Accessibility to Healthcare in South Africa

Countrywide Infrastructure Assessment

In Collaboration with the World Bank

Team Members: Andri Vidarsson, Yitong Lu, Charles Shaviro

Mentors: Mersedeh Tariverdi, Miguel Nunez-del-Prado

Introduction

Quality infrastructure is essential for effective healthcare services, especially during major disruptions caused by natural disasters or pandemics. Resilient systems for water, electricity, and transport are critical to ensure reliable access to healthcare, and a large body of research makes clear that access to quality healthcare is a prerequisite for sustainable growth in developing countries [Rentschler et al., 2021]. The onset of climate change and the greater frequency and intensity of natural disasters humanity will need to weather in the 21^{st} century make it imperative that policymakers invest in the essential infrastructure that enables access to healthcare, and that researchers conduct sophisticated analyses to guide these investments to be as effective as possible.

In this project, we focus on transportation infrastructure, specifically the road network and healthcare infrastructure of South Africa, and build a framework to assess accessibility to healthcare facilities on a national scale, as well as to evaluate the criticality of individual road segments. Our primary analysis emphasizes normal conditions; however, we are in the early stages of extending this work to consider accessibility under disrupted scenarios caused by natural disasters. Moreover, without loss of generality, the framework and methods we develop can be readily applied to other essential services beyond healthcare and adapted to different countries or regions with minimal preprocessing, provided that the necessary data is available.

Related Work

Our work builds on prior research conducted by the World Bank, in particular the innovative *Measuring Accessibility to Public Services and Infrastructure Criticality for Disasters Risk Management* [Tariverdi et al., 2023]. In the paper, Tariverdi et al. introduce a new user-centric measure for estimating infrastructure criticality and urban accessibility to critical public services, and we build off their approach here. But previous studies, including the World Bank's work in similar projects, primarily focus on individual cities or more contained regions. In contrast, our framework is designed to assess accessibility and criticality on a national scale, modeling and assessing the entire road network under both normal and disrupted conditions. We hope that the computational pipeline we build in this work can be used to inform regional planning, investment prioritization, and myriad weighty policy decisions, ultimately enhancing the effectiveness of investments in infrastructure resilience on a broader scale.

Data

The essential data for our work includes road network data, population density data, and healthcare facility locations. We integrate these datasets into a graph structure that serves as the foundation for our simulations and assessments. The road network, representing roads and intersections, is primarily sourced from publicly available platforms, with *OpenStreetMap* (OSM) [contributors, 2023] being the main resource for constructing the model graph. Population density data was obtained from the United Nations Office for the Coordination of Humanitarian Affairs (OCHA), while data on healthcare facility locations was provided by the World Bank.

Although our framework was designed to be versatile and applicable to any country with sufficient data on road networks, population distribution, and healthcare facilities, we use South Africa as a case study to build out, test, and refine our approach. We demonstrate the model's effectiveness in evaluating infrastructure resilience at a national scale and confirm the framework's adaptability to broader applications in other regions.

Methods

Graph Construction and Data Preparation

The road network data is provided as a *GeoPackage* [Open Geospatial Consortium, 2014], represented as a collection of line strings. We process this data and construct a graph structure in which nodes represent intersections and dead ends and edges represent the road segments that connect intersections.

Healthcare locations are mapped to corresponding nodes on the graph using efficient nearest-neighbor search implemented with k-d trees [Bentley, 1975]. We process population density data using the H3 spatial indexing system [Uber Technologies, 2018] with resolution 5, so as to divide South Africa into 4,809 hexagons capturing the geographic distribution of the population.

Simulation Pipeline

We establish a scalable pipeline that can simulate journeys from populated areas to healthcare facilities. For each population hexagon, n origin points are sampled, where n is proportional to the logarithm of the squared population density. By using this sampling approach, we ensure a balanced representation of trips in both urban and rural areas.

From each origin point, we simulate journeys to the five nearest hospitals, which are identified using k-d trees. The shortest paths, based on travel time, are calculated using Dijkstra's algorithm [Dijkstra, 1959]. This approach enables us to model realistic travel scenarios across the entire country and its 12^{th}-largest-in-the-world road network [Wikipedia contributors, 2023].

The entire pipeline is implemented in PySpark [Zaharia et al., 2016] on a multi-core setup, leveraging implicit parallelization to efficiently process the large-scale data and simulations.

