Improving the Accuracy of Intersection Counts and Densities for Measuring Urban Street Network Compactness and Resilience

December 2022

A Research Report from the Pacific Southwest Region University Transportation Center

Geoff Boeing, University of Southern California
Jaehyun Ha, University of Southern California
Yuquan Zhou, University of Southern California

Pacific Southwest Region UTC
University Transportation Center

USC University of Southern California

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. PSR-CA23-3459	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A
4. Title and Subtitle Improving the Accuracy of Intersection Counts and Densities for Measuring Urban Street Network Compactness and Resilience		**5. Report Date** 2022-12-31 **6. Performing Organization Code** N/A
7. Author(s) Geoff Boeing, https://orcid.org/0000-0003-1851-6411 Yuquan Zhou, https://orcid.org/ 0000-0001-6984-4373 Jaehyun Ha, https://orcid.org/ 0000-0003-2292-7416		**8. Performing Organization Report No.** TBD
9. Performing Organization Name and Address METRANS Transportation Center University of Southern California University Park Campus, RGL 216 Los Angeles, CA 90089-0626		**10. Work Unit No.** N/A **11. Contract or Grant No.** USDOT Grant 69A3551747109 Caltrans 65A0674
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology 1200 New Jersey Avenue, SE, Washington, DC 20590		**13. Type of Report and Period Covered** Final report (Jan 2022-Dec 2022) **14. Sponsoring Agency Code** USDOT OST-R
15. Supplementary Notes		

16. Abstract

Intersection counts are ubiquitous in transportation planning practice and research. They are frequently normalized by area to calculate intersection density, the most common measure of compact street network design in planning practice. However, due to the nature of typical street network data (centerlines) and the typical tools used to count intersections (desktop geographic information systems [GIS]), traditional methods of counting intersections can significantly overcount them. This project addresses this long-standing problem of intersection count bias. First, it develops and distributes an algorithm to automatically and correctly calculate intersection counts and densities anywhere in the world, using a novel topological consolidation method. Second, it conducts a worldwide empirical assessment of traditional intersection counting methods' bias to quantify the importance of measurement bias and to validate our algorithm. Third, it assesses this bias's impact on resilience simulations' results and identifies the street network design characteristics that are most related to resilience. In transportation planning, innumerable downstream models and measures — from energy efficiency certification schemes to resilience simulations — rely on intersection counts as input data. A full accounting of input data bias and better methods to overcome misrepresentations of intersections are necessary for data-driven, evidence-based planning for sustainable transportation networks that support active and resilient travel.

17. Key Words street networks, resilience, intersection density	**18. Distribution Statement** No restrictions.		
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 88	**22. Price** N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Contents

About the Pacific Southwest Region University Transportation Center

The Pacific Southwest Region University Transportation Center (UTC) is the Region 9 University Transportation Center funded under the US Department of Transportation's University Transportation Centers Program. Established in 2016, the Pacific Southwest Region UTC (PSR) is led by the University of Southern California and includes seven partners: Long Beach State University; University of California, Davis; University of California, Irvine; University of California, Los Angeles; University of Hawaii; Northern Arizona University; Pima Community College.

The Pacific Southwest Region UTC conducts an integrated, multidisciplinary program of research, education and technology transfer aimed at *improving the mobility of people and goods throughout the region.* Our program is organized around four themes: 1) technology to address transportation problems and improve mobility; 2) improving mobility for vulnerable populations; 3) Improving resilience and protecting the environment; and 4) managing mobility in high growth areas.

U.S. Department of Transportation (USDOT) Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

California Department of Transportation (CALTRANS) Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the United States Department of Transportation's University Transportation Centers program, in the interest of information exchange. The U.S. Government and the State of California assumes no liability for the contents or use thereof. Nor does the content necessarily reflect the official views or policies of the U.S. Government and the State of California. This report does not constitute a standard, specification, or regulation. This report does not constitute an endorsement by the California Department of Transportation (Caltrans) of any product described herein.

Disclosure

Principal Investigator and others conducted this research titled, "Improving the Accuracy of Intersection Counts and Densities for Measuring Urban Street Network Compactness and Resilience" at the University of Southern California. The research took place from Jan 2022 to Dec 2022 and was funded by a grant from CalTrans in the amount of $100,000. The research was conducted as part of the Pacific Southwest Region University Transportation Center research program.

Acknowledgements

The authors wish to thank Caltrans for funding this research.

