TEXAS A&M UNIVERSITY

STAT 636 Final Report

Lauren Wilcox, Ayaan Omair

December 11, 2025

1 Introduction

Texas has some of the most congested and weather–vulnerable roads in the United States. City planners need to know which links in the road network matter the most: which roads keep the system moving, which closures hurt the most, and which roads should be repaired first after a disruption. In this project we use tools from network science to study two sub–networks of the Texas road system: the suburban area around Leander & Cedar Park, and the road network in Downtown/Central Austin.

Our goals are to: (1) identify critical roads and intersections, (2) measure how robust each network is to different closure scenarios, and (3) propose a greedy order in which to repair closed roads to recover efficiency as quickly as possible.

1.1 Dataset

We use an OpenStreetMap Texas extract from Geofabrik [2], specifically the `gis_osm_roads_free_1.shp` roads shapefile. Each row in the shapefile is a road segment with a geometry (polyline) and tags such as road class (`fclass`), name, and maximum speed limit (km/h).

We convert this shapefile to a directed road network graph for analysis using a Python script (originally developed with the help of ChatGPT). In this graph:

- Line endpoints become nodes (intersections or end points).
- Line segments become directed edges (roads), with attributes for length, road class, and name.

The full Texas network has roughly 5.8 million intersections and 7.0 million road segments, so we focus on two smaller regions

- Leander & Cedar Park: about 8,500 nodes and 14,000 edges.
- Downtown/Central Austin: about 4,300 nodes and 7,800 edges.

2 Analysis Questions

Our analysis is guided by three questions:

1. Critical Roads

Which roads are most important for maintaining travel efficiency, and how much does efficiency drop when one segment is removed?

2. Network Robustness

How rapidly does network efficiency degrade under failures, and how does resilience differ between random and targeted closures?

3. Order of Road Repair

After a major disruption, in what order should roads be repaired to restore efficiency as quickly as possible?

3 Methods

All three analysis questions depend on comparing how the network performs under different road availability scenarios. A single metric is required to quantify these changes in system-wide performance. Global efficiency provides this measure by capturing how easily travel can occur between all pairs of nodes in the network. Each method in the following sections is built around evaluating how global efficiency responds to road removals or restorations.

3.1 Global Efficiency

Global efficiency is the metric used to evaluate how well a road network supports movement between all pairs of locations. Introduced by Latora and Marchiori [1], it summarizes how easily travel occurs across the entire system when travel time is used as the cost.

For a graph G with N nodes, global efficiency is defined as

$$E(G) = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d(i,j)},$$

where $d(i,j)$ is the shortest-path travel time between nodes i and j. Shorter travel times result in larger contributions to the sum, so the metric increases when the network is easier to traverse and decreases when routes lengthen.

The value of $E(G)$ is most useful in comparison, such as evaluating the network before and after a road closure. A decrease indicates longer system-wide travel times, while an increase reflects better accessibility. Classical centrality measures identify important intersections but do not quantify how much travel worsens when a segment fails. Global efficiency captures system-wide performance and responds directly to disruptions, making it suitable for identifying critical roads, assessing robustness, and determining effective repair strategies.

Example

Consider three intersections A, B, and C with travel times $A \leftrightarrow B = 4$, $B \leftrightarrow C = 4$, and $A \leftrightarrow C = 8$ minutes. The global efficiency is computed from all reciprocal shortest-path times.

If the road between A and B closes, the shortest path becomes $A \to C \to B = 12$ minutes, reducing the efficiency by 26.6%.

$$E(G) = \frac{1}{3(2)} \left(\frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} \right) = \frac{5}{24} \approx 0.2083,$$

$$E(G_{\text{closed}}) = \frac{1}{6} \left(\frac{1}{12} + \frac{1}{4} + \frac{1}{8} + \frac{1}{12} + \frac{1}{4} + \frac{1}{8} \right) = \frac{11}{72} \approx 0.1527,$$

$$\frac{E(G) - E(G_{\text{closed}})}{E(G)} = 0.266.$$

3.2 Critical Roads

The analysis of critical roads addresses the question of which road segments exert the largest influence on overall travel efficiency. Identifying these segments begins with computing the global efficiency of the full network to establish a baseline measure of travel performance. Each road segment is then removed individually, and the global efficiency is recomputed on the modified network. The change in efficiency quantifies that segment's contribution to system-wide connectivity, with larger drops indicating roads that play a more central role in maintaining short travel paths. Evaluating this efficiency change for all segments allows the network to be ranked from most to least critical.

