreading process to reach a more steady behavior.

There is no mark on March 24th in Figure 8 b), because the number of new cities with confirmed cases is negligible in the period. Interestingly, on March 31st, a week later, the high-frequency oscillations start to diminish in SP. A few days further, after April 7th, the betweenness centrality with η_d starts to be a bad predictor for BRWN and then for SP.

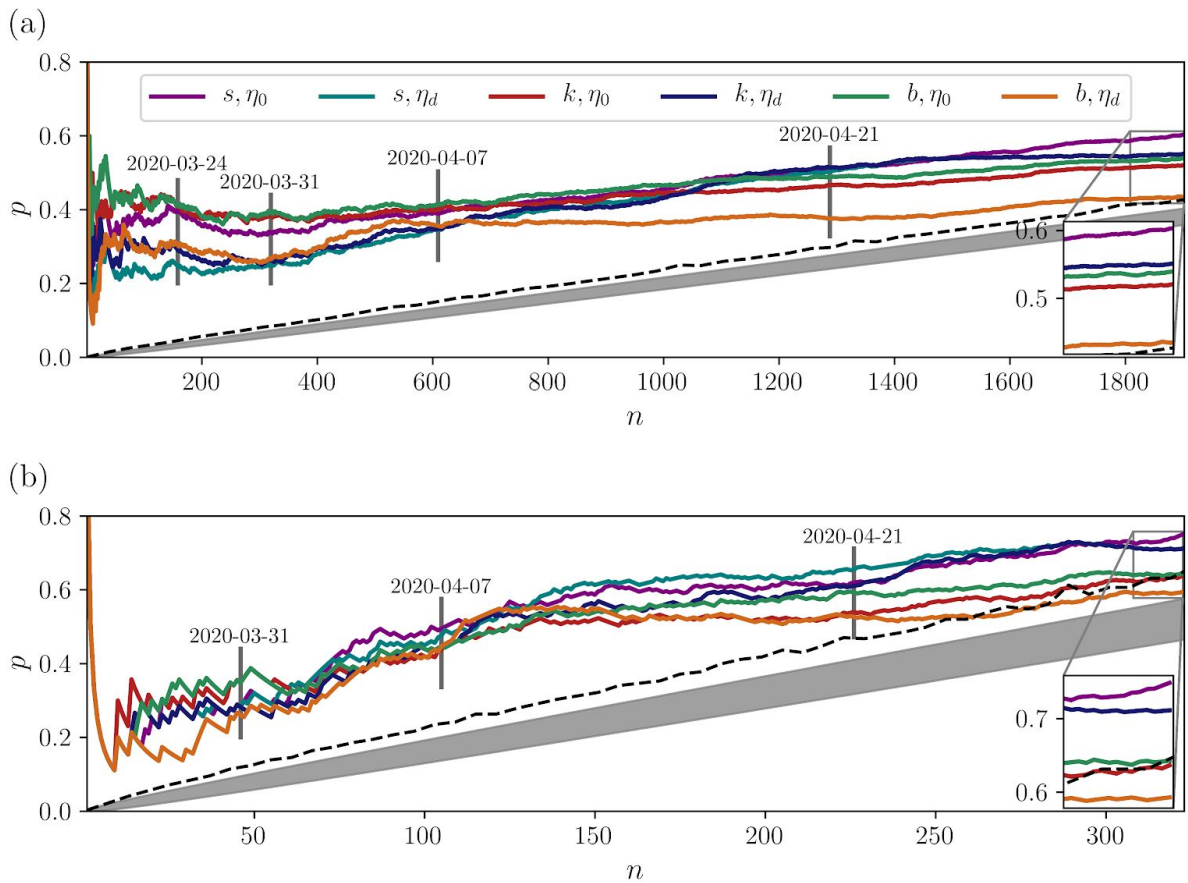

Figure 8: Rate of the first $n \in [1, X]$ ranked cities (s, k, and b) that have at least one patient with COVID-19 in: a) BRWN ($X = 1902$) and b) SP ($X = 323$), for thresholds η_0 and η_d, as in Figure 1. The gray region indicates 95% of 10^5 simulations for each n and the dashed line is the maximum simulated rate. The inset is a zoomed area of the last days until May 1st. In both panels, the s and k with η_d are overlapping in the end.