Accessibility and Criticality Analysis

We assess accessibility by computing the average travel time of simulated journeys originating from each population hexagon. This metric provides a spatially granular measure of how easily healthcare facilities can be accessed across the country.

We evaluate the criticality of the road network by identifying road segments that most frequently occur on the simulated paths. A segment is defined as a connected sequence of 2 to 10 nodes, and if a critical path is entirely contained within another, only the larger segment is retained (to avoid redundancy). To ensure fairness across regions, we establish a minimum support threshold of 0.25% for all trips within each province, preventing major cities from disproportionately dominating the analysis. Additionally, road segments in the immediate vicinity of hospitals can be excluded, as their criticality is trivial given their direct physical proximity to healthcare facilities.

We also implement a second method for computing road segment criticality that involves summing the populations of hexagons that rely on a given road segment as part of their shortest travel-time path. Let C_r denote the criticality of road segment r, and p_{h_i} denote the population of hexagons h_i. The criticality is calculated as:

$$C_r = \sum_{h_i | h_i \cap r \neq \emptyset} h(p_{h_i})$$

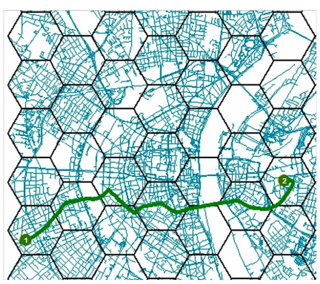
where $h(\cdot)$ is a weighting function applied to the population, allowing for flexible prioritization based on different considerations. For example, $h(p_{h_i}) = p_{h_i}$ means that we treat each person equally (similar to $\|\cdot\|_1$), whereas $h(p_{h_i}) = 1$ means we treat each hexagon as equally important (similar to $\|\cdot\|_\infty$). This flexibility enables exploration of different population weighting schemes that can better balance the significance of population density in the criticality assessment.

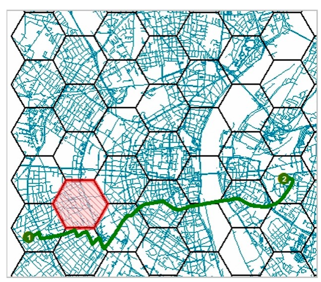
Regional Criticality

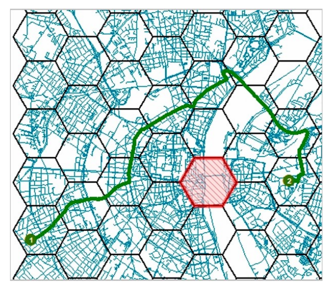
In addition to computing the criticality of individual road segments, we can also evaluate it on a regional level. More exactly, we can assess how the blockage of a region (for example, one of our hexagons) caused by an emergency or natural disaster might affect the accessibility of healthcare facilities for people living in the vicinity of the blockage.

In our computation, we first derive the blockage-influenced subgraph, G_{h^-}, by removing the road network contained within a blocked hexagon h from the original comprehensive graph G. We then recalculate the shortest path journeys of populations in affected hexagons to nearby healthcare facilities and sum up the increase in the shortest travel times. Let $T_{h_i}^{G_{h^-}}$ denote the new travel time for people in hexagon h_i in the new subgraph G_{h^-}, and $T_{h_i}^{G}$ denote the original travel time. The regional criticality, C_h, is computed as:

$$C_h = \sum_{h_i \in h^-} (T_{h_i}^{G_{h^-}} - T_{h_i}^{G}) * h(p_{h_i})$$







In the three figures above, the green line indicates the shortest path to a healthcare location, the grey-blue lines represent the road network, and the red tiles are blocked. The left figure represents the route without any blockage introduced, the middle figure shows a blockage with lower impact, and the right figure shows a blockage with high impact due to the removal of a river-spanning bridge from the road network [Rohr et al., 2020].

Results

We present a variety of maps that describe South Africa's population density, road network, and its landscape of healthcare accessibility and road network criticality.

Accessibility

The maps displayed in Figure 1 and Figure 2 show a clear relationship between population density and average travel time to healthcare facilities. Large cities like Pretoria, Johannesburg, and Cape Town have shorter travel times due to better infrastructure and many more hospitals, needed to serve their large populations. Conversely, rural areas, particularly in the Northern and Eastern Cape, face significantly longer travel times, highlighting disparities in healthcare access. These dynamics emphasize the need for targeted infrastructure improvements to reduce travel times in under-served regions, as well as the multi-facted considerations and complex dynamics involved in investment decisions over such a large country or region.