Abstract

Intersection counts are ubiquitous in transportation planning practice and research. They are frequently normalized by area to calculate intersection density, the most common measure of compact street network design in planning practice. However, due to the nature of typical street network data (centerlines) and the typical tools used to count intersections (desktop GIS), traditional methods of counting intersections can significantly overcount them. This project addresses this long-standing problem of intersection count bias. First, it develops and distributes an algorithm to automatically and correctly calculate intersection counts and densities anywhere in the world, using a novel topological consolidation method. Second, it conducts a worldwide empirical assessment of traditional intersection counting methods' bias to quantify the importance of measurement bias and to validate our algorithm. Third, it assesses this bias's impact on resilience simulations' results and identifies the street network design characteristics that are most related to resilience. In transportation planning, innumerable downstream models and measures — from LEED-ND certification to resilience simulations — rely on intersection counts as input data. A full accounting of input data bias and better methods to overcome misrepresentations of intersections are necessary for data-driven, evidence-based planning for sustainable transportation networks that support active and resilient travel.

Improving the Accuracy of Intersection Counts and Densities for Measuring Urban Street Network Compactness and Resilience

Executive Summary

Intersection counts are ubiquitous in transportation planning practice and research. In particular, they are frequently normalized by area to calculate intersection density, the most common measure of compact street network design in planning practice for sustainable transport, active travel, and (alongside connectivity) networks resilient to perturbation. However, due to the nature of typical street network data (centerlines) and the typical tools used to count intersections (desktop GIS), recent research suggests that traditional methods of counting intersections overcount them by 16% on average. Worse, they do so unevenly in different kinds of places, and by as much as 33% in urban areas in certain countries such as Spain and Australia. The nature of this bias means that individual neighborhoods with the most car-centric complex intersections are often the most misrepresented as being more compact and fine-grained than they are in reality. Researchers and practitioners rely on intersection counts and densities as foundational inputs and require accurate measurements to represent the real world accurately to plan resilient and sustainable transportation infrastructure accordingly.

This project addresses the longstanding problem of intersection count bias by 1) developing and distributing an algorithm to automatically and correctly calculate intersection counts and densities anywhere in the world, 2) conducting a worldwide empirical assessment of traditional intersection counting methods' bias to quantify the importance of measurement bias, and 3) assessing this bias's impact on resilience simulations.

This project develops and validates an algorithm that consolidates the multiple network nodes representing complex street intersections. It merges network nodes within some design tolerance of one another and then reconnects network edges correctly to the merged node to maintain the network topology. For each urban area in the world, we assign the design tolerance value for consolidation using estimates of 1) the urban area's average road width and 2) the world median weighted road width from 1990-2015 using data provided by the Atlas of Urban Expansion. After stratifying 8,910 cities by world regions, population size, and intersection density, we validate the consolidation results of 32 cities. The algorithm consolidates intersections with a true positive rate above 95% and true negative rate around

90% for both parameters. The algorithm performed best in cities located in Land Rich Developed Countries and when correcting redundant nodes caused by divided roadways.

Next, using the original and consolidated networks, we measure the resilience of street networks focusing on how quickly networks become disconnected and how network perturbation affects trip efficiency. We conduct the network perturbation simulations by systematically attacking 1) nodes with the highest betweenness centrality, 2) nodes at the lowest elevations, and 3) random nodes. Then we measure how street network resilience differs across cities and describe how unconsolidated networks might lead to over or underestimation of street network resilience. Among the three perturbation methods, our results demonstrate that attacking the nodes with high betweenness centrality (that is, the most "important" nodes in the network) may lead to particularly severe network disruption and longer trip distances. Finally, we identify the street network design characteristics that are most related to resilience. Our results suggest that street networks with lower node degrees, greater circuity, and an over-reliance on important nodes are more vulnerable to network perturbations.

In transportation planning, innumerable downstream models and measures — from LEED-ND certification to resilience simulations — rely on intersection counts as input data. A full accounting of input data bias and better methods to overcome misrepresentations of intersections are necessary for data-driven, evidence-based planning for sustainable transportation networks that support active and resilient travel.

1. Introduction

Street intersections, which are junctions where two or more roads meet, are an important part of the street network characteristics that contribute to urban resilience, such as compactness, centrality, connectivity, walkability, and safety (Boeing, 2017; Sharifi, 2019). Despite street intersections' apparent simplicity, counting them can be nuanced and challenging in practice. Typically, analysts count either planar line intersections (using geometric data like that from TIGER/Line) or topological intersections (using topological data like that from OpenStreetMap), but both methods can exhibit substantial bias. The former ignores network nonplanarity and includes overpasses and underpasses in the count (*Boeing, 2020*). The latter accounts for nonplanarity but continues to overcount complex intersections, such as the intersection of two divided roadways (where two by two centerlines yield a four intersection count), roundabouts, slip lanes, complicated interchanges, etc. This overcount of intersections can in turn bias downstream indicators of compactness, fine-grain, and resilience.