Following, we quantitatively evaluate the curves from Figure 8 and others with different thresholds, to check exactly which η better captures the spreading process of COVID-19 in the mobility network. Figure 9 displays the integral R of each of those curves with $\eta \in [0, \mu + 2\sigma]$, in which μ is the average flow of the network and σ is the standard deviation. The η_d is marked with the vertical line, showing to be a good threshold in SP, but bad in BRWN. While for SP the strength is always the best measure, there is a certain oscillation in BRWN, where both the s and b are the best predictors for small threshold, switching to b at $\eta \approx 45$ and then to k at $\eta \approx 110$. The best prediction is given by betweenness with $\eta \approx 60$, and similar results are captured by both s and b at η_0. When it comes to the SP network, the η_d captures the exact point where s has the best outcome.

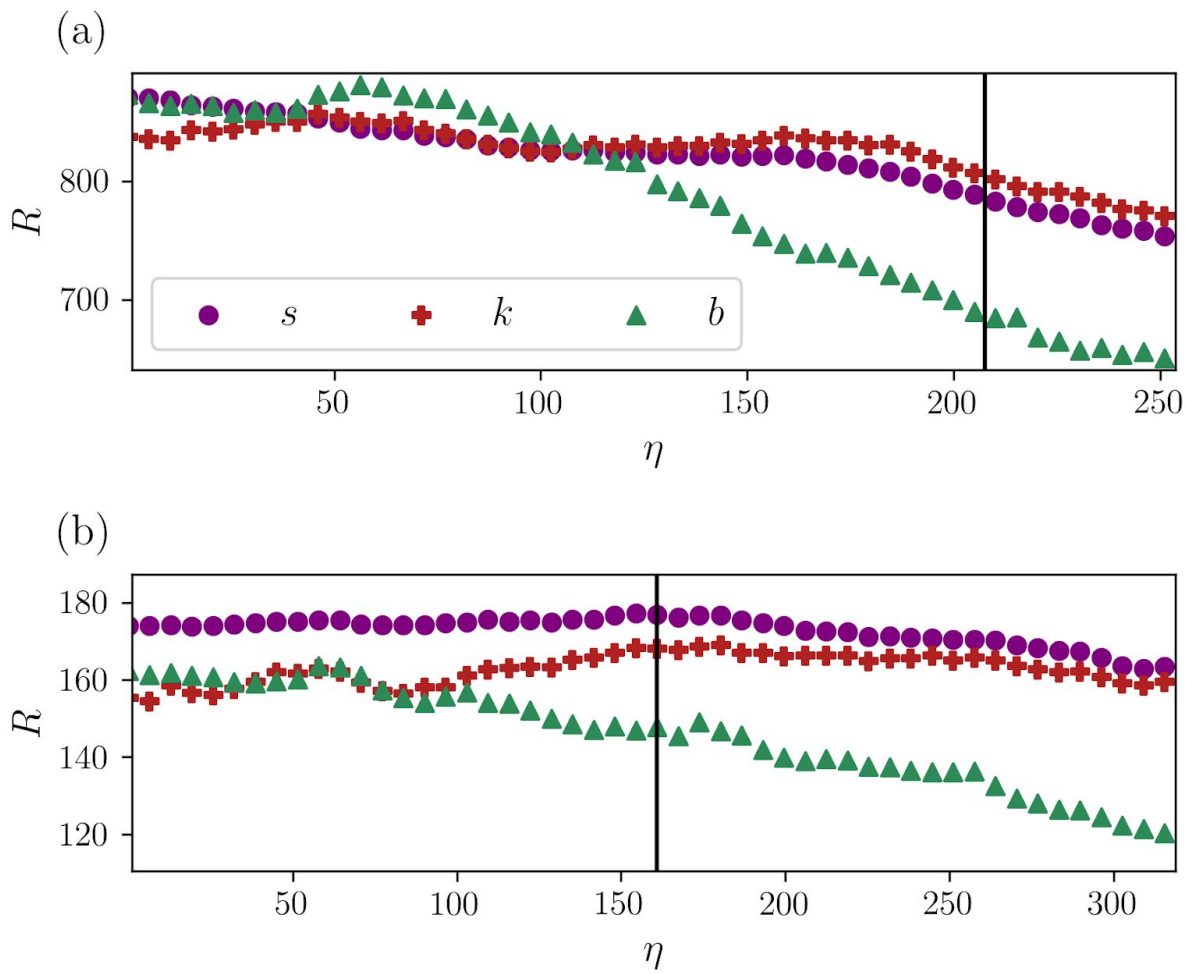

Figure 9: Integral R of each curve of Figure 8 and for other thresholds within $\eta \in [0, \mu + 2\sigma]$, in which μ is the average flow of the network and σ is the standard deviation: a) BRWN, and b) SP. Both η_0 and η_d are included in the interval and the integral related to η_d is marked with the vertical line. Notice that here the abscissa is η and not n.

Table 2 enumerates the first twenty ordered cities according to the best-evaluated measures and compares them side-by-side with the first twenty cities with COVID-19 cases in the BRWN network. The best measures for SP are compared with each other in Table 3 as well. In both networks, the metrics present high-frequency oscillations in the beginning as shown in Figure 8, but still have some correspondences with the first confirmed cases.

TABLE 2

Cities with at least one case of COVID-19 in Brazil (BRWN) in the order they were documented[10], side-by-side with the top-ranked cities regarding s, k *and* b *for* η_0 *and* η_d. *The best combination is* s *with* η_0 *(second column). Matching cities are colored alike.*