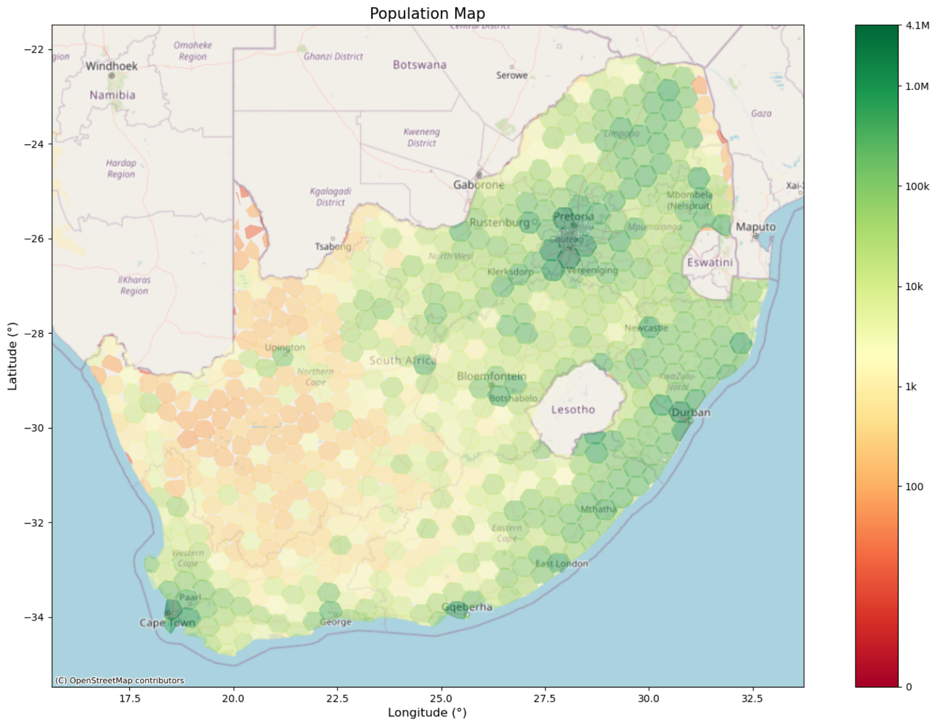


Figure 1: Population distribution in South Africa.

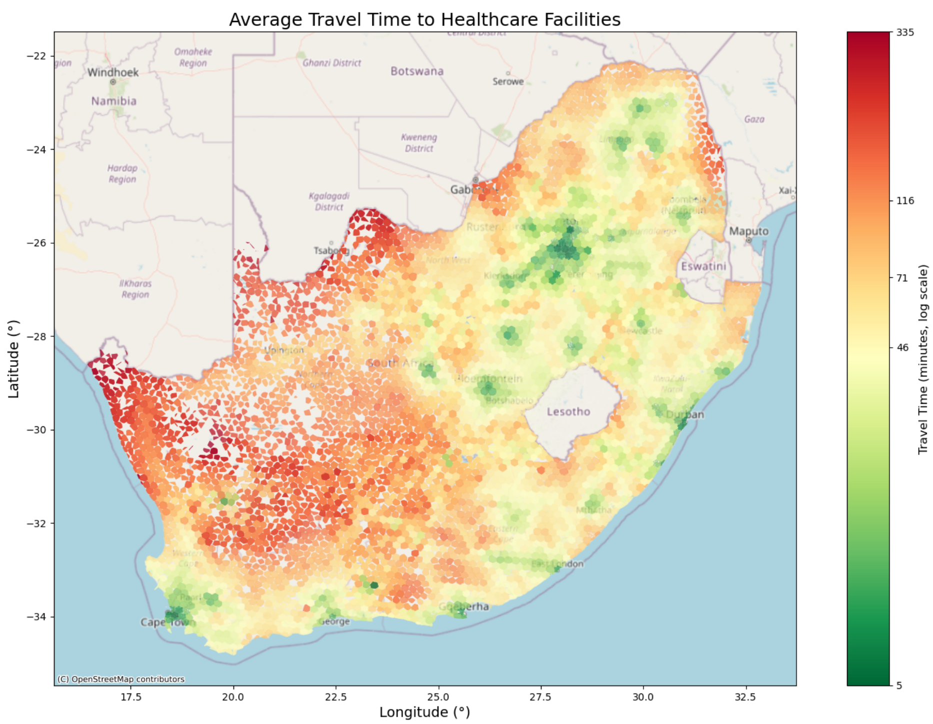


Figure 2: Average travel time to healthcare facilities in South Africa.

