Gozzi et al. *EPJ Data Science* (2024) 13:69
https://doi.org/10.1140/epjds/s13688-024-00508-8

EPJ Data Science
a SpringerOpen Journal

RESEARCH Open Access

Bridging the digital divide: mapping Internet connectivity evolution, inequalities, and resilience in six Brazilian cities

Nicolò Gozzi[1,2*], Niccolò Comini[2] and Nicola Perra[3,2,4]

*Correspondence: nicolo.gozzi@isi.it
[1]ISI Foundation, Turin, Italy
[2]The World Bank Group, Washington, D.C., USA
Full list of author information is available at the end of the article

Abstract

We investigate the evolution of Internet speed and its implications for access to key digital services, as well as the resilience of the network during crises, focusing on six major Brazilian cities: Belo Horizonte, Brasília, Fortaleza, Manaus, Rio de Janeiro, and São Paulo. Leveraging a unique dataset of Internet Speedtest® results provided by Ookla®, we analyze Internet speed trends from 2017 to 2023. Our findings reveal significant improvements in Internet speed across all cities. However, we find that prosperous areas generally exhibit better Internet access, and that the dependence of Internet quality on wealth have increased over time. Additionally, we investigate the impact of Internet quality on access to critical online services, focusing on e-learning. Our analysis shows that nearly 13% of catchment areas around educational facilities have Internet speeds below the threshold required for e-learning, with disadvantaged areas experiencing more significant challenges. Moreover, we investigate the network's resilience during the COVID-19 pandemic, finding a sharp decline in network quality following the declaration of national emergency. We also find that less wealthy areas experience larger drops in network quality during crises. Overall, this study underscores the importance of addressing disparities in Internet access to ensure equitable digital services and enhance network resilience during crises.

Keywords: Internet speed evolution; Digital inequality; Socioeconomic disparities; Digital services access

1 Introduction

The widespread availability of Internet connectivity has transformed several aspects of our lives. From communication and commerce to education and entertainment, access to reliable, fast, and affordable Internet connectivity has become a critical factor for promoting economic and social development [1–3]. Despite the large overall improvements in technology and adoption witnessed over the last decades, we still observe huge gaps in access to digital services and varying levels of digital literacy. The COVID-19 pandemic has shown the impact of such digital divide and highlighted the importance of addressing it. Indeed, during the acute phases of the crisis, as numerous activities rapidly migrated online, unequal access to a reliable Internet connection affected the possibility to carry out activities from home, increasing the possible exposures to the virus for the unconnected [4–7]. Particularly clear is the negative impact of Internet connectivity disparities on educational achievement, access to tele-medicine, and adoption of remote working [8–15]

In this context, we aim to investigate how Internet connectivity has evolved over the past years across regions and socioeconomic strata, its impact on the access to key services, and its resilience to extraordinary events such as the COVID-19 Pandemic. As a case study, we consider six major Brazilian cities: Belo Horizonte, Brasília, Fortaleza, Manaus, Rio de Janeiro, and São Paulo. Brazil reports one of the highest GINI index in the world [16] and inequality has been one of the main issues affecting its socioeconomic development for decades. On the other hand, Brazil can compete with the most advanced areas in the world when it comes to digital capabilities. Indeed, it hosts cloud services of some of the most important providers, and it is home to several high-tech startups. However, the inequality observed in the socioeconomic dimension, is also reflected in the digital sector. For instance, while the overall average broadband fixed access for every 100 inhabitants is 24 [17] there is a significant heterogeneity among states. Some Brazilian States such as Santa Catarina (36.15) outperform OECD countries (e.g., Italy, 32.1) while others such as Acre (13.8), Amazonas (13.8), and Maranhao (9.9) report remarkably lower figures. Additionally, access and usage of digital tools is far from being inclusive and several areas, even within wealthier states, face a dramatic digital inequality. For these reasons, Brazil constitutes a perfect representation of the complex socioeconomic dynamics and challenges that public and private sector face in addressing the digital gap.

To quantify Internet quality and its evolution in these cities, we leverage a unique dataset provided by Ookla consisting of nearly 100M geolocalized Speedtest results, collected in the time window spanning from 2017 to 2023. We split the analysis in two parts. In the first, we focus on characterising the spatio-temporal evolution of Internet connectivity by exploring differences across socioeconomic indicators. In the second part instead, we study Internet connectivity indicators in the catchment areas of educational activities and quantify the resilience of the digital infrastructure during the COVID-19 Pandemic.

We find significant improvements in Internet quality across all cities considered between 2017 and 2023. Interestingly, we observe a trend towards a more homogeneous distribution of Internet speed, indicating reduced dispersion over the years. However, despite this increased homogeneity, we find an increasing correlation between Internet speed and wealth, with wealthier areas experiencing better Internet access and with this gap widening over time. Furthermore, we also find a noticeable increase in spatial autocorrelation of Internet quality over the years, with the emergence of clusters characterized by high and low speeds.

Furthermore, our analysis reveals that approximately 13% of catchment areas around education facilities experience Internet speeds insufficient for accessing key digital services such as e-learning. Additionally, these areas tend to exhibit lower wealth, suggesting a compounding effect of inequality.

Finally, we assess the impact of the stress placed on the network following the declaration of the COVID-19 national emergency in Brazil. We find that, on average, this caused a –20% in download speed across all cities, with values ranging from –7% in Brasília to almost –30% in Manaus. Our findings indicate that this impact was more pronounced in less wealthy areas compared to more wealthy ones.

Overall, this study highlights that while the evolution of Internet quality showed an overall progress, disparities persist, with socioeconomic factors playing significant roles. Addressing these disparities is crucial to ensure equitable access to digital services and to enhance network resilience in times of crisis. This study demonstrates that despite the resources allocated by the public and private sector to the strengthening of the Brazilian digital infrastructure, investments are still needed, particularly in the less affluent areas.