	COVID-19	s, η_0	k, η_0	b, η_0
1	3550308 São Paulo	3550308 São Paulo	3550308 São Paulo	3550308 São Paulo
2	3300407 Barra Mansa	3106200 Belo Horizonte	3106200 Belo Horizonte	3106200 Belo Horizonte
3	2910800 Feira de Santana	2927408 Salvador	3509502 Campinas	5208707 Goiânia
4	3304557 Rio de Janeiro	2800308 Aracaju	5300108 Brasília	5300108 Brasília
5	5300108 Brasília	3509502 Campinas	5208707 Goiânia	3304557 Rio de Janeiro
6	3547304 Santana de Parnaíba	2304400 Fortaleza	3304557 Rio de Janeiro	4314902 Porto Alegre
7	3122306 Divinópolis	4314902 Porto Alegre	4314902 Porto Alegre	2304400 Fortaleza
8	2704302 Maceió	2504009 Campina Grande	2927408 Salvador	4106902 Curitiba
9	4303905 Campo Bom	2910800 Feira de Santana	2910800 Feira de Santana	3509502 Campinas
10	4314902 Porto Alegre	2611606 Recife	4106902 Curitiba	2408102 Natal
11	4305108 Caxias do Sul	2604106 Caruaru	4104808 Cascavel	2910800 Feira de Santana
12	4105508 Cianorte	5208707 Goiânia	3543402 Ribeirão Preto	2927408 Salvador
13	4106902 Curitiba	3304557 Rio de Janeiro	2304400 Fortaleza	2211001 Teresina
14	3515707 Ferraz de Vasconcelos	2604007 Carpina	4115200 Maringá	3543402 Ribeirão Preto
15	4205407 Florianópolis	2211001 Teresina	4304705 Carazinho	2111300 São Luís
16	5208707 Goiânia	3543402 Ribeirão Preto	5201108 Anápolis	2504009 Campina Grande
17	3131307 Ipatinga	4106902 Curitiba	3549805 São José do Rio Preto	2611606 Recife
18	2408102 Natal	2608909 Limoeiro	2933307 Vitória da Conquista	2507507 João Pessoa
19	3303302 Niterói	2610608 Paudalho	2903201 Barreiras	2800308 Aracaju
20	2611606 Recife	2510808 Patos	5002704 Campo Grande	3549805 São José do Rio Preto

TABLE 3

Cities with at least one case of COVID-19 in the State of São Paulo (SP) in the order they were documented[10], side-by-side with the top-ranked cities regarding s, k and b between η_0 and η_d. The best combination is s with η_d (second column). Matching cities are colored alike.

