EPA1352. Advanced Simulation

Assignment 3. Building Components for Data-Driven Simulation

Group 17

Ivan Temme (4955196)

Hidde Scheuer (4607325)

Philip Mueller (5809703)

Madalin Simion (5838363)

Nachiket Kondhalkar (5833884)

1. Introduction

This report aims to expand on the analysis conducted in assignment two on the impact of bridge unavailability and maintenance on traffic travelling. In this assignment, the previous mesa model is updated and overlayed with Network X, and the analysis is done on the N1 and N2 highways and their side N roads. To reduce the influence of modeller bias (Keller & Hu 2019), a data-driven modelling approach was employed in this study.

2. Data preparation & Model Design

Most of the data preparation and cleaning process remained the same as the previous assignments, with the exception of some additions that will be the addressed and justified in this report.

2.1. Dataframe size and merging

Two new additions were made to the data cleaning process. Firstly, the cleaning process now includes a larger number of roads, including the N1, N2 with all of their crossroads and sideroads. Additionally, the bridges on these roads are also considered. To accomplish this, it was necessary to filter all roads that cross N1 or N2 and verify that their last LRP (LRPE) was more than 25km away from their start (LRPS). Secondly, intersections between roads needed to be explicitly identified. While most intersections were properly labelled in the original data, some were not In these cases, overlapping points of different roads were either found or approximated, and defined as the intersection by selecting the closest LRP point of the road that should intersect. For more information on the procedure please consult the *data_preprocessing.ipynb file*.

2.2. SourceSink and NetworkX

To enable vehicles to move over the road network, the Source and Sink components used during model generation should be replaced with SourceSink. When a vehicle is generated at a source, it will be directed to a random sink location in the network. To facilitate this, a NetworkX model of the road network is generated. In this model, each bridge is presented as a node, each road as an edge and can return the shortest path between two nodes.

To optimize the process of determining the shortest path between an origin and a destination for a vehicle using the NetworkX model, the discovered paths are saved into dictionary path_ids_dict_complex. When a new vehicle needs to travel the same path, it first checks this dictionary if path already exists. If so, the vehicle can use it directly. Otherwise, the NetworkX model will be used to compute the shortest path. This approach reduces the need for repetitive calculations and speeds up the process of determining the shortest path for a vehicle.

2.3. Batch Run

The *batchrunner.py* file was added to the project, which utilizes Mesa's batchrunner (https://mesa.readthedocs.io/en/stable/apis/batchrunner.html) to run all scenarios sequentially. Each scenario runs 10 iterations for a period of 5 days. The probability of bridges breaking according to category is passed to the batchrunner. The output is saved for every scenario and combined in the file, called *scenarios.csv.*

For this process we have also utilised the functionality of tensorflow to optimise the functionality of the code in order to process the calculations on the GPU. This functionality has also been used to optimise calculations for the analysis.

3. Results

The *batchrunner.py* file contains scenarios 0 to 4 with varying probabilities of bridges breaking down for different condition-categories (A, B, C, D). After analysing the simulation outcomes of these scenarios, important findings have been summarized in several plots. Driving time increases with the higher likelihood of bridge breakdowns and higher centrality of collapsed bridges.

In the following, scenario 0 will refer to the baseline scenario, in which no bridge is being modelled as collapsed. Table 1 shows the probability of bridges collapsing. Probabilities are skewed towards Cat D bridges. Delay times are modelled based on probability functions too. Delay times are increasing with increasing bridge length.

Scenario	Cat A %	Cat B %	Cat C %	Cat D %
0	0	0	0	0
1	0	0	0	5
2	0	0	5	10
3	0	5	10	20
4	5	10	20	40

Table 1: a of bridges breaking down, depending on their category and the modelled scenario.