2 Results

2.1 Internet speed evolution analysis

As a first step, our research aims to analyze the evolution of Internet quality, specifically measured by fixed download speed, across six major Brazilian cities: Belo Horizonte, Brasília, Fortaleza, Manaus, Rio de Janeiro, and São Paulo. To accomplish this, we leverage a unique dataset consisting of $\sim 100M$ Internet Speedtest results. The data covers the period between 2017 and 2023. Furthermore, it is geolocalized and provides the download/upload speed (i.e., Megabits per second) and latency in milliseconds for fixed networks. In the Additional file 1 we show results considering mobile networks, which we also discuss below.

It is important to highlight from the start how the data serves only as a proxy of Internet quality. Indeed, due to the details of the software/tool used to make a measurement, possible bottlenecks in home networks (e.g., routers), the number of devices connected to a specific network, and selection biases (e.g., tests might be done when users are experiencing connectivity issues or when users need to connect in a new location and/or by more digitally aware users) the outcome of tests might differ from the real Internet speed [7, 18]. Nevertheless, Ookla is the *canonical* network performance testing service. It is widely used to infer the features of Internet connectivity across and within regions by academic and governmental institutions [18–20]. Furthermore, as described below, our analysis aggregates Speedtest results within specific geographical cells thus averaging among many measurements. This allows to reduce the possible impact of the more technical issues mentioned.

To ensure uniform spatial coverage we partition the geographical area of each city into hexagonal cells, creating a regular grid (see Fig. 6A). Then, we calculate the Internet speed within each of these units as function of time. This approach allows to explore different resolutions and finer scales with respect to administrative partitions. We also compute a proxy measure for wealth in each of these unit using the Relative Wealth Index (RWI) provided by Meta [21]. For a more detailed description of our methodology, please refer to Sect. 4.

Figure 1 shows the evolution of Internet speed, from 2017 to 2023 in the six cities. Across the board, our analysis reveals a significant improvement in Internet speed throughout all cities over the past six years. Specifically, Belo Horizonte exhibits the highest median download speed (176*mbps*) in 2023, followed by São Paulo (146*mpbs*), Manaus (116*mbps*), Rio de Janeiro (114*mbps*) Fortaleza (111*mbps*), and Brasília (105*mbps*). On the other hand, Manaus experienced the highest growth during the period, marking a +1200% increase, followed by Belo Horizonte (+1012%), Rio de Janeiro (+719%), Fortaleza (+685%), Brasília (+677%), and São Paulo (+463%). Furthermore, in the same plot we show the coefficient of variation of the distribution of Internet speed within each city across the years. The coefficient of variation is a measure of dispersion defined as the ratio between standard

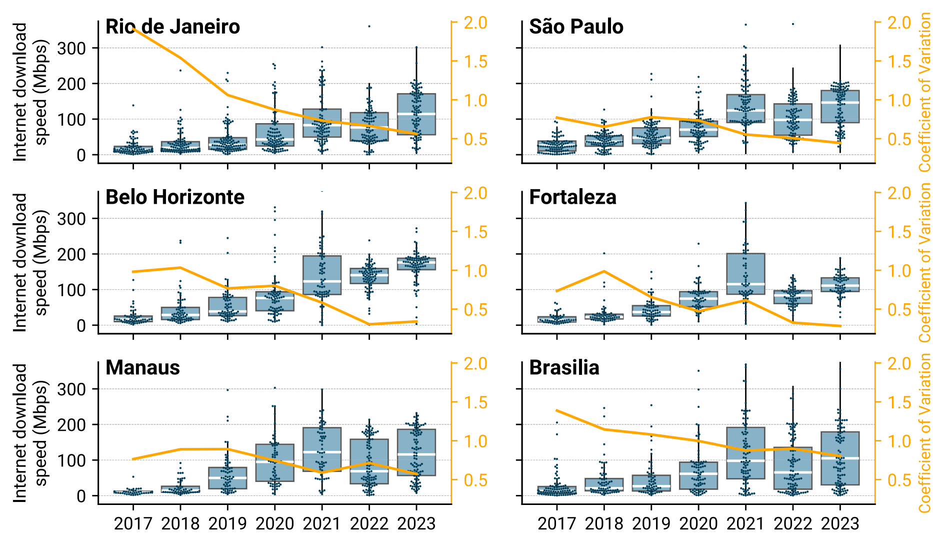

Figure 1 Evolution of Internet speed, expressed in download speed (Mbps), across the six cities considered. Boxplots show the distribution of Internet speed within each hexagonal unit in each city. The orange line indicates the evolution of the coefficient of variation of the Internet speed distribution over the years

deviation and average of a statistical distribution. Our findings indicate a decreasing trend in the coefficient of variation across the six cities, suggesting a persistent trend towards a more homogeneous distribution of Internet speed. However, we acknowledge differences among the cities examined. Brasília exhibits the highest dispersion in Internet speed distribution in 2023 ($CV = 0.80$), while Fortaleza the lowest ($CV = 0.28$). More quantitatively, in 2023, the ratio between the 3rd and 1st quartiles of Internet speed is 5.8 in Brasília, whereas it is only 1.4 in Fortaleza.

It is important to highlight how, despite a general trend towards homogenisation, the data still reveals persistent and even increasing disparities across socioeconomic strata. Figure 2 shows the logarithm of the ratio between the average Internet speed measured in cells with wealth higher than the 75th quantile and those with wealth lower than the 25th quantile. This metric is meant to compare and highlight the differences between the wealthiest and the poorest units. A value close to zero indicates similar Internet quality for both wealthy and less wealthy areas, while positive (negative) values denote better Internet quality for the more wealthy (less wealthy). As detailed in Sect. 4, the wealth of each unit is calculated using the Relative Wealth Index (RWI) provided by Meta [21].