	COVID-19	s, η_d	k, η_d	b, η_0
1	3550308 São Paulo	3550308 São Paulo	3550308 São Paulo	3550308 São Paulo
2	3547304 Santana de Parnaíba	3509502 Campinas	3509502 Campinas	3509502 Campinas
3	3515707 Ferraz de Vasconcelos	3543402 Ribeirão Preto	3543402 Ribeirão Preto	3549805 S. José do Rio Preto
4	3529401 Mauá	3537602 Peruíbe	3549805 S. José do Rio Preto	3543402 Ribeirão Preto
5	3547809 Santo André	3549805 S. José do Rio Preto	3552205 Sorocaba	3541406 Presidente Prudente
6	3548708 S. Bernardo do Campo	3546801 Santa Isabel	3506003 Bauru	3506003 Bauru
7	3548807 S. Caetano do Sul	3530607 Mogi das Cruzes	3541406 Presidente Prudente	3552205 Sorocaba
8	3518800 Guarulhos	3525904 Jundiaí	3529005 Marília	3538709 Piracicaba
9	3505708 Barueri	3552205 Sorocaba	3555000 Tupã	3529005 Marília
10	3509502 Campinas	3538709 Piracicaba	3549904 S. José dos Campos	3502804 Araçatuba
11	3513009 Cotia	3549904 S. José dos Campos	3538709 Piracicaba	3501608 Americana
12	3524709 Jaguariúna	3541406 Presidente Prudente	3525904 Jundiaí	3504008 Assis
13	3534401 Osasco	3541000 Praia Grande	3554102 Taubaté	3511102 Catanduva
14	3549805 S. José do Rio Preto	3527207 Lorena	3550605 São Roque	3524808 Jales
15	3549904 S. José dos Campos	3550605 São Roque	3546801 Santa Isabel	3554102 Taubaté
16	3552502 Suzano	3548906 São Carlos	3540705 Porto Ferreira	3549904 S. José dos Campos
17	3554102 Taubaté	3554102 Taubaté	3537602 Peruíbe	3525904 Jundiaí
18	3556453 Vargem Gde. Paulista	3551009 São Vicente	3527207 Lorena	3516200 Franca
19	3519071 Hortolândia	3542602 Registro	3524303 Jaboticabal	3524303 Jaboticabal
20	3530607 Mogi das Cruzes	3551702 Sertãozinho	3518404 Guaratinguetá	3547809 Santo André

Regarding Table 2, some cities are captured by the three measures but do not appear in the first column, namely Fortaleza (CE), Salvador (BA), Campinas (SP), Ribeirão Preto (SP) and Belo Horizonte (MG), but they soon had patients with COVID-19. Interestingly, the city of Feira

de Santanta (BA) appears in all columns - it is the second-largest city of the state and connects the capital to the countryside of Bahia[1]. Oppositely, the city of João Pessoa, capital of Paraíba state (PB) does not appear in the top 20 of the second column (best measure - see Figures 7 and 8), but two other cities from the state do, namely Campina Grande (PB) and Patos (PB). Campina Grande and Patos are among the five richest cities of Paraíba[2]. Note that within the context of an epidemic, such cities are potential super spreaders along with the states' capitals. Five cities of Pernambuco state (PE) appear in the second column (best measure - see Figures 7 and 8), namely Caruaru, Carpina, Limoeiro, Paudalho, and Recife. Pernambuco is currently ranked as the second state in the number of confirmed cases of the Northeast region[(10)].

Table 3, as in Table 2, also displays cities that are captured by the three rightmost columns that do not appear in the first, showing their high level of vulnerability: Ribeirão Preto, Jundiaí, Sorocaba, Piracicaba, and Presidente Prudente. They all have documented cases before May 1st, though. Our study also captured the most influential cities that had cases already in the beginning, like São Paulo, Campinas, São José dos Rio Preto, São José dos Campos and Taubaté. Other cities appear in the second column (best metric) but not in the first: Praia Grande, São Vicente, São Carlos, Registro, Sertãozinho.

Due to their importance in mobility, many cities of Table 3, especially in the second column, appear in the report[(17)] on the vulnerability of microregions of São Paulo state to the SARS-CoV-2 pandemic of April 5th either as potential spreaders or places with a high probability of receiving new cases. They all have notified cases by May 1st and some have the highest numbers of São Paulo state[3].