3.1. Driving and Delay Times

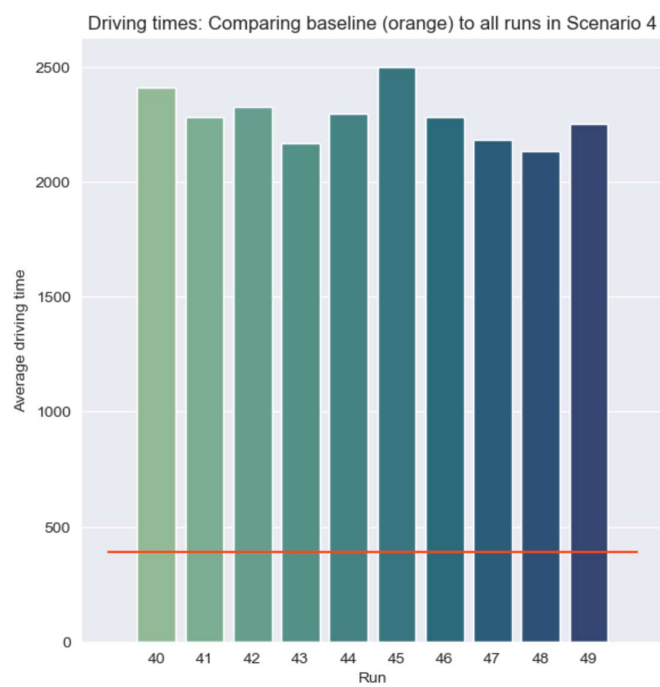

Figure 1: Comparing driving times between baseline and scenario 4. Blue bars depict measured driving time for each run. The first digit in runs refers to the scenario, the second digit refers to the run inside that scenario. The red line depicts driving time in baseline.

Figure 1 shows that more collapsed bridges add to driving time. Driving time is equal across all runs in the baseline scenario since no bridge is collapsed. Since bridge collapses are based on probabilities. Driving times are varying between runs. The spread of driving times is increasing with each scenario. Average driving time increases roughly 6-fold across all scenarios.

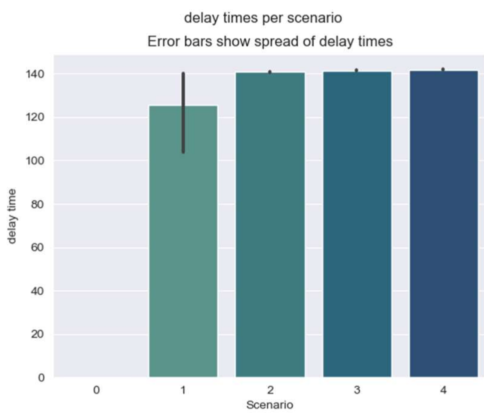

Figure 2: Delay times in hours. Coloured bars depict average delay time over all runs, the error bars depict range between maximum and minimum delay time over all runs.

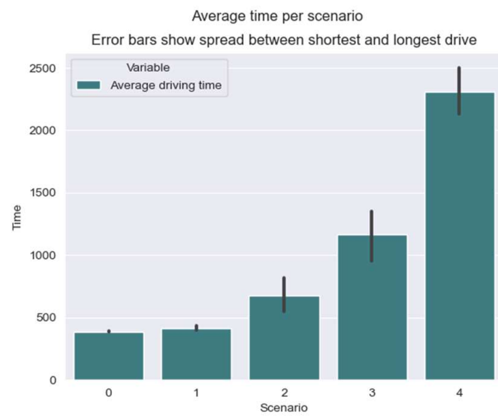

Figure 3: Driving times in hours. Coloured bars depict Average driving time over all runs, the error bars depict range between maximum and minimum driving time over all runs.

s can be seen in Figure 2, the baseline scenario has no delay time. Delays between scenarios 1, 2, 3, and 4 vary little. As can be seen in Figure 2, scenario 1 shows the largest range of delay times, roughly between 100 and 140 hours. Driving times are increasing exponentially with each new scenario. AS can be seen in Figure 3, the baseline driving time is at about 380 hours. Interestingly, average driving time does not increase massively between scenario 0 and 1. Furthermore, although delay time stays relatively stable between scenarios 2, 3, and 4, that is not the case for driving times. Increase delay time still seems to lead to increased driving time. The relationship seems to be not proportional, however. The result suggest that Total driving time is influenced by an underlying factor. We suspect centrality to play a role. Centrally located bridges, if collapsed, should have a higher impact on driving times than bridges of less network centrality. In the following, we will investigate positioning of bridges and intersections.