Across various years and cities, our analysis reveals consistent trends: i) wealthy areas generally experience better Internet quality, ii) disparities across areas increased. In the case of Manaus and São Paulo cells characterized by higher RWI features better Internet quality across the whole time horizon under study. This trend is observed also in Brasilia with the exception of 2017. In Rio de Janeiro instead, only in the last two years Internet quality in wealthy cells was better with respect to less wealthy areas, though the negative values are closer to zero. Finally in Belo Horizonte and Fortaleza, the values are overall smaller with respect to the other cities though positive in the last years. The association between Internet quality and RWI is supported by the Pearson correlation coefficient between Internet speed and RWI, shown in Fig. 2. The coefficient has increased across all

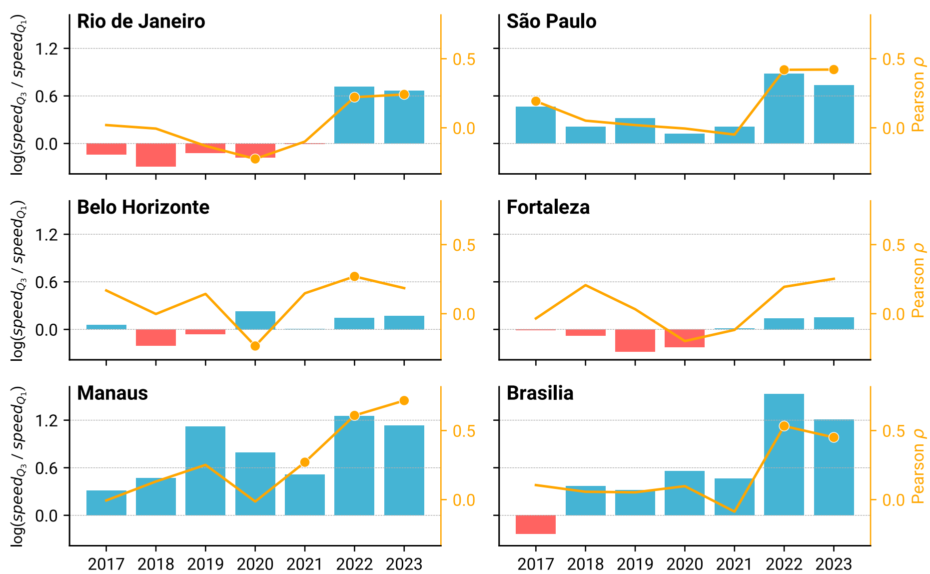

Figure 2 Logarithm of the ratio between Internet speeds measured in spatial units with wealth higher than the 75th quantile and those with wealth lower than the 25th quantile. The orange line represents the Pearson correlation coefficient between Internet speed and RWI of different spatial units. Circles indicate where the coefficient is significant at the 5% level

cities in recent years, with all cities showing a positive correlation as of 2023, which is significant at 5% level with the exception of Belo Horizonte and Fortaleza.

In the Additional file 1 we repeat the analyses presented in Fig. 1 and Fig. 2 for mobile networks (Figure S3 and Figure S4). Also in that case, we find a significant overall improvement of speed over the period considered. Interestingly, we find that, while mobile speed and wealth are also positively correlated. However, the observed trend is decreasing in time, contrasting the findings for fixed networks.

In the case of Rio de Janeiro, we extend our analysis to include tests conducted both inside and outside *favelas*. The results of this analysis are presented in the Additional file 1 (Figure S1). *Favelas* are informal, densely populated urban settlements in Brazil, typically characterized by substandard housing and a lack of basic services, arising from socioeconomic disparities and rapid urbanization. Not surprisingly, we find that tests performed within a *favela* generally exhibit lower Internet speeds. Additionally, this disparity has increased over the years. In 2017, the median speed of tests conducted inside and outside *favelas* was 13.7 Mbps and 14.3 Mbps, respectively, reflecting a 4% difference. By 2023, these speeds had changed to 40.1 Mbps and 94.2 Mbps, respectively, resulting in a 57.4% difference.

To investigate whether Internet speed has become more spatially autocorrelated over time, we calculate the Moran's I statistic for download speed in each hexagonal unit across various cities for each year within the study period [22]. The Moran's I quantifies the degree of spatial autocorrelation of a quantity, indicating the extent to which similar values cluster or disperse across geographical units. More in detail, a positive (negative) Moran's I indicates spatial autocorrelation (dispersion) in the dataset, meaning that similar (dissimilar) values tend to cluster together in space. Our analysis reveals the emergence of spatial clusters characterized by high or low Internet speed. This finding is exemplified