To attenuate this intersection overcounting problem, this research project developed a novel and reusable toolkit to topologically consolidate complex intersections in street network models to bring them closer to actual real-world street networks. The algorithm improves the accuracy of the street network in the following aspects. First, it consolidates multiple nodes comprising complex intersections due to divided roadways, slip lanes, roundabouts, etc. that should belong to the same intersection in the real world. Second, unlike geometric consolidation, it prevents the false consolidation of topologically remote but spatially proximate nodes, such as unconnected intersections on overpasses and underpasses.

This study used two parameterizations to consolidate complex intersections across 8,910 worldwide urban street networks. One parameterization is customized for each city, calculated by spatially matching it to the closest city with road width information. The second is a uniform parameterization representing the estimated global median road width. We then stratified all the cities and randomly selected 32 for consolidation validation. The results demonstrate that the algorithm effectively corrected the street network and could be a useful tool for future research and practice. Finally, we investigated the resilience of the original and consolidated street networks and explored how the consolidated models shift our understanding of street network resilience through a series of simulations.

This technical report is structured as follows. First, it starts with a literature review that describes the study's context and current research gaps. Second, it describes our research process and data sources. Third, it presents the results of our intersection consolidation, validation, and resilience simulations. Finally, it concludes with a discussion of how these methods can be used and how they impact understandings of street network resilience.

2. Background

2.1 The importance of intersection counts

Street intersections are closely associated with measures of street network resilience, defined as the ability to keep the city connected and accessible even if the street network is disrupted (Mattsson and Jenelius, 2015). Those indicators include intersection density, connectivity, disconnectedness, centrality, circuity, efficiency, and so on. Greater street network connectedness and resilience can be indicated by greater intersection density, a shorter average distance between intersections, and a higher mean node degree of intersections (Barrington-Leigh and Millard-Ball, 2015; Salat et al., 2010; Sharifi, 2019). Street connectedness, which can influence people's decisions to walk and take public transit, is related to pedestrian accessibility and urban walkability, and thus has ties to public health (Ewing and Cervero, 2010). Street network connectivity also affects people's driving behaviors. More connected street networks can lower the vehicle miles traveled (VMT) by households (Boarnet et al., 2004).

Recently, scholars have begun to focus on the inverse of connectivity — disconnectedness, a major characteristic of urban sprawl, which has a positive association with increased VMT, congestion delays, energy use, and greenhouse gas (GHG) emissions (Barrington-Leigh and Millard-Ball, 2017 ; Boarnet et al., 2004; Litman, 2003). The more disconnected street networks are less resilient, as trips themselves can easily become at best inefficient and at worst impossible to complete when facing enough street network disruption. In recent years, a number of studies have assessed street network sprawl on a large scale. Barrington-Leigh and Millard-Ball (2015) gave a quantitative history of urban sprawl in the United States and defined it as having a low mean node degree of intersections and a high percentage of dead ends. In 2020, Barrington-Leigh and Millard-Ball expanded their study to a worldwide scale and analyzed the global time series of urban street network sprawl by generating an integrated Street-Network Disconnectedness index (SNDi) based on the mapped nodes (intersections) and edges (street segments) in the world, and found that in large parts of the world, recent urban growth has resulted in inflexible and disconnected street networks (Barrington-Leigh and Millard-Ball, 2020). In 2021, Boeing modeled and analyzed street networks based on their griddedness, orientation order, straightness, 4-way intersections, and intersection density in every US census tract and found similar results to Barrington-Leigh and Millard-Ball (2015) that the street networks in the US grew more disconnected between the 1940s and 1990s and started to return to more connected griddedness over the past 20 years (Boeing, 2021a).

Modeling intersections correctly is essential for measuring these values that depend on them. Furthermore, the representation of intersections contributes to the measurement of centrality, which is a critical indicator of urban resilience. High-centrality intersections are of great importance to the urban street network, and their disruption can result in a loss of reachability (Sharifi, 2019). Wang (2015) compared the resilience of the street networks of London and

Beijing by measuring betweenness centrality, closeness centrality, network efficiency, and simulating random attacks. He found that the gridded street network in Beijing is more resilient than alternatives.

As these studies show, the geometry and topology of street intersections contributes significantly to street network resilience, connectivity, disconnectedness, and centrality. The potential miscount of intersections, and in turn downstream measures, could bias our understanding of the street network's performance.