Road Criticality

Figure 3 illustrates the critical road segments in the Free State province. Critical road segments are identified based on their frequent occurrence in simulated journeys to healthcare facilities. Critical paths are largely concentrated around major cities such as Bloemfontein, reflecting the central role of these roadways in connecting rural areas to urban healthcare centers.

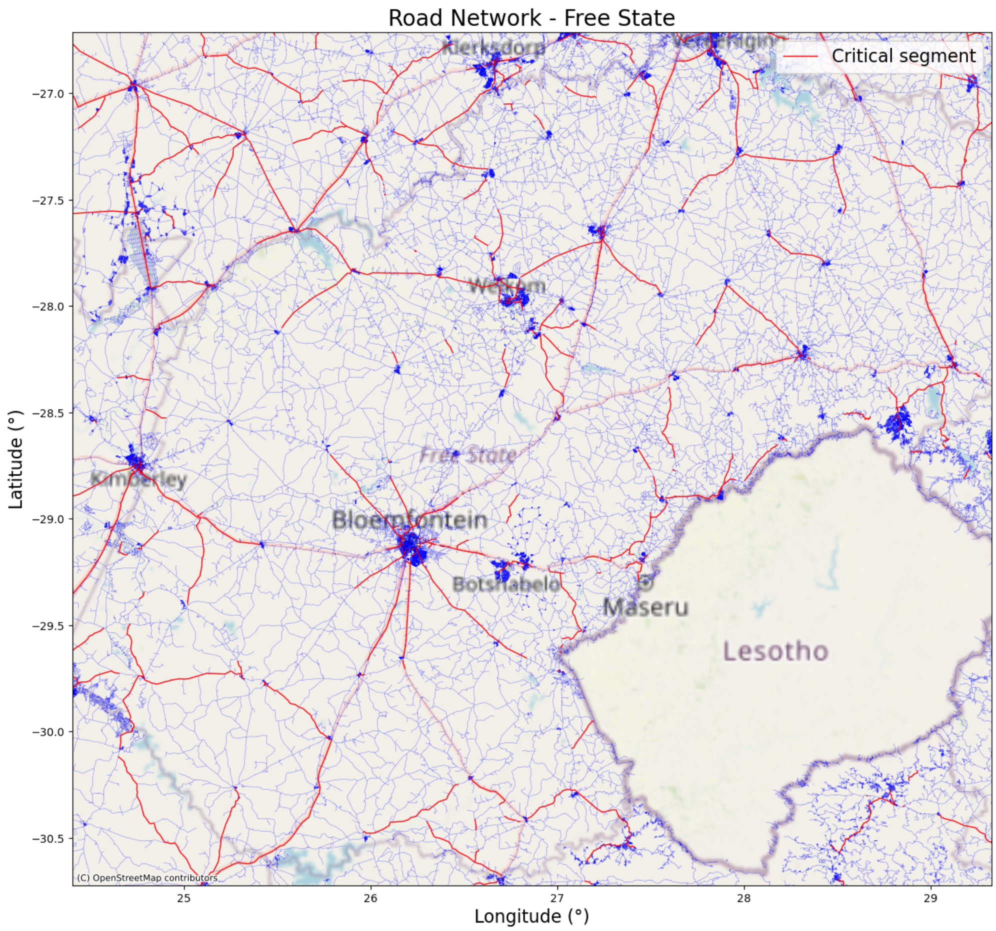


Figure 3: Critical road segments in the Free State province, South Africa. The critical segments, shown in red, highlight the most frequently used road segments for simulated journeys to healthcare facilities.

Region-Based Criticality

Figure 4 illustrates the criticality calculated via our regional/hexagon-based criticality assessment in the Cape Town area. Green regions represent areas with zero criticality that would not impact healthcare access if the road segments contained within them were blocked. On the other hand, darker red regions indicate the areas with the highest criticality, which are typically located near major healthcare facilities and along critical highway segments with significant importance for accessibility.