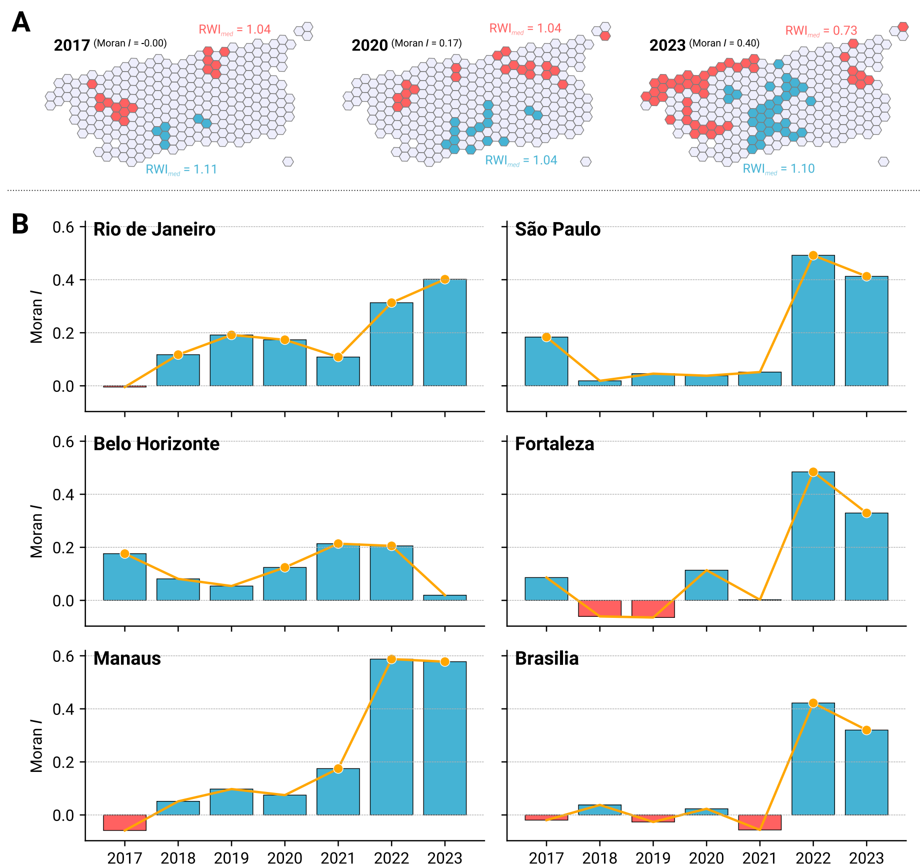

Figure 3 Spatial Clustering of Internet Speed. **A)** Distribution of spatial units with significant local Moran's *I* in Rio de Janeiro in 2017, 2020, and 2023. Clusters of low (high) Internet speed are shown in red (blue). Figure also reports the median RWI of high/low Internet speed clusters. **B)** Evolution of global Moran's *I* in each city between 2017 and 2023. Circles indicate where the statistic is significant at the 5% level

in Fig. 3A where we present the results for Rio de Janeiro in 2017, 2020, and 2023 (in the Additional file 1 we show results also for other cities). The global Moran's *I* values exhibit a notable increase from approximately 0 in 2017 to 0.17 in 2020 and further to 0.40 in 2023. Visual inspection of the maps also reveals the emergence of spatial clusters of high and low Internet speed over the years. Specifically, the maps indicate units where the local Moran's *I* statistic — measuring the spatial clustering pattern of individual observations — is significant at the 5% level, with units colored to denote low-speed (red) or high-speed (blue) clusters. The figure also shows the median RWI for high- and low-speed clusters. We observe that the difference between the two is more pronounced in 2023, with the high-speed cluster having a median RWI of 1.10, compared to 0.73 for the low-speed cluster. This supports the pattern of increasing correlation between Internet speed and wealth, as discussed earlier. Furthermore, we analyze the evolution of the global Moran's *I* across different cities over the six-year period. The findings observed in the case of Rio de Janeiro are consistent across cities, with the statistic generally demonstrating an increase over the years. In more details, we observe how in all cities, with the exception of Belo Horizonte, the last two years show the highest values of Moran's *I*. Also, we note how in Brasilia, Fortaleza, and São Paulo, the global Moran's *I*, measured considering data collected in 2023, is smaller with respect to the previous year. The decreasing trend in the last year is also observed, though to a lesser extent, also in the case of Manus and in Belo Horizonte, though in the latter the value obtained it is not significant at the 5% level. In the Additional file 1 we repeat this analysis for mobile network (see Figure S5). Also in this case we find positive and significant spatial autocorrelation of mobile Internet speed, event tough the temporal trend is less clear.

2.2 Access to e-learning

In the second part of our analysis we investigate whether the disparities in Internet quality highlighted in the previous section may impact access to key services. While acknowledging the diversity and variety of these, in the following we use e-learning as a concrete and arguably important example. Indeed, as mentioned in the Introduction, extant research has highlighted the positive relationship between Internet quality and educational attainment [9]. We note how e-learning is a general term referring to both synchronous and asynchronous learning activities. These span from access to dedicated platforms to ability of exploring broader online resources for homework. Our approach is as follows. First, we gather the locations of educational facilities across the six cities under consideration using data from OpenStreetMap [23]. Next, we conduct a Voronoi tessellation for each city, with the positions of educational facilities as centroids. This process allows us to obtain the catchment area of each educational facility. By construction a catchment area describes the closest educational entity for people living in that region. Subsequently, we compute the Internet speed within each catchment area by aggregating the download speeds of all tests performed within. For this analysis, we consider the most recent data from 2023. Our aim is to focus solely on recent data to accurately characterize the current disparities in access to essential services. Additionally, we calculate the RWI for each area. Further details on our methodology are available in Sect. 4. In Fig. 4, we present the distribution of fixed download speeds across all catchment areas in the six cities. We also highlight a threshold of 80 Mbps (approximately 10 megabytes per second) as the minimum speed required to access e-learning services [24]. Remarkably, across all cities we find that nearly 13% of catchment areas have speeds below this threshold. Nonetheless, we observe a significant variability across cities. In Belo Horizonte, none of the catchment areas exhibit an Internet speed below the 80 Mbps threshold. Following closely is Brasília, with only 4.8% falling below, then São Paulo (6.8%), Fortaleza (7.4%), and Manaus (8.3%). In stark contrast, nearly 24% of catchment areas in Rio de Janeiro fall below this threshold.
activities. These span from access to dedicated platforms to ability of exploring broader online resources for homework. Our approach is as follows. First, we gather the locations of educational facilities across the six cities under consideration using data from OpenStreetMap [23]. Next, we conduct a Voronoi tessellation for each city, with the positions of educational facilities as centroids. This process allows us to obtain the catchment area of each educational facility. By construction a catchment area describes the closest educational entity for people living in that region. Subsequently, we compute the Internet speed within each catchment area by aggregating the download speeds of all tests performed within. For this analysis, we consider the most recent data from 2023. Our aim is to focus solely on recent data to accurately characterize the current disparities in access to essential services. Additionally, we calculate the RWI for each area. Further details on our methodology are available in Sect. 4. In Fig. 4, we present the distribution of fixed download speeds across all catchment areas in the six cities. We also highlight a threshold of 80