2.2 The miscount of intersections

Even though street intersections are important to street network analysis, they are hard to define and count. Counting and analyzing street networks using potentially incorrect methodologies may provide biased and inconsistent results. Existing studies used street centerline intersections to represent real-world street intersections and adopted GIS tools such as ArcGIS and QGIS to calculate intersection density from geometric data such as publicly accessible TIGER/Line shapefiles (Wang et al., 2020; Xue et al., 2020). However, counting street intersections using the above method may be problematic. On the one hand, the completeness, consistency, accuracy, and recency of its public geometric data may be doubtful (Boeing, 2022; Frizzelle et al., 2009). On the other hand, the majority of these researchers count the intersections of planar lines in their models of 2-D (two-dimensional) street networks. But in reality, due to underpasses, overpasses, and tunnels, the real-world street network is not planar. Most of the time, imposing a planar model is used due to its computational simplicity and efficiency, as well as the approximate planarity of the real world. However, it generates false nodes at any line intersections that are not real-world intersections. Boeing (2020) investigated the disparity between the planar street network model and the nonplanar reality. The study examined street networks in the cores of 50 cities worldwide and found that planar models overestimate intersection counts due to the presence of false nodes. This result indicated that planar models can distort urban form and urban resilience analysis (Boeing, 2020). Moreover, as cities have grown more vertical and complex over the past century, modern and automobile-dominant cities show more nonplanarity than traditional old cities or towns (Bruyns et al., 2021). Therefore, for instance, in parts of Los Angeles, a metropolis dominated by automobiles, planar models overcounted intersections by 72%. (Boeing, 2020).

Other research uses topological models with open data sources, such as OpenStreetMap (OSM), an online crowdsourced free world mapping collaboration, and attaches elevations to the nodes to analyze the street network and count intersections (Boeing, 2017, 2022). While this method could potentially address the above nonplanarity and data inconsistency problems, it still overcounts complex intersections, such as the intersection of two divided roads (where two by two centerlines yield four intersections' count), roundabouts, slip lanes, complex junctions, etc. (Barrington-Leigh & Millard-Ball, 2020).

This misrepresentation of the intersection in the real world would lead to biased street network analysis results and resilience estimates. Consider modeling a disturbance at a two-by-two divided roadway intersection, for instance. The street network would erroneously appear resilient if the model yielded four nodes and the disruption needed to target all four nodes to disconnect the street network. In reality, however, there is only one intersection, and perturbing a single node is sufficient to disconnect the street network. The correct representation of the street network is needed for accurate street network analysis.

2.3 Endeavors to address intersection miscounting

At the current stage, nonplanarity could be addressed by employing computational approaches to distinguish planar line crossings such as overpasses and underpasses by incorporating elevation data or topological connection data (Boeing, 2020). However, in large-scale street network modeling, the overcounting of intersections in circumstances such as divided roadways and complex junctions has not yet been adequately addressed. Therefore, automating the accurate count of street intersections has remained an ongoing challenge (Li et al., 2020). Researchers have been attempting to address the miscount of intersections and automatically consolidate redundant nodes since the end of the twentieth century. Their work primarily aims for consolidating redundant nodes in interchanges, that is, junctions where travelers (usually motorists) can switch from one road to another using one or more ramps without directly crossing any traffic stream (Zhou and Li, 2015).

2.3.1 Clustering method

Mackaness & Mackechnie (1999) conducted one of the first studies to detect and simplify complex road junctions, which mainly used the clustering analysis method. They viewed complex junctions as a scale-dependent issue and proposed a method for the identification and simplification of road junctions. The method first detects relatively dense regions of nodes using spatial hierarchical clustering methods geometrically on a 2-D level and then generalizes them using graphic theory by dividing the cluster into subgraphs, removing internal nodes, and calculating the centroid of the remaining nodes. The challenge of this method is setting the right level of threshold parameters, including the distances between two points or clusters and the least number of points in a cluster. To add, this method has the limitation of only considering the internal geometric properties of the network (Mackaness and Mackechnie, 1999).

Built on the above clustering algorithm, instead of clustering all ordinary intersection nodes in Mackaness and Mackechnie (1999) to detect highway interchanges, Touya (2010) clustered fork and y-nodes, which were seen as two main characteristics of highway interchanges (Touya, 2010).

Zhou and Li (2015) investigated the aforementioned two clustering-based automatic interchange detection approaches and concluded that the recognition of distinctive road intersections is essential for effective detection. Therefore, Zhou and Li (2015) extended Touya (2010)'s two characteristics of interchanges and determined nine types of road intersections, including T, y, Y, arrow, cross, X, fork, K, and multi-leg shaped road intersections, to improve the identification accuracy and achieve better validation results (*Zhou and Li, 2015*).