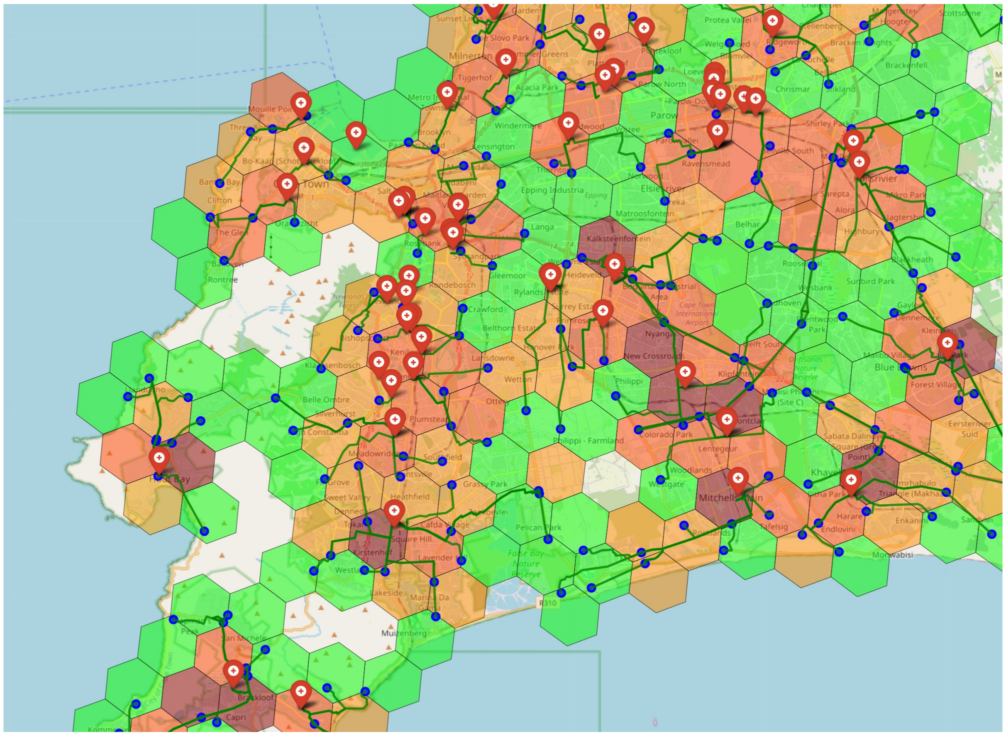


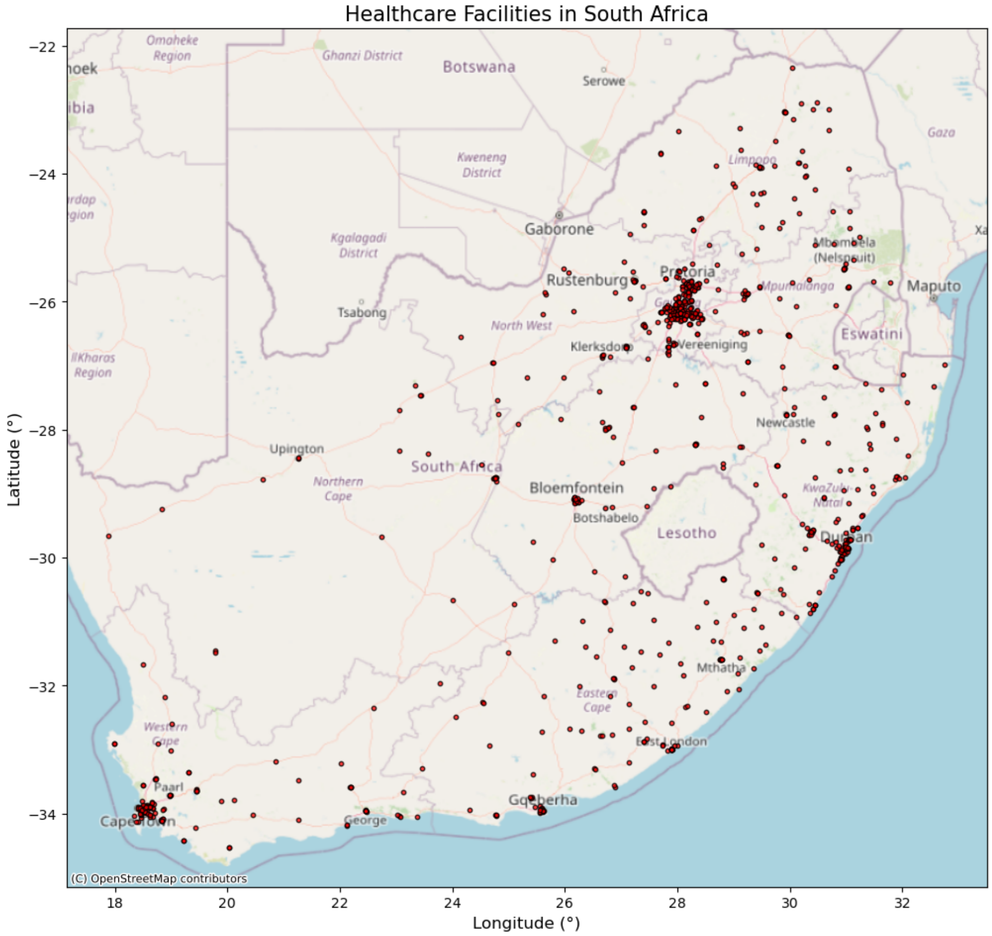
Figure 4: Criticality of each region in the area of city of Cape Town.