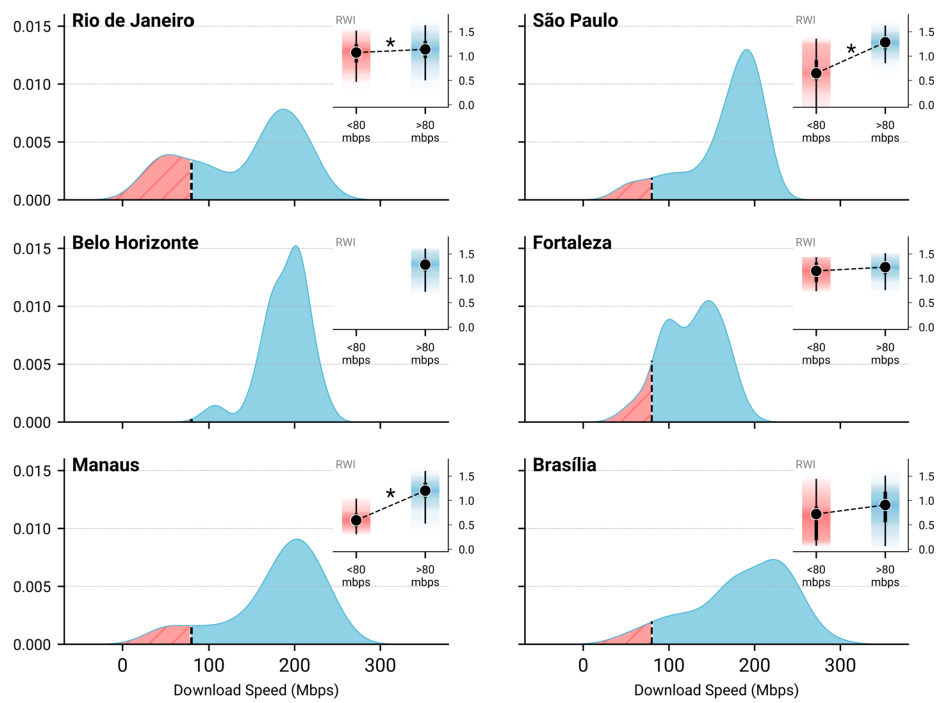

Figure 4 Internet Speed in Catchment Areas of Education Facilities. Distribution of Internet speed, measured as download speed (Mbps), in catchment areas of education facilities across all cities (2023). The portion of the distribution where speed is lower than 80 Mbps is colored in red. In the inset of each plot, RWI distribution of catchment areas of education facilities featuring a download speed above and below 80 Mbps is shown. The asterisks indicate a significant difference between the two RWI distributions assessed using t-test with significance level of 5%

Mbps (approximately 10 megabytes per second) as the minimum speed required to access e-learning services [24]. Remarkably, across all cities we find that nearly 13% of catchment areas have speeds below this threshold. Nonetheless, we observe a significant variability across cities. In Belo Horizonte, none of the catchment areas exhibit an Internet speed below the 80 Mbps threshold. Following closely is Brasília, with only 4.8% falling below, then São Paulo (6.8%), Fortaleza (7.4%), and Manaus (8.3%). In stark contrast, nearly 24% of catchment areas in Rio de Janeiro fall below this threshold.

Additionally, in the inset of each plot in Fig. 4, we display the RWI distribution of catchment areas below and above the 80 Mbps threshold. Across all cities, our analysis indicates that, on average, catchment areas below the threshold are 15% less wealthy than areas above the threshold. We assess the differences in RWI distribution between the two cases using a t-test, finding a significant difference in the case of Rio de Janeiro, São Paulo, and Manaus (significance level 5%). This observation points to a compounding effect of inequality. Indeed, students facing higher challenges in accessing key digital services such as e-learning may already be foreclosed from other opportunities due to their socioeconomic disadvantage.

2.3 Network resilience during crises

Finally, we aim to investigate the resilience of the network to external shocks and the potential heterogeneous impacts of such events. As a case study, we consider the COVID-19 pandemic. With infections and deaths surging worldwide and restrictions being imposed,

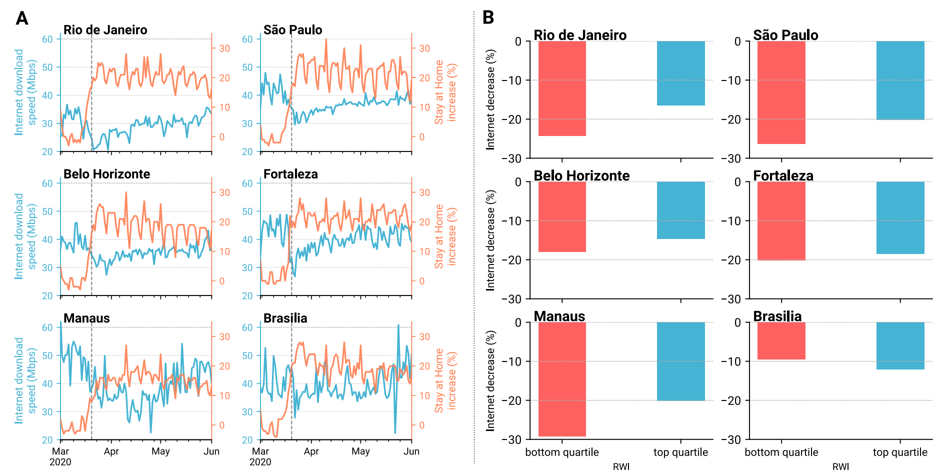

Figure 5 Network Resilience During the COVID-19 Pandemic. **A)** Daily median download speed in the six cities between March and June 2020. The vertical dashed line indicates when Brazil declared the national emergency. The percentage change in individuals staying at home as measured via Google Community Mobility Reports is also shown. **B)** Drop in Internet speed following the national emergency declaration in the top and bottom quartiles of the RWI in each city

the world moved online to maintain essential activities. Arguably, such an unprecedented surge in demand may have affected network quality. In Fig. 5A, we present the median daily download speeds in the six cities between March and June 2020. Additionally, we mark the date when Brazil declared a national emergency with a vertical dashed line and we show the increase in the percentage of individuals staying at home measured using data from the COVID-19 Community Mobility Reports published by Google [25]. Across all cities, we observe a sharp decline in network quality, as measured by download speed, following the declaration of the national emergency. Concurrently, the fraction of population staying at home increased. After the initial drop, we observe a gradual recovery, with download speeds approaching pre-emergency levels by June 2020. Among the cities considered, Manaus experienced the most significant drop in median download speed computed in periods March 1st-March 20th (pre-emergency) and March 20th-April 1st (post-emergency), with a decline of –29%, while Brasília showed the lowest drop at –7%. The other cities experienced declines ranging from Rio de Janeiro (–25%), Fortaleza (–21%), São Paulo (–19%), to Belo Horizonte (–16%).