3.3 Network Robustness

Assessing the robustness of the network involves measuring how travel efficiency deteriorates as roads are progressively removed. The analysis begins by computing the global efficiency of the intact network to establish a baseline. Road segments are then removed one at a time, and the efficiency is recalculated after each removal. The study uses two removal strategies. In the first, segments identified as highly impactful in the critical-roads analysis are removed in decreasing order of importance. In the second, segments are removed at random. Tracking efficiency under both strategies produces degradation curves that can be compared directly. A network that remains functional under random removals but shows rapid efficiency loss under targeted removals is structurally fragile. This comparison quantifies how dependent the system is on its highest-impact roads and how quickly mobility declines under different failure patterns.

3.4 Order of Road Repair

Determining an effective order of road restoration requires evaluating how much each closed segment contributes to recovering network performance. The process begins with a damaged network in which several roads are unavailable, and the global efficiency of this degraded system is computed to establish a starting point. Each closed segment is then restored temporarily, and the resulting efficiency is measured. The segment that yields the largest improvement becomes the next road to restore. Once this segment is permanently added back to the network, the procedure is repeated with the remaining closed roads. The greedy restoration process continues until all roads are returned. The final sequence identifies the repair order that produces the greatest efficiency gains at each step and offers a practical guideline for prioritizing recovery efforts after major disruptions.

4 Results

4.1 Leander & Cedar Park

1. Critical Roads

   Which set of roads should we protect so the system keeps working best?

   This analysis identifies the road segments that contribute most to maintaining efficient travel. Figure 1 shows the twenty segments that produce the largest drop in global efficiency when removed from the network. The segment with the greatest impact is West Parmer Lane, highlighted in red at the bottom of Figure 1. The removal of this road reduces global efficiency by 1.26 percent. Several other high-impact segments appear along East Whitestone Boulevard, the major east–west corridor in the figure. Their prominence indicates that this corridor plays a central role in supporting short travel paths across the network. Together, these results provide a clear view of the roads the Leander & Cedar Park system relies on most.

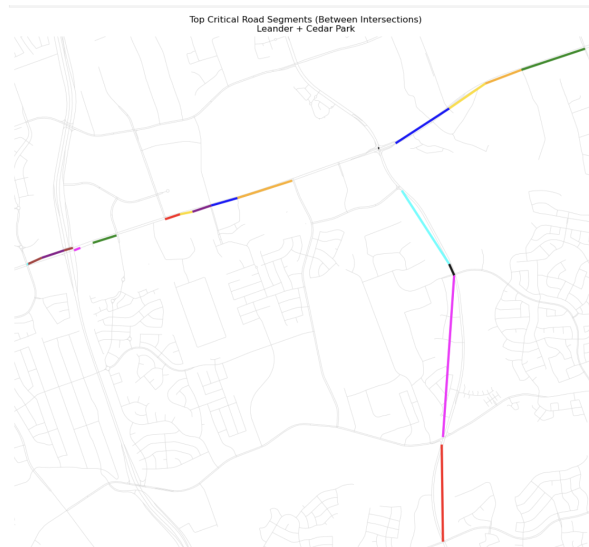

Figure 1: Top-ranked critical road segments

2. Network Robustness

   How fast does the network become less efficient when critical roads close?

   This part of the analysis examines how quickly the network loses efficiency when roads fail. Two scenarios were tested. In the first, road segments were removed at random. In the second, only the most critical roads identified earlier were removed. The results are shown in Figure 2.
   The robustness curve for random removals, shown by the orange line, stays nearly flat, which means the network remains stable even after several closures. In contrast, the curve for targeted removals, shown by the blue line, declines sharply once the highest-impact roads are removed. This pattern shows that the network is resilient to typical, uncoordinated disruptions but becomes fragile when failures occur on a small set of key corridors. The difference between the two curves highlights how unevenly mobility is supported across the network.