Both s and b with η_0 pose good results at the beginning of the pandemics for the BRWN network, but s alone started to be the best predictor from the end of April. The most important cities, due to their high flow of travelers and their role in the most used routes, are reached first, followed by those with smaller flows, probably because of the interiorization of the virus - the outbreak reaching the countryside cities. This behavior is even more pronounced in SP, in which s under η_d is the best option at first, neglecting lower flow venues, but the η_0 started to be the best option from the end of April.

In the ongoing pandemics, from May 1st, the s index with η_0 is currently the best predictor and may help to figure out which countryside cities are about to receive new cases. Moreover, it may help in the following waves of the disease. In the case of another pandemic, one could first compute the strength of the networks according to the last updated data from IBGE and identify the top-ranked cities. In Brazil, it is enough checking on strength at the original data, as we presented, since it produces similar results as the betweenness centrality and is computationally cheaper to obtain. Regarding the state of São Paulo, one better checks on the strength index with threshold η_d in the first weeks and only then switch to η_0. As our results

[1] Source (IBGE): https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/25278-ibge-divulga-as-estimativas-da-populacao-dos-municipios-para-2019

[2] https://www.ibge.gov.br/explica/pib.php

[3] http://www.saude.sp.gov.br/resources/cve-centro-de-vigilancia-epidemiologica/areas-de-vigilancia/doencas-de-transmissao-respiratoria/coronavirus/coronavirus010520_65situacao_epidemiologica.pdf

show, the correspondence has statistical significance and, along with other information about the regions such as where are the first notified cases, the pandemic could be closely traced.

FINAL REMARKS

We present a complex network-based analysis in the Brazilian inter-cities mobility networks towards the identification of cities that are vulnerable to the SARS-CoV-2 spreading. The networks are built with the IBGE terrestrial mobility data from 2016 that have the weekly flow of people between cities. The cities are modeled as nodes and the flows as weighted edges and the geographical graphs, (geo)graphs, are visualized within Geographical Information Systems.

Two scales are investigated, the Brazilian cities without the North region, and the state of São Paulo. The former does not account for the North due to the high number of fluvial routes and some intrinsic local characteristics that are not represented with the terrestrial data. The state of São Paulo is important in the ongoing pandemic since the first documented case was in the state capital and it is currently one of the main focus of the virus spreading.

Three network measures are studied, namely the strength, degree, and betweenness centrality, under several flow thresholds to account for different mobility intensities, ranging from the original flow data to networks with only the edges with higher weights. We verified that the strength has the best matching to the cities with COVID-19 confirmed cases. Moreover, the strength measure with the original flows showed to be the best option for Brazil. Oppositely, a more restricted threshold culminates in better correspondences at the beginning of the pandemic in SP. Probably due to the interiorization of the spreading process, a transition is observed after a certain point, when the original flows have better results as the connections to smaller cities are only present when they are accounted for.

Surprisingly, some countryside cities such as Campina Grande (state of Paraíba), Feira de Santana (state of Bahia), and Caruaru (state of Pernambuco) have higher strengths than some states' capitals. Furthermore, some cities from the São Paulo state such as Presidente Prudente and Ribeirão Preto are captured in the top-rank positions of all the analyzed network measures under different flow thresholds. Their importance in mobility is crucial and they are potential super spreaders like the states' capitals.

As future work, we intend to analyze aerial and fluvial mobility data as well, as they include valuable information about the transport of people and goods. The former is fundamental to the discussion of the dynamics for the Brazilian North region, especially the Amazon, and the latter captures long-range connections. Lastly, one could check for correspondences between the networks' measures and data from other epidemic outbreaks.

AUTHOR'S CONTRIBUTION

V. L. S. Freitas and L. B. L. Santos conceived the original idea, methodology, collected the data, interpreted results, wrote the paper and developed the methodology. J. Feitosa and C. S. N. Sepetauskas developed the methodology. All authors reviewed and approved the final version. The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We would like to thank Jussara Angelo (Fiocruz, Rio de Janeiro) for the valuable discussions.