Since 1999, detection methods based on clustering analysis have changed and improved. The hard part of this approach is setting parameters like the number of intersections in a cluster, the number of threshold road angles, and the distance between points or clusters. Setting the wrong parameters would lead to wrong results that would have to be checked by hand to see if they are right. For example, for complex interchanges, if the cluster threshold is too loose, we could potentially take intersections that are actually not a part of the interchanges into account. Adversely, if the cluster threshold is too tight, we may end up not identifying some complex interchanges. Due to the different sizes and shapes of cities around the world, it is not practical to use the same threshold with these clustering-based methods to find intersections on a large regional or global scale.

2.3.2 Deep learning method

With the rapid development of deep learning, especially the use of convolutional neural networks (CNNs), there have been significant breakthroughs in image classification and target detection, thus introducing new methods and opportunities for complex junction detection (Carleo and Troyer, 2017; Lecun et al., 2015). In recent years, a number of researchers have utilized advanced deep learning techniques to detect complicated interactions. Li et al. (2019) proposed a target detection model of deep learning Faster-Region Convolutional neural network (Faster-RCNN) that uses raster representation of vector data and convolutional neural networks (CNNs) to learn target features and identifies the location of an overpass by a Region Proposal Network (RPN) (Li et al., 2019). This study preprocesses the road data from OSM and performs data screening, labeling, geometric metrics calculation, and data enhancement and conversion. Then, it uses the data to train Faster-RCNN, optimizes parameters and structure, and uses the best model to find complex junctions. The authors validated the method in Beijing, Shanghai, Shenzhen, Wuhan, Zhengzhou, and Lanzhou, China. The result shows that the accuracy rate of overpass identification using the proposed method reached more than 90%, proving the effectiveness of this method.

Li et al. (2020) defined a complex junction as a set of stacked intersecting roads with at least three junction nodes, including dual roadways and ramps, and developed a CNN-based method based on the GoogLeNet model to improve the recognition accuracy for complex junctions (Li et al., 2020). Using the Delaunay triangulation clustering algorithm, the method first identifies the center point and spatial range of the complex junctions in the training vector sample.

Second, they augmented vector training samples through simplification, rotation, and mirroring. Finally, the vector sample data were transformed into raster images, and the GoogLeNet model was trained to learn the characteristics. The result shows that the GoogLeNet model has high calculation efficiency in recognizing complex junctions in the test sample.

Both of these two studies above show that deep learning models are effective in complex junction detection. However, in the preprocessing step, these two methods need to turn vector data into raster data to find complex intersections, which may make it harder to do topological and geometric analysis. Additionally, this detection method may be too technical for urban planners.

2.3.3 Topological method

Researchers have also used topological methods to generalize street networks and solve the overcounting problem of intersections. Yang et al. (2021) identified complex junctions using topological method, which overcomes the previous threshold setting problems of the clustering analysis methods and maintains the topological validity of the simplified road network (Yang et al., 2021). The authors pointed out that complex junctions are made up of primary (major roads that connect places across geographic space) and secondary roads, which could be clearly identified by transforming an ordinary segment-based representation of a road network into a stroke-based representation and then to a dual topological representation. Originally, the concept of road stroke was introduced by Thomson and Richardson (1999), which is defined as a long continuous road composed of several segments and has the advantage of good continuity, meeting the cognition criteria of Gestalt principles (Thomson and Richardson, 1999 ; Yu et al., 2019). The paper tested their method on San Diego city, by first preprocessing the data in ArcGIS Desktop to assemble road strokes, pair dual roadways, categorize primary or secondary type of roads based on the road attribute data. Then, it transforms the streets into a dual network of the stroke-based representation of the dataset and uses the proposed algorithm to find complex intersections. The result shows that none of the complex junctions that needed to be found were missed or added by mistake.

The preprocessing portion of this method is critical because insufficient processing or errors, such as misclassification of the roads, stroke assembly failure, and dual-line pairing problems, would significantly reduce the accuracy. However, not every street network dataset will contain embedded road features. Therefore, the preprocessing step remains a challenge for the method's generalizability. Yang et al. (2021) also pointed out that there are no known error-free algorithms for the preprocessing operations of data enrichment, i.e., stroke assembly, counterpart pairing, and road classification, and manual check and correction after preprocessing is necessary. In addition, Boeing (2017) argues that dual graphs overlook important street network spatial characteristics, including length, shape, width, and circuity, and are thus not ideal for analyzing the performance and functionality of street networks.