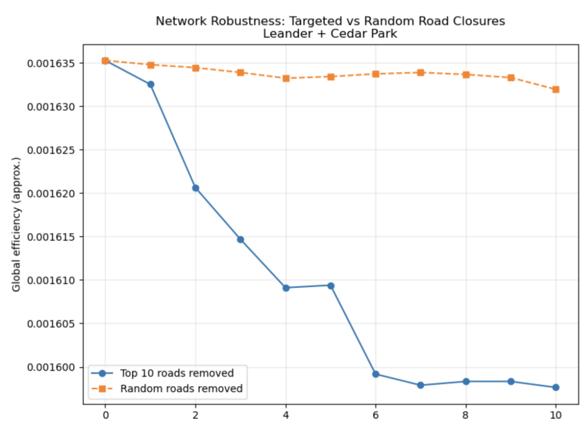

Figure 2: Network robustness in Leander–Cedar Park

3. Order of Road Repair

   After a big disruption, which closed roads should we fix first? What's the best repair order for time saving?

   This analysis identifies an effective order for restoring closed roads after a major disruption. Starting from a worst-case scenario where the top twenty critical segments are removed, each closed road is tested to see how much it improves global efficiency when reopened. The roads are then ranked by their impact, so they are restored in the order that most quickly recovers network performance. As expected, West Parmer Lane, the most critical segment identified earlier, appears first in the repair sequence.

   Most of the high-priority repairs fall along the same major corridors highlighted in the critical-roads analysis, since these roads support many of the shortest travel paths in the area. Roads that provide larger efficiency gains appear early in the sequence, while those with smaller effects appear later. In a few cases, the algorithm selects a middle segment of a corridor to repair before fixing the endpoints. This happens because the repair strategy is greedy. At each step, it chooses the road that gives the largest immediate increase in global efficiency, without considering whether that choice leads to the best overall sequence. A globally optimal ordering would require evaluating many combinations of repair sequences, which quickly becomes computationally infeasible for networks of this size.

   This effect is most noticeable in the Leander & Cedar Park region because many of the major arterial roads are stored in OpenStreetMap as long chains of short, curved segments. When these corridors are broken into dozens of small pieces, some internal segments naturally appear more influential than their neighbors. In contrast, Austin's downtown grid has cleaner and more uniform geometry, so this behavior is much less pronounced there.

4.2 Downtown Austin

1. Critical Roads
   Which set of roads should we protect so the system keeps working best?
   For central Austin, our edge-closure experiment shows that a small set of roads carries most of the connectivity. When we close one road segment at a time and measure the percentage drop in global efficiency, the largest drops (0.8%) occur on several segments of Guadalupe Street near downtown and the UT campus, indicating that many of the fastest routes go through that corridor. Additionally, the next most critical links are West and East Cesar Chavez Street (a key east–west riverfront corridor) and a short segment of Purple Heart Trail / I-35 (***Figure 3***).

   Overall, the roads that should be prioritized for protection in central Austin are Guadalupe Street, the Cesar Chavez corridor, and the Purple Heart Trail / I-35 segment, since disruptions there hurt network performance the most.

2. Network Robustness
   How fast does the network become less efficient when critical roads close?
   We measured how fast the network becomes less efficient by removing roads one at a time and tracking global efficiency. When we remove a few random road segments in central Austin, the efficiency curve is almost flat—closing 4–5 random edges barely changes the average travel efficiency. In contrast, when we remove 4–5 critical edges (highest edge betweenness, like segments on Guadalupe and Cesar Chavez), efficiency drops much more quickly (***Figure 4***). Thus, the network loses efficiency gradually under random failures, but becomes less efficient much more rapidly when those few key roads are closed.

3. Order of Road Repair
   After a big disruption, which closed roads should we fix first? What's the best repair order for time saving?
   For a disruption scenario where several top critical segments are closed, the greedy repair procedure ranks the roads by how much each repair improves efficiency. In Austin, the highest-priority repairs tend to be segments along *Guadalupe Street* and *West Cesar Chavez Street*, followed by important

freeway connectors (***Figure 5***). This suggests that, after a major event, reopening these corridors first would restore network performance the fastest.

5 Challenges

5.1 Expensive Computation

A major pitfall in this project was the sheer size of the Texas road network. The full OpenStreetMap graph for Texas contains on the order of ~ 5 million nodes (intersections) and ~ 7 million edges (road segments), which makes it infeasible to run expensive algorithms such as weighted betweenness and closeness centrality or repeated global efficiency calculations on the entire state graph.

Even after restricting to large metro areas such as Austin and Houston, the subgraphs still had around 1–2 million nodes/edges. At this scale, many of our centrality and robustness computations would either take an extremely long time or exhaust available memory on a typical laptop.