REFERENCES

1. E. Estrada, The structure of complex networks: theory and applications, Oxford University Press, 2012.
2. A. L. Barabási, et al., Network science, Cambridge university press, 2016.
3. H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, M. Tomasini, Human mobility: Models and applications, Physics Reports 734 (2018) 1–74.
4. R. M. Lana, M. F. C. Gomes, T. F. M. Lima, N. A. Honorio, C. T. Codeço. The introduction of dengue follows transportation infrastructure changes in the state of Acre, Brazil: A network based analysis. PLoS Negl Trop Dis 11(11) (2017)
5. J. R. Angelo, T. H. Katsuragawa, P. C. Sabroza, L. A. S. de Carvalho; L. H. P. da Silva, C. A. Nobre. The role of spatial mobility in malaria transmission in the Brazilian Amazon: The case of Porto Velho municipality, Rondônia, Brazil (2010-2012). PLoS One, v. 12 (2017), p. e0172330
6. H. Saba, M. A. Moret, F. R. Barreto, M. L. V. Araújo, E. M. F. Jorge, A. S. Nascimento Filho, J. G. V. Miranda, Relevance of transportation to correlations among criticality, physical means of propagation, and distribution of dengue fever cases in the state of bahia, Science of The Total Environment (2018)
7. L. B. L. Santos, L. M. Carvalho, W. Seron, F. C. Coelho, E. E. Macau, M. G. Quiles, A. M. Monteiro, How do urban mobility (geo) graph's topological properties fill a map?, Applied Network Science 4 (1) (2019) 1–14.
8. F. C. Coelho, R. M. Lana, O. G. Cruz, C. T. Codeco, D. Villela, L. S. Bastos, A. P. y Piontti, J. T. Davis, A. Vespignani, M. F. Gomes, Assessing the potential impacts of covid-19 in Brazil: Mobility, morbidity and impact to the health system, medRxiv.
9. Worldometer, Coronavirus in number (2020). Accessed on May 15th, 2020. Available at: <https://www.worldometers.info/coronavirus/>.
10. W. Cota. Monitoring the number of COVID-19 cases and deaths in Brazil at municipal and federative units level. (2020). Scientific Electronic Library Online (SciELO). DOI https://doi.org/10.1590/SciELOPreprints.362
11. IBGE - Instituto Brasileiro de Geografia e Estatística, Censo demográfico 2010: Resultados gerais da amostra. (2020). URL <https://censo2010.ibge.gov.br/resultados.html>

12. IBGE. (2017). Ligações rodoviárias e hidroviárias: 2016. Coordenação de Geografia, Rio de Janeiro, ISBN 9788524044175.
13. L. Santos, M. Costa, S. T. R. Pinho, R. F. S. Andrade, F. R. Barreto, M. Teixeira, M. L. Barreto, Periodic forcing in a three-level cellular automata model for a vector-transmitted disease, Physical Review E 80 (1) (2009) 016102.
14. M. Barthélemy, Spatial networks, Physics Reports 499 (1) (2011) 1–101.
15. L. B. Santos, A. A. Jorge, M. Rossato, J. D. Santos, O. A. Candido, W. Seron, C. N. de Santana, (geo) graphs-complex networks as a shapefile of nodes and a shapefile of edges for different applications, arXiv preprint arXiv:1711.05879 (2017).
16. W. Ceron, L. B. Santos, G. D. Neto, M. G. Quiles, O. A. Candido, Community detection in very high-resolution meteorological networks, IEEE Geoscience and Remote Sensing Letters (2019).
17. P. R. Guimarães Jr, D. Muniz, L. Giacobelli, K. Maia, M. Gaiarsa, A. P. Assis, P. Santana, E. M. Santana, I. Birskis-Barros, L. Medeiros, V. Velásquez, G. Burin, F. Marquitti, W. Dáttilo, M. Cantor, P. Lemos-Costa, R. Raimundo, C. Andreazzi, M. Pires, M. Côrtes, E. Migon. Vulnerabilidade das microrregiões do Estado de São Paulo à pandemia do novo coronavírus (SARS-CoV-2). (2020). Scientific Electronic Library Online (SciELO). DOI https://doi.org/10.1590/SciELOPreprints.49