2.3.4 Open problem

The aforementioned existing methodologies laid the groundwork for the rectification and simplification of street network models. However, the majority of them primarily addressed the consolidation of particularly complex intersections, such as highway interchanges, and did not address other types of intersection overcount issues, such as divided roadways, roundabouts, and slip lanes. In addition, they have inherent restrictions that preclude them from being applied to urban street networks worldwide. The clustering method is limited by the need to specify the proper parameters and utilizes geometric roadway networks that may contain typologically distant intersections. The approaches for deep learning require the transition between vector and raster datasets as well as labor-intensive manual labeling. The topological technique requires laborious preprocessing measures. Transportation practitioners and researchers need a new, reusable, and generalizable method to consolidate all kinds of redundant intersections and produce more accurate street network models. This study's methods aim to address this problem.

3. Methods

This study develops a new topological method for complex intersection consolidation. We develop and distribute an algorithm to automatically consolidate complex intersections for accurate modeling and analysis anywhere in the world. Then we conduct a worldwide empirical assessment of traditional intersection counting methods' bias, to quantify the importance of measurement bias. Finally, we assess this bias's impact on network resilience simulations around the world.

3.1. Input data

All of these input data sources are detailed (including contents and sources) in the Data Management Plan at the conclusion of this document and are briefly listed and summarized here to explain our analytical methods. Note that all of these data are publicly accessible for reuse at the locations listed in the Data Management Plan.

3.1.1. Global Urban Street Networks GraphML

The global urban street networks GraphML data set is the primary input dataset used for intersection consolidation and resilience simulation. It includes urban street network models in GraphML file format for 8,914 cities from 178 countries and regions in the world — representing every urban area in the world, per the Global Human Settlement Layer (more details below).

3.1.2. Global Urban Street Networks Indicators

The Global Urban Street Networks Indicators file includes 8,910 cities' metrics on street networks, population, and area. We used the total resident population in 2015, and the area within urban center boundary to stratify cities for subsequent validation.

3.1.3. Global Human Settlement Layer 2015, multitemporal and multidimensional attributes

The Global Human Settlement Layer (GHSL) dataset is supported by the Joint Research Centre (JRC) and the Directorate General for Regional Development (DG REGIO) of the European Commission. It generated global spatial data on human presence around the world using automatic spatial data mining methods and has the information on urban center coordinates and urban population. We used GHSL data to match urban centroid coordinates to the 8,910 studied cities in the global urban street networks indicators file.