3.2. Betweenness centrality

Betweenness centrality is a widely used measure that captures an actors role in allowing information to pass from one part of the network to the other (Golbeck, 2015). In our case, the actor we are interested in are the intersections and bridges, which act as links between the different roads, and subsequently the traffic that passes through them.

3.2.1. Intersections

Take the intersection with id 2756. The intersection is connecting the roads N1 and N102. This intersection is very important for the flow of traffic through the network. The betweenness in this case measures the percentage of shortest paths that must go through this specific node (Golbeck, 2015). If an intersection of high betweenness centrality were to be blocked, it is likely the case that the average drive time would increase significantly, since the shortest path is not available anymore. Whether there are further intersections between these two roads also has an impact on overall driving time. Intersection 6956 is connecting only regional roads but still shows high betweenness centrality. This intersection most likely does not contribute to overall driving time as much as the other listed intersections. It can be assumed that most traffic is router through the large national roads.

road 1	road 2	id	betweenness_centrality
N1	N102	2756	0,513072066
N1	N104	6956	0,496662342
Z1031	Z1034	6956	0,496662342
N2	R211	3141	0,492144605
R211	R310	3141	0,492144605

Table 2: 5 Intersections with highest betweenness centrality in the network.

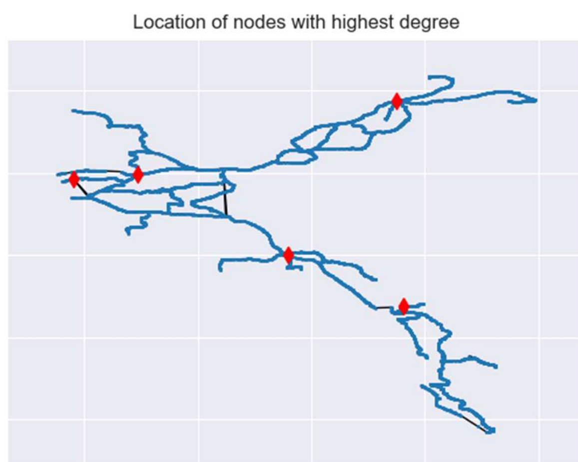

Figure 4: Location of the nodes with the highest degree

3.2.2. Bridges

Amtoli Slab Culvert Bridge has the highest betweenness centrality. Almost 48.6% of the shortest paths go through this bridge. Thankfully, the condition of the bridge is good (A), which does not endanger the flow of traffic. Mostapur Bridge, however, is in a condition of major elemental damage (C), while having almost just as high of a centrality, at almost 48.5%. This bridge is of more interest to be repaired or strengthened. Its likelihood of collapse is high and its contribution to traffic flow – and thus potential network delay time – is also high. All 5 bridges of highest betweenness centrality are on N1, which is the economically most important road in Bangladesh.

road	id	condition	bridge_name	betweenness_centrality
N1	1213	A	AMTOLI SLAB CULVERT	0,485825884
N1	1217	A	DHANPUR RCC GIDER BRIDGE	0,485593787
N1	1229	B	PADUAR BAZER RCCGIDER	0,484886256
N1	1231	C	Mostapur Bridge	0,484766694
N1	1238	A	RAZAPARA RCC GIDER BRIDGE	0,484344541

Table 3: 5 bridges with highest betweenness centrality in the network.

3.3. Closeness centrality

Closeness centrality is a measure of how central a node (in this case, a bridge) is within a network. The node with highest closeness centrality is the node with most direct connections to other nodes in the network (Golbeck, 2013).