Furthermore, in Fig. 5B, we illustrate these drops for the top and bottom quartiles of the RWI. More in detail, we select all spatial units in top and bottom RWI quartile and compute median daily Internet speed in these two groups. Then, we aggregate speed in pre- and post-emergency periods, and compute the percentage differences shown in Figure. We observe that, with the exception of Brasília, more wealthy areas experienced smaller drops compared to less wealthy areas. These differences are also significant across all cities when compared using a t-test with a 5% significance level. When combined with the previous findings, this suggests that besides experiencing slower Internet speeds, less wealthy areas may also face more significant drawbacks during extraordinary stress on the network.

In the Additional file 1 we repeat this analysis for mobile network. Also in that case we find that mobile Internet speed was significantly affected by the stress put on the network following national emergency declaration, even tough we do not observe a clear divide in the drops experienced by more and less wealthy areas (see Figure S6).

3 Discussion

In this study, we analysed the spatio-temporal evolution of Internet speed in six Brazilian cities spanning the years 2017 and 2023. Our analysis revealed a significant increase in Internet speed across all cities, along with a trend towards more uniform distribution. However, we also identified the emergence of spatial clusters characterized by high/low Internet speed. Furthermore, we found an increasing correlation between Internet speed and measures of wealth, indicating that more wealthy areas tended to experience higher Internet speeds over time. Such inequality pattern was also reported by the analysis done in the case of the *favelas* in Rio de Janeiro, which revealed an increasing Internet speed gap with the rest of the city.

To further characterize the impact of such disparities, we considered two case studies. In the first one, we focused on the Internet speed in catchment areas around educational facilities in each city. Notably, we find that, as of 2023, approximately 13% of these areas may have encountered challenges in accessing key digital services such as e-learning. We observed significant variations among cities, with Rio de Janeiro reaching a peak of 24% of these areas falling below the threshold for e-learning. Additionally, we showed that these areas tend to be less wealthy, suggesting a potential compounding effect of inequality, where regions already facing limited access to opportunities may also encounter challenges in digital access.

In our second case study, we examined the unprecedented stress placed on the network due to the shift online driven by the COVID-19 pandemic. Our analysis showed a significant decrease in Internet speed across all cities following the declaration of national emergency. Moreover, we found that less wealthy areas generally experienced more pronounced declines in Internet connectivity during the early weeks of the COVID-19 crisis. This result is even more concerning when combined with findings from a recent study that has shown how access to a fast Internet is an effective measure in case of exogenous shocks such as the pandemic to limit the exposure to infections [7].

In the main text, we focused our analysis on fixed network. The main reason for this choice is that, in the case of mobile Internet speed tests it is less reliable to assume that the individual performing the test resides in the area being assessed. Furthermore, mobile Internet speeds are generally lower, which can prevent activities such as reliable video calls, which is a key focus of the section on e-learning access within this paper. Nonetheless in the Additional file 1 we repeat some of the analyses in the case of mobile networks. Overall, our findings are confirmed also in this case. Additionally, we also show that fixed and mobile Internet speeds tend to positively correlate in different cities. Interestingly, however, in the case of mobile network we observe that over time, the correlation between wealth and speed showed a gradual decline. This phenomenon could be attributed to the higher demand in economically disadvantaged regions for more affordable connectivity options, such as mobile connections. Consequently, the evolution of mobile Internet may have diverged from that of fixed Internet due to distinct demands and consumer segmentation. Although this may provide a short-term fix, it could prevent long-term digital growth and access.

The present study comes with limitations. First, we used data from Internet Speedtest results, which is only a proxy for Internet speed. As discussed, due to several factors, the outcome of tests might differ from the real Internet. Nevertheless, Ookla is widely used by academic and official institutions to measure Internet connectivity. Additionally, our methodology aims at attenuating some of the possible issues deriving from the heterogeneous use of this service, as detailed in Sect. 4 and in Ref. [7]. In the Additional file 1 we conducted an additional analysis on the sample size of Internet measurements, demonstrating that it does not impact the robustness of the results (see Figure S2). We also provide the total number of Internet measurements (both fixed and mobile) across different cities, years, and RWI segments (see Table S1 and Table S2).

Second, we use only proxy data to measure wealth. Indeed, we consider the Relative Wealth Index published by Meta [21] to characterize wealth at the desired spatial granularity, nonetheless such data come with inherent limitations, as is the case with all proxy measures. Lastly, Internet speed and wealth are linked by a feedback loop that we do not fully characterize due to data availability. As a result, our study mostly focuses on associations over time and space rather than causation or providing comprehensive explanations of the current landscape.

Since 2020, about 28 USD billion have been invested in the telecom sector in Brazil [26]. Despite the significant amount of resources, the underlying efforts were not enough to provide a level playing field for all Internet users. This study, indeed, has shown how the poorest segments of population still experience a slower Internet connectivity compared to the most wealthy and how this gap may widen in case of exogenous shocks. Such disparity can have a significant impact on the socioeconomic development of the country and requires a joint work of policy makers and the private sector to be solved. Specific policies at the local level should be promoted to improve connectivity in the poorest areas of towns, favoring the penetration of fiber to the Home (FTTH) technology, the affordability of high-speed Internet packages and devices, the development of specific digital skills through dedicated training and awareness programs. All these measures will support a more equal access to the Internet, ensuring that all individuals have access to a fast, affordable, and reliable Internet connection.

4 Materials and methods

4.1 Measuring Internet speed

We characterize Internet quality using as proxy Speedtest Intelligence® data by Ookla [27]. Speedtest apps offer free analyses of Internet performance metrics. The tests are geolocalized and provide download/upload speed (expressed in Megabits per second). Here, following a common practice, we consider download speed as a metric to assess the quality of Internet. The dataset includes nearly 100M tests performed between 2017 and 2023, divided as follows: 47.5M in São Paulo, 24.1M in Rio de Janeiro, 8.1M in Belo Horizonte, 6.5M in Brasília, 6.1M in Fortaleza, and 4.9M in Manaus.