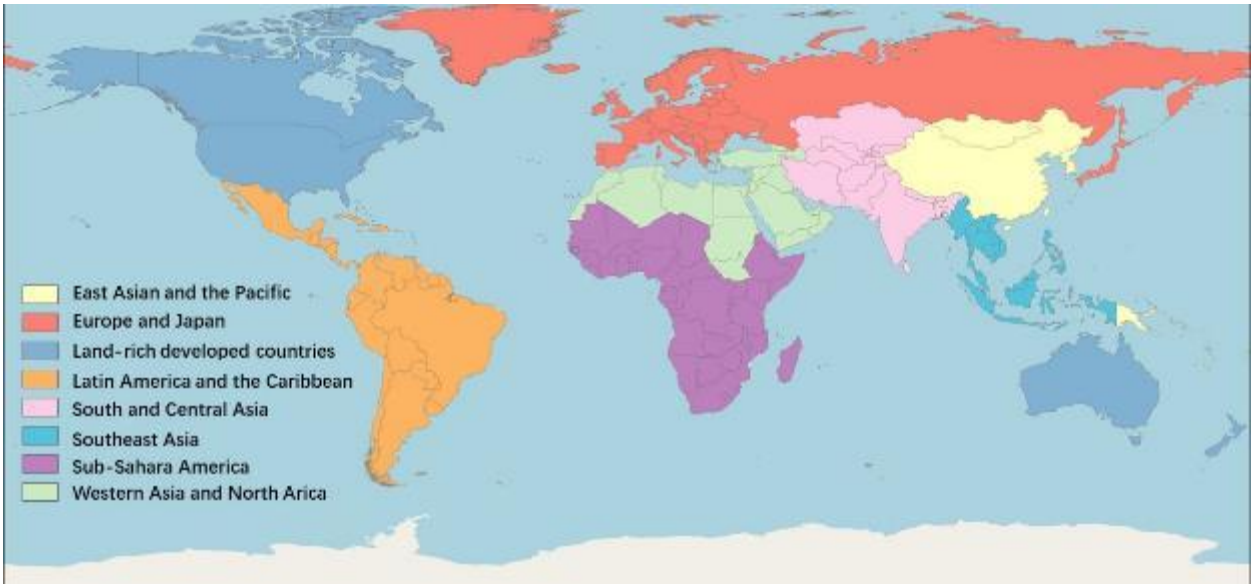


Figure 1. Sampled 200 cities in eight world regions

3.1.4. Atlas of Urban Expansion

When figuring out the best design tolerance parameterization for a city, the street width is a key point of reference. However, there is no existing comprehensive street width data at the city level, but there is a sample dataset from the Atlas of Urban Expansion. The Atlas of Urban Expansion dataset is a result of collaboration between UN-Habitat, NYU, and the Lincoln Institute of Land Policy, and it contains the average street width of 200 cities, stratified across eight world regions (Figure 1). The eight world regions were defined by the Atlas of Urban Expansion project, which largely mirrors divisions in United Nations' World Urbanization

Prospects, with minor changes, considering regions' economic development level and urban expansion pattern (U.N. Population Division, 2014, Angel et al., 2016a, 2016b).

For each city, this dataset includes two indicators of their road width - the average road width of the urban areas that were built before the 1990s, and the average road width of areas built between around 1990 and 2015. In addition, the dataset also includes the average density of all arterial roads in urban areas built before the 1990s and between around 1990 and 2015. Table 1 shows that on average cities in land-rich developed countries have the widest streets among eight world regions.

Our research team discovered that the five cities' coordinates were recorded incorrectly in this dataset (Lagos, Nigeria; Port Elizabeth, South Africa; Johannesburg, South Africa; Oyo, Nigeria; and Kigali, Rwanda) and we corrected it in the dataset for our subsequent analysis.

Table 1. Average road width and number of sampled cities by world region (Atlas of Urban Expansion)

Region	Averaged by city: Average Road Width (meters) Pre-1990	Averaged by city: Average Road Width (meters) 1990 - 2015	Number of Sampled cities
East Asia and the Pacific	8.72	7.49	42
Europe and Japan	8.10	6.30	34
Land-Rich Developed Countries	11.16	11.26	18
Latin America and the Caribbean	10.17	7.49	26
South and Central Asia	7.66	6.5	32
Southeast Asia	8.01	6.39	15
Sub-Saharan Africa	8.79	6.09	18
Western Asia and North Africa	9.65	8.35	15

3.1.5. Countries geometries

ArcGIS Hub offers a shapefile defining the geometries of the world's countries. We accessed and downloaded this shapefile on July 13, 2022, and used them to assign the world countries to the eight world regions.

3.2. Intersection consolidation

3.2.1 Algorithm

The intersection consolidation algorithm merges multiple nodes representing a single complex intersection into a single node and reconnects the network's edges to maintain a correct topology. It operates according to seven general steps. First, it geometrically buffers the graph's nodes with the radius set to a user-selected design tolerance (discussed in detail below), then merges overlapping buffers to represent node clusters. Second, it attaches each graph node to its cluster of merged nodes via a spatial join. Third, if a cluster contains multiple components (i.e., it is not fully connected), the algorithm moves each subcomponent to its own cluster so as not to subsequently link nodes that are not truly connected in the real-world (e.g., nearby dead-ends or surface street intersections below a bridge). Fourth, the algorithm creates a new empty graph. Fifth, it adds a new node for each cluster of merged nodes and adds the appropriate node-level attributes to each. Sixth, it creates a new graph edge from cluster to cluster for each edge that existed in the original graph. Seventh, for every cluster containing more than one node from the original graph, regenerates the edge geometries to extend to the new node's coordinates, and then updates the edge length accordingly. We have packaged and distributed this algorithm in the OSMnx package at https://pypi.org/project/osmnx/

3.2.2 Parameterization

Given the well-known bias-variance trade-off in statistical modeling, there will be compromises between false positive and false negative consolidations. Figures 2 through 5 illustrate why selecting the proper parameterization is essential and why street width data offers an ideal method for parameterization. For instance, a divided 20m by 20m road intersection yields four nodes (a,b,c,d) whereas a one-divided 20m by 10m road intersection yields two nodes (e, f) (Figure 2). In reality, a,b,c,d on the left are viewed as one intersection, and e and f on the right are considered the other intersection. For example, in Figure 3, we merge nodes within a tolerance radius of 3 meters. This parameter is too conservative to consolidate these nodes together – resulting in a false negative case. In figure 4, the parameter is set to 15m (approximately the street width), where the buffers of a,b,c,d overlap, and the algorithm consolidates these four nodes into a single node A. Same for nodes e and f, the algorithm consolidates them into node B. In this instance (figure 4), the algorithm is performing as expected, which is a true positive case. In figure 5, the parameter is set too aggressively - at 30 meters, the buffers of two distinct junctions overlap, and the algorithm incorrectly consolidates

these two intersections into node C. It is a false positive, as the algorithm incorrectly consolidates distinct intersections.