As a result, we narrowed our focus to smaller study areas: a suburban region (Leander & Cedar Park) and the Downtown/Central Austin network. These smaller graphs (thousands of nodes instead of millions) made it practical to compute centrality measures, simulate edge closures, and approximate global efficiency.

A natural direction for future work would be to develop faster methods that can handle the full Texas network, for example by analyzing the state in spatial "chunks" and combining the results, using more efficient low-level graph libraries, or designing multi-scale approaches that summarize the network at a coarser level before zooming in on critical regions.

5.2 False Nodes

When we first extracted the Leander and Cedar Park road network from OpenStreetMap, we found that many intersections were represented by several nodes located only a few meters apart. This occurs because OpenStreetMap stores roads as many small segments, especially around curves, lane splits, and data source boundaries. As a result, one real intersection often appeared in the data as multiple separate points. If left uncorrected, this artificially inflated the number of intersections and made the importance calculations unreliable. A single real intersection could show up several times in the top rankings simply because its geometry was fragmented.

To address this, we applied a node merging procedure that combined any points within 10 to 15 meters into a single true intersection. This step produced a cleaner road network, removed duplicate critical locations, and made the analysis much more realistic. This issue was particularly prominent in the Leander and Cedar Park area because many of the major roads there are long curving arterials that OpenStreetMap represents as sequences of very short polylines. That fragmentation created clusters of nodes only a few meters apart, all representing a single real intersection. Merging nearby points into a single intersection cleaned up the network and prevented the analysis from repeatedly flagging the same physical location as several "critical" spots. Downtown Austin, by contrast, has a more regular grid with straighter road geometries, so it did not exhibit the same level of node duplication and did not require the same merging step.

6 Conclusion

In this project, we modeled parts of the Texas road network as weighted graphs and used global efficiency, centrality, and simulation to study how well different areas support travel. We found that both the Leander/Cedar Park suburbs and Downtown Austin are fairly robust to random road closures, but much more vulnerable when a small number of critical corridors (like East Whitestone in the suburbs and Guadalupe Street / Cesar Chavez in Austin) fail. Overall, the results suggest that a relatively small set of roads carries a disproportionate share of connectivity, and protecting or prioritizing those links could make the network more resilient to disruptions.

References

[1] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Physical Review Letters*, vol. 87, no. 19, 2001.

[2] Geofabrik GmbH, "OpenStreetMap data extracts for GIS formats," accessible at https://download.geofabrik.de/osm-data-in-gis-formats-free.pdf and https://download.geofabrik.de/north-america/us/texas.html

[3] Github link containing code we used https://github.com/lauren-fuller/STAT-636-Texas-Roads