We preprocess the data by excluding all tests displaying a download speed of 0 Mbps, as these typically represent failed tests and do not provide informative insights into the actual network quality. Additionally, to limit the impact of outliers, we filter out tests with a download speed > 2 Gigabits per second, as this threshold is regarded the maximum value for broadband technology. After preprocessing, we compute Internet speed following a procedure similar to the one presented in Ref. [7].

To compute Internet speed in a geographical area g over a timeframe (t_1, t_2), we gather all tests conducted within that area during that period. Then, we calculate the median of the results obtained from tests conducted by individual users. In other words, for each user u, we compute the associated download speed as follows:

$$Mbps^{u,g}_{(t_1,t_2)} = med_i(Mbps^{u,g}_{i,(t_1,t_2)})$$

This step is taken to prevent bias caused by users who utilize the service more frequently than others. Finally, the download speed associated to area g in timeframe (t_1, t_2) is calculated as the median download speed across all users:

$$Mbps^{g}_{(t_1,t_2)} = med_u(Mbps^{u,g}_{(t_1,t_2)})$$

4.2 Measuring wealth
We assess the socioeconomic status of different geographical regions using the Relative Wealth Index (RWI) from Meta's Data for Good Program [21]. This index, made publicly available in 2021, offers micro-estimates of the relative standard of living within countries. It is built considering non-traditional data sources such as satellite imagery and privacy-preserving Facebook connectivity data, and it is validated by Meta through ground truth measurements obtained from the Demographic and Health Surveys. The RWI covers approximately 93 low and middle-income countries globally, providing data at a high spatial resolution ($2.4km^2$ micro-regions). In this study, we aggregate the RWI at the desired geographical resolution by computing the average RWI of all micro-regions contained in the considered geography. Specifically, for each hexagonal unit, we select the RWI tiles whose centroids fall within that unit and compute the average RWI of those tiles, which we then assign to the corresponding hexagonal unit.

4.3 Hexagonal grid
We partition the area of each city considered into an hexagonal grid. This allows us to obtain a regular uniform spatial grid. More in detail, we follow a standard discrete global grid system for indexing geographies into a hexagonal grid, called $H3$, initially developed by Uber [28]. We consider a resolution of 7 such that hexagonal units have an area of approximately $5.2km^2$. In this way, we balance granularity and sparsity, ensuring a reasonable combination of both. Indeed, at resolution level 6 the average area covered is approximately 36.1 km^2, which would result in an aggregation that is too coarse for our purposes. On the other hand, at resolution level 8, the average area is only 0.7 km^2, making the granularity too fine compared to the Relative Wealth Index (RWI) resolution. This would make it impossible to unambiguously associate an RWI tile with a specific hexagon. Figure 6A illustrates the resulting hexagonal grid for Manaus.

4.4 Voronoi tessellation
We collect the location data of educational facilities within the six cities under investigation from OpenStreetMap [23]. We include all entities categorized with the tag "amenity=school". We note that the tags 'college' and 'university' are excluded from our analysis. Indeed, our primary focus is on lower-grade schools, which have a broader impact on the entire population across all socio demographic groups. In contrast, higher

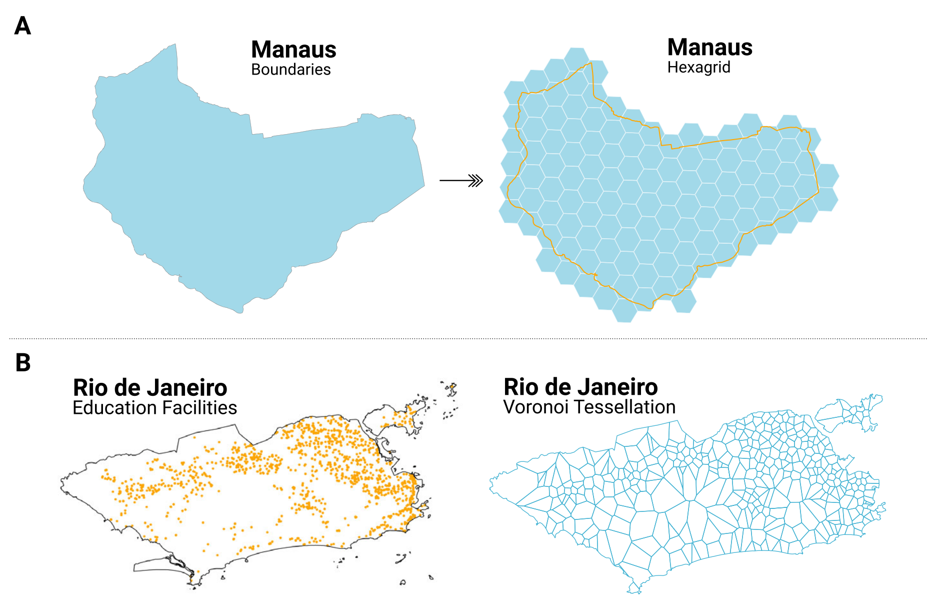

Figure 6 **A)** Boundaries of Manaus and obtained hexagonal grid. **B)** Location of education facilities and obtained Voronoi tessellation for Rio de Janeiro

education often interacts with other socioeconomic dimensions in complex ways. For this reason, we chose to isolate lower-grade schools from higher education institutions to avoid conflating different factors. Each facility is then condensed to its centroid, so that all facilities are represented by a unique set of coordinates. To prevent excessive fragmentation, we merge facilities located within a $1km$ radius of each other. Subsequently, we employ Voronoi tessellation on the resulting centroids. This process generates a Voronoi cell for each centroid, including all points on the plane that are closer to that seed point than to any other. This approach allows us to define the catchment areas of each educational facility. Figure 6B illustrates the location of education facilities and the resulting Voronoi tessellation for Rio de Janeiro.

Abbreviations
Mbps, Megabits per second; RWI, relative wealth index; FTTH, fiber to the home; USD, United States Dollar; CV, coefficient of variation.

Supplementary information
Supplementary information accompanies this paper at https://doi.org/10.1140/epjds/s13688-024-00508-8.