3.3.1. Intersections

All intersections with highest closeness centrality are connecting roads to N1. This is anticipated since N1 is a particularly central and economically important road in Bangladesh, indicating that nodes on N1 have generally high closeness centrality. Table 4 indicates that N1 is a well-connected road. Traffic over N1 seems to be easily rerouted. The high number of intersections on N1 is important, since the number of bridges – and thus potential delay points – is also high on N1. Comparing tables 2, and 4, it can be seen that the listed intersections do not overlap. Further research is needed to investigate what measure of centrality is adding most insights for the client, e.g., what intersection is most important to reroute traffic and minimize total driving time.

road 1	road 2	id	closeness_centrality
N1	Z1042	1158	0,001794412
N1	Z1402	1133	0,001789187
N1	N102	1210	0,001789173
N1	N102	2756	0,00178901
N1	R301	6015	0,001753364

Table 4: 5 intersections with highest closeness centrality in the network.

3.3.2. Bridges

For Bridges, it can be seen that the one which have the highest closeness centrality are situated on the N1 road, which is to be expected. It would be uncommon if a bridge on a side road would be the most central in the network. From the top five bridges sorted by closeness, Aital Banga is of most interest, given its advanced state of deterioration and the high closeness centrality score, relative to the other bridges. Like with betweenness centrality all 5 most central bridges are located on N1. Note that there is no overlap between the bridges listed in tables 3, and 5. Further research is needed to assess what criticality measure should be consulted in this use case, e.g., what measure shows highest correlation to the overall measured driving time. Jafino et.al. (2020) provide different approaches to selecting measures of centrality.

road	id	condition	bridge_name	closeness_centrality
N1	1156	B	ILLIOT BAZER	0,001793967
N1	1163	B	KHAT GHOR BOX CULVERT	0,001793847
N1	1167	B	BAKRA BOX CULVERT	0,001793396
N1	1152	C	AITAL BANGA BOX CULVERT	0,001793077
N1	1173	B	MATIA BAZER	0,001792719

Table 5: 5 bridges with highest closeness centrality in the network.

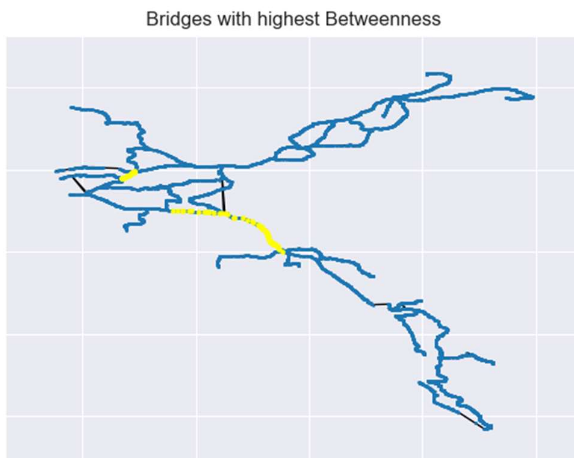

Figure 5: bridges with the Highest betweenness

3.4. Bonus

Upon analysing the shapefile for roads, we created a dataframe that comprised individual coordinates for every point in the shapefile. Processing the entire shapefile was not practical and a filter was created using the max and min values of latitude and long for the model data. This Dataframe was then used to break down individual coordinate locations into a pandas dataframe.

This Dataframe was then used to create a map to observe the coincidence of points on the shapefile with the points identified as intersections. On observing this map, we could observe that the location of all intersections was located in clusters of points, as described in the shapefile. This means that they were geographically close to the shapefile features but there was no quantification for the same.

Our plan was to measure the distance between the intersection and the shapefile points. However, upon observation it was decided that this method was highly inefficient, and we chose not to pursue this process further in the interest of time.