Appendix

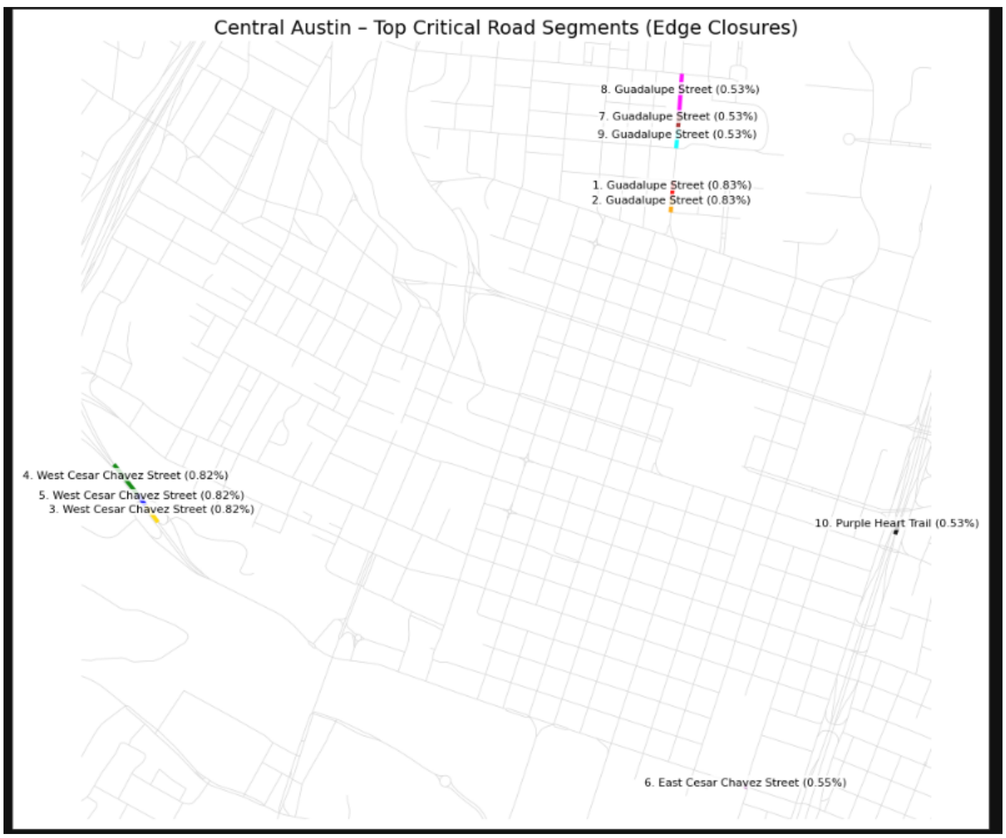

Figure 3: Most Critical Road Segments in Downtown Austin

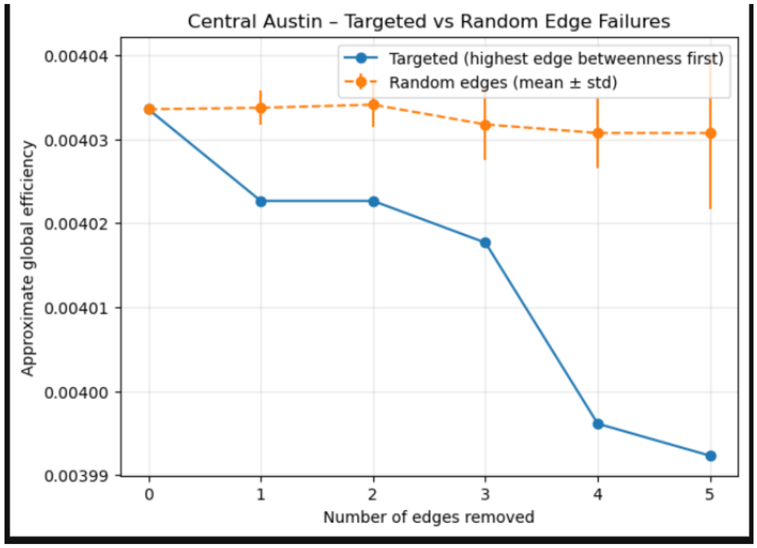

Figure 4: Network Robustness in Downtown Austin

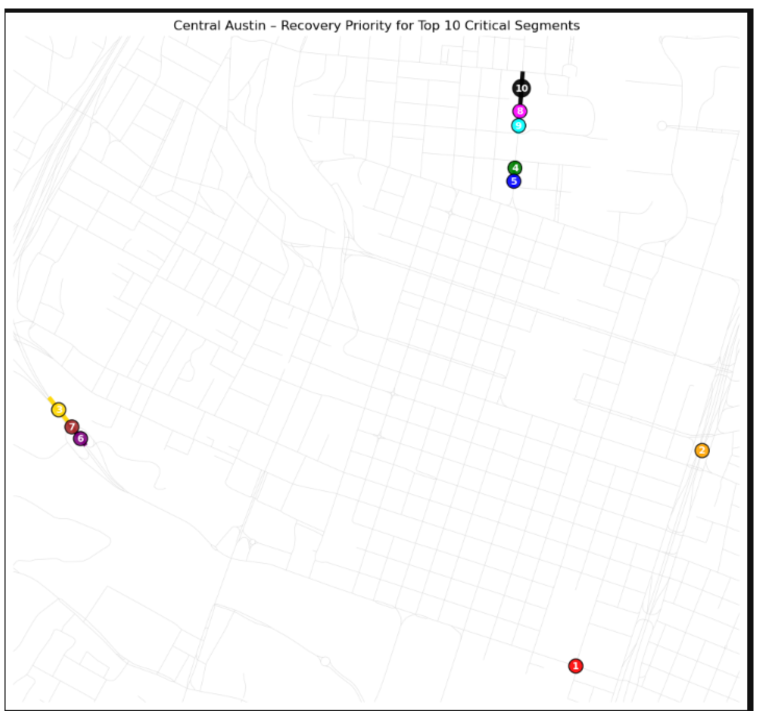

Figure 5: Recovery Strategy for top 10 critical roads in Downtown Austin