Additional file 1. (PDF 995 kB)

Acknowledgements
All authors thank Ookla, The World Bank and the Development Data Partnership. All authors thank James Carroll, Katherine Macdonald, and Luciano Charlita De Freitas for their support and review.

Author contributions
All authors designed the study, wrote and approved the manuscript. N.G. performed the analyses.

Funding
The research is partially funded by The World Bank Digital Development Partnership. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the

Executive Directors of the World Bank or the governments they represent. N.G. acknowledges support from the Lagrange Project of the Institute for Scientific Interchange Foundation (ISI Foundation) funded by Fondazione Cassa di Risparmio di Torino (Fondazione CRT).

Data Availability
All the dataset used are publicly available and referenced within the text with the exception of the internet Speedtest measurements from Ookla. The Ookla dataset is instead proprietary and cannot be shared publicly.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]ISI Foundation, Turin, Italy. [2]The World Bank Group, Washington, D.C., USA. [3]School of Mathematical Sciences, Queen Mary University of London, London, UK. [4]The Alan Turing Institute, London, UK.

Received: 5 July 2024 Accepted: 4 November 2024 Published online: 08 November 2024

References

1. Mora-Rivera J, García-Mora F (2021) Internet access and poverty reduction: evidence from rural and urban Mexico. Telecommun Policy 45(2):102076
2. Medeiros V, Ribeiro RSM, do Amaral PVM (2021) Infrastructure and household poverty in Brazil: a regional approach using multilevel models. World Dev 137:105118
3. Hernan G, Viecens MF (2017) Connected for development? Theory and evidence about the impact of Internet technologies on poverty alleviation. Dev Policy Rev 35(3):315–336
4. Mariscal EV, Elbittar A, Cave M, Guerrero R, Garcia-Zaballos A, Iglesias E, Webb W (2020) The impact of digital infrastructure on the consequences of COVID-19 and on the mitigation of future effects. SSRN
5. Chiou L, Tucker C (2020) Social distancing, Internet access and inequality. Working Paper 26982, National Bureau of Economic Research
6. Mariscal EV, Elbittar A, Cave M, Guerrero R, Garcia-Zaballos A, Iglesias E, Webb W (2020) The Impact of Digital Infrastructure on the Consequences of COVID-19 and on the Mitigation of Future Effects. IDB Institutions for Development Sector Connectivity, Markets, and Finance Division Discussion Paper No. IDB-DP-827
7. Gozzi N, Comini N, Perra N (2023) The adoption of non-pharmaceutical interventions and the role of digital infrastructure during the COVID-19 pandemic in Colombia, Ecuador, and El Salvador. EPJ Data Sci 12(1):18.
8. Torres A, Domańska-Glonek E, Dzikowski W, Korulczyk J, Torres K (2020) Transition to on-line is possible: solution for simulation-based teaching during pandemic. Med Educ
9. Bauer JM, Hampton KN, Fernandez L, Robertson CT (2020) Overcoming Michigan's homework gap: the role of broadband Internet connectivity for student success and career outlooks. IRPN: Innovation & Information Management (Topic)
10. Taddei M, Bulgheroni S (2020) Facing the real time challenges of the covid-19 emergency for child neuropsychology service in Milan. Res Dev Disabil 107:103786
11. Taylor J, Taylor R (2021) Decreasing work-related movement during a pandemic. Location analytics and the implications of the digital divide. Int J Dev Iss 20:293–308
12. Soomro KA, Kale U, Curtis R, Akcaoglu M, Bernstein M (2020) Digital divide among higher education faculty. Int J Educ Technol Higher Educ 17(1):21
13. Azubuike OB, Adegboye O, Quadri H (2021) Who gets to learn in a pandemic? Exploring the digital divide in remote learning during the covid-19 pandemic in Nigeria. Int J Educ Res Open 2:100022
14. Eruchalu CN, Pichardo MS, Bharadwaj M, Rodriguez CB, Rodriguez JA, Bergmark RW, Bates DW, Ortega G (2021) The expanding digital divide: digital health access inequities during the COVID-19 pandemic in New York city. J Urban Health 98(2):183–186
15. Watts G (2020) COVID-19 and the digital divide in the UK. Lancet Digit Health 2(8):e395–e396
16. GINI Index, The World Bank (2024) https://data.worldbank.org/indicator/SI.POV.GINI. Accessed: 2024-04-23
17. Anatel (2024) https://informacoes.anatel.gov.br/paineis/acessos. Accessed: 2024-04-23
18. Feamster N, Livingood J (2020) Measuring Internet speed: current challenges and future recommendations. Commun ACM 63(12):72–80
19. Vakataki'Ofa S, Aparicio CB (2021) Visualizing broadband speeds in Asia and the Pacific
20. Ford GS (2021) Form 477, Speed-Tests, and the American Broadband User's Experience. Speed-Tests, and the American Broadband User's Experience (March 31, 2021)
21. Chi G, Fang H, Chatterjee S, Blumenstock JE (2022) Microestimates of wealth for all low-and middle-income countries. Proc Natl Acad Sci 119(3):e2113658119
22. Zhou X, Lin H (2008) Moran's I. In: Encyclopedia of GIS. Springer, Boston, pp 725–725
23. Humanitarian OpenStreetMap - Education Facilities. (2024) https://data.humdata.org/dataset/?dataseries_name=HOTOSM+-+Education+Facilities. Accessed: 2024-03-28

24. FCC increases broadband speed benchmark (2024) https://docs.fcc.gov/public/attachments/DOC-401205A1.pdf. Accessed: 2024-04-23
25. Google LLC (2020) Google COVID-19 Community Mobility Reports. https://www.google.com/covid19/mobility/. Accessed: 2021-08-01
26. Conexis (2024) Statistics. https://conexis.org.br/numeros/estatisticas/. Accessed: 2024-04-23
27. Ookla for Good (2024) https://www.ookla.com/ookla-for-good. Accessed: 2024-05-23
28. H3 discrete global grid system (2024) https://h3geo.org/. Accessed: 2024-04-23

Publisher's Note