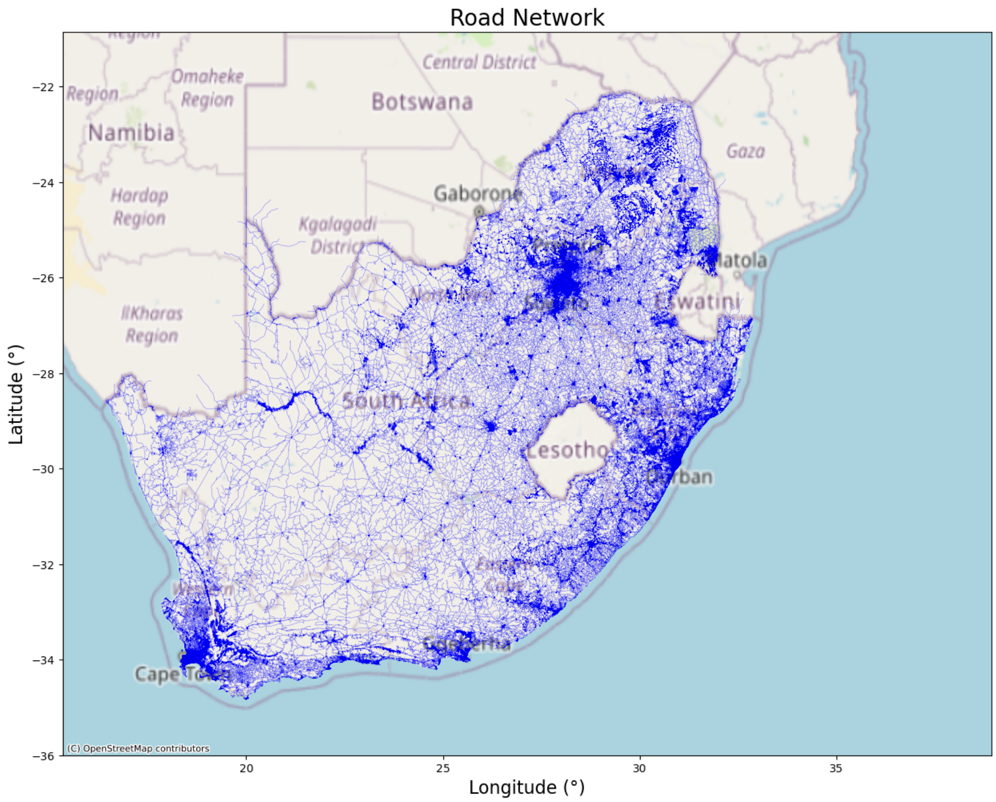
Next Steps

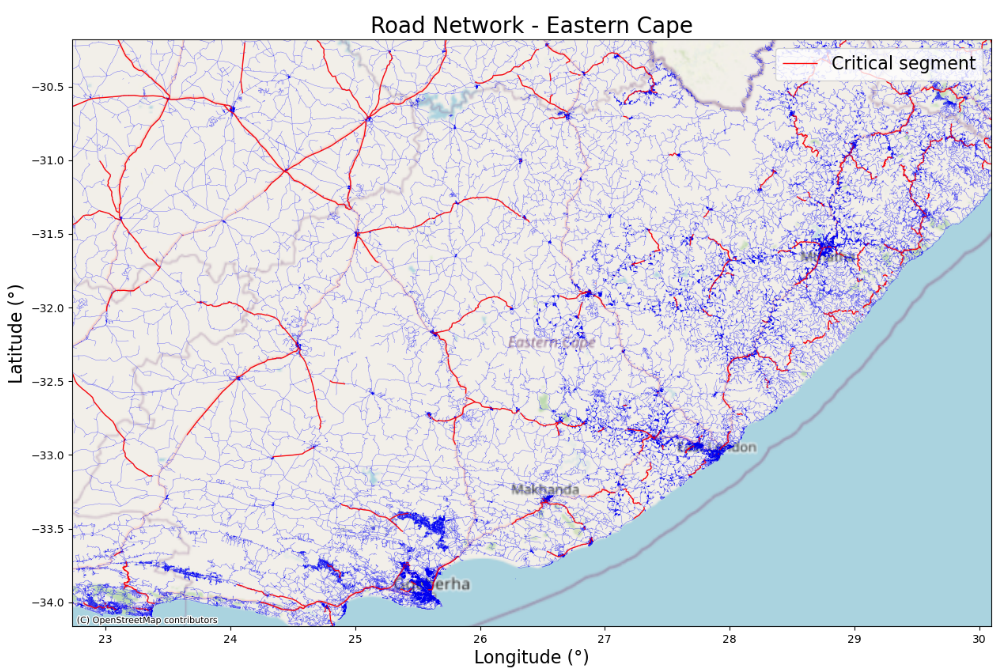
There are clear pathways over which to expand this research. One direction lies in the opportunity to design a more detailed and data-driven analysis of resilience on the national and province-by-province level. This can be accomplished via the incorporation of historical hazard map data or state-of-the-art hazard risk assessments that can realistically simulate network perturbation in disaster scenarios and allow researchers to meaningfully assess resilience. Additionally, further functionality can be built out to increase the generalizability of our computational pipeline to other countries and regions.

Appendix

We briefly include additional appendix figures generated in our analysis.




Road Network
Latitude (°)
Longitude (°)
−22
−24
−26
−28
−30
−32
−34
−36
20
25
30
35
Namibia
Botswana
Gaborone
South Africa
Lesotho
Cape Town
Omaheke Region
Hardap Region
Kgalagadi District
Central District
Gaza
IIKharas Region
(C) OpenStreetMap contributors


Road Network - Eastern Cape
Critical segment
Latitude (°)
Longitude (°)
−30.5
−31.0
−31.5
−32.0
−32.5
−33.0
−33.5
−34.0
23
24
25
26
27
28
29
30
Eastern Cape
Makhanda
(C) OpenStreetMap contributors

References

[Bentley, 1975] Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517.

[contributors, 2023] contributors, O. (2023). Openstreetmap api. Accessed: 2023-09-18.