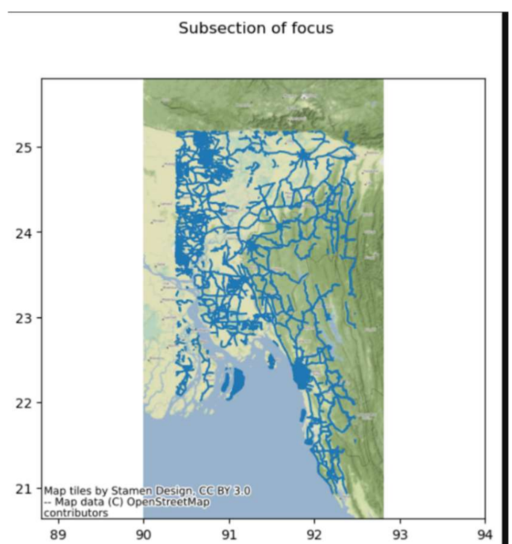

Figure 6: Subsection of focus

We decided to use the shapefile directly and use the geopandas.overlay function along with the buffer function to observe the intersection of these point. Here, we observed that for a buffer of 0.05 (5.56 km approx.) there are only 44 intersections that are close to geographical features out of 87 intersections identified in the model. This means that either the method identified to clean the data or the process of creating intersections may have certain flaws in it. Further research should be directed towards checking the correctness of identified intersections and finding the optimum buffer value.

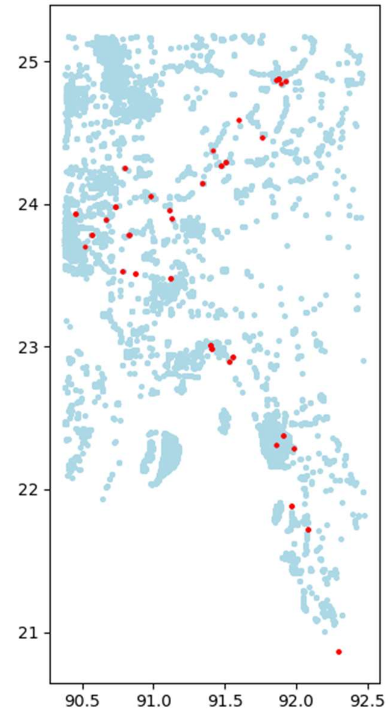

Figure 7: Occurrence of intersection and feature

What bridges should be prioritized by the client depends on several factors. Bridges of lower quality – and thus higher probability of collapse – should be considered first. Bridges of high centrality should also be considered first, in order to minimize the rerouting distance and resulting added driving time. further research is needed in order to find the optimum measure of centrality for this simulation. A good centrality measure shows high correlation with added driving time.

The goal here is to decrease overall driving time in the simulation. When extrapolating model results, ethical questions may arise. For instance, instead of decreasing overall driving time, certain bridges that are critical to connecting remote areas to the highway network might have to be prioritized. A relatively high overall driving time in the simulation might thus be acceptable. Jafino et.al. (2020) provide a section on ethical and spatial considerations for choosing centrality measures.

4. Limitations

There are multiple avenues for improving the model. Firstly, the process of creating the intersections is computationally intensive. One way to optimise this could be creating flags where the identification of sideroads takes place.

Another limitation of the analysis is that it excluded bridges from the dataset due to missing LRPs. Moreover, the approach of eliminating 'Left' side bridges may result in the exclusion of potential bridges that do not have a corresponding counterpart on the ‘Right’ side in the dataset in the future. This method needs to be monitored when scaling the model in the future.

The small sample size of only 10 iterations per scenario may result in abrupt and jarring shifts caused due to statistical variance of sample size. If the number of iterations had been increased, the transitions between scenarios would have been smoother.

Computational complexity is also a big hurdle to processing this assignment, which necessitated the use of GPUs to obtain meaningful results. Moreover, we are uncertain as to why network generated circular edges at certain points between nodes. The fact that they seem to apparently coincide with some intersections was not coincidence. However, we were unable to resolve this issue.

References

- Rhino Car Hire. (2023). Drive Left or Right - Which Countries Drive On the Left or Right. https://www.rhinocarhire.com/Drive-Smart-Blog/Drive-Left-or-Right.aspx
- Keller, N., & Hu, X. (2019). Towards data-driven simulation modelling for mobile agent-based systems. *ACM Transactions on Modeling and Computer Simulation* (TOMACS), 29(1), 1-26
- Jafino B.A., Kwakkel, J., Verbraek, A. (2020). Transport network criticality metrics: A comparative analysis and a guideline for selection (Transport Reviews), 40:2, 241-264, DOI: 10.1080/01441647.2019.1703843