[Dijkstra, 1959] Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271.

[Open Geospatial Consortium, 2014] Open Geospatial Consortium (2014). Ogc geopackage encoding standard. Technical Report 12-128r13, Open Geospatial Consortium. Version 1.0.

[Rentschler et al., 2021] Rentschler, J., Tariverdi, M., Desjonqueres, C., and Mercadante, J. (2021). Frontline: Preparing healthcare systems for shocks from disasters to pandemics.

[Rohr et al., 2020] Rohr, A., Priesmeier, P., Tzavella, K., and Fekete, A. (2020). System criticality of road network areas for emergency management services—spatial assessment using a tessellation approach. *Infrastructures*, 5(11):99.

[Tariverdi et al., 2023] Tariverdi, M., del Prado, M. N., Leonova, N., and Rentschler, J. (2023). Measuring accessibility to public services and infrastructure criticality for disasters risk management. *Scientific Reports*, 13(1):1569.

[Uber Technologies, 2018] Uber Technologies, I. (2018). H3: A hexagonal hierarchical geospatial indexing system. https://h3geo.org.

[Wikipedia contributors, 2023] Wikipedia contributors (2023). List of countries by road network size. Accessed: 2023-12-04.

[Zaharia et al., 2016] Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., and Stoica, I. (2016). Apache spark: A unified engine for big data processing. In *Communications of the ACM*, volume 59, pages 56–65.