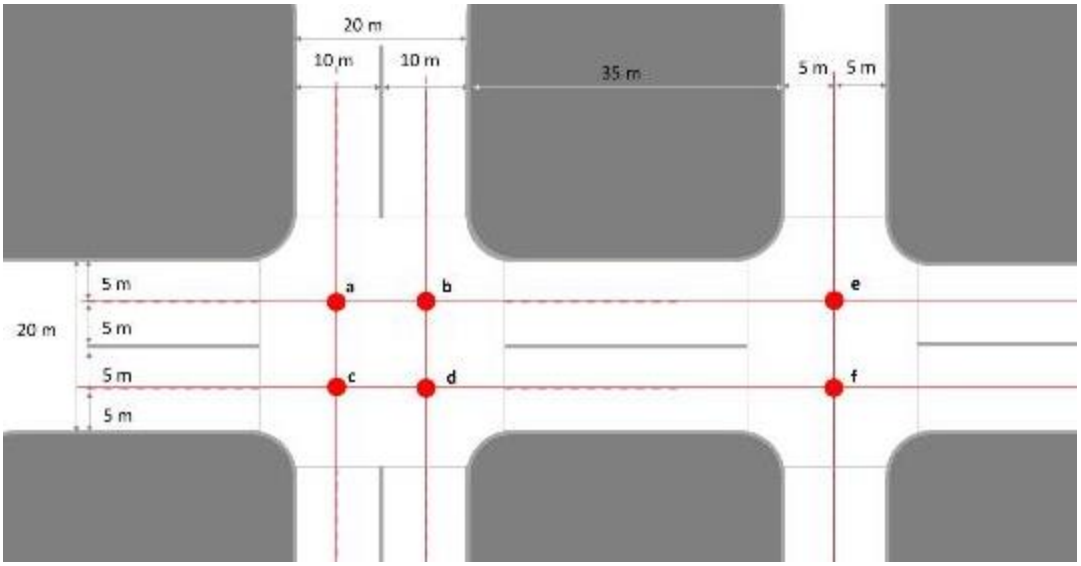


Figure 2. Example intersection needs consolidation

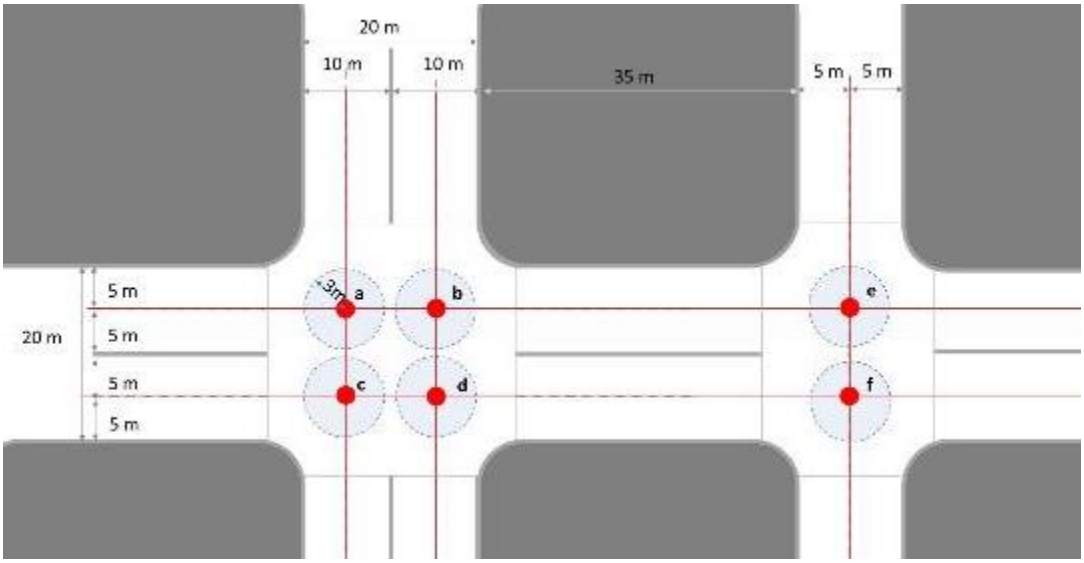


Figure 3. Example consolidated by a 3m buffer (false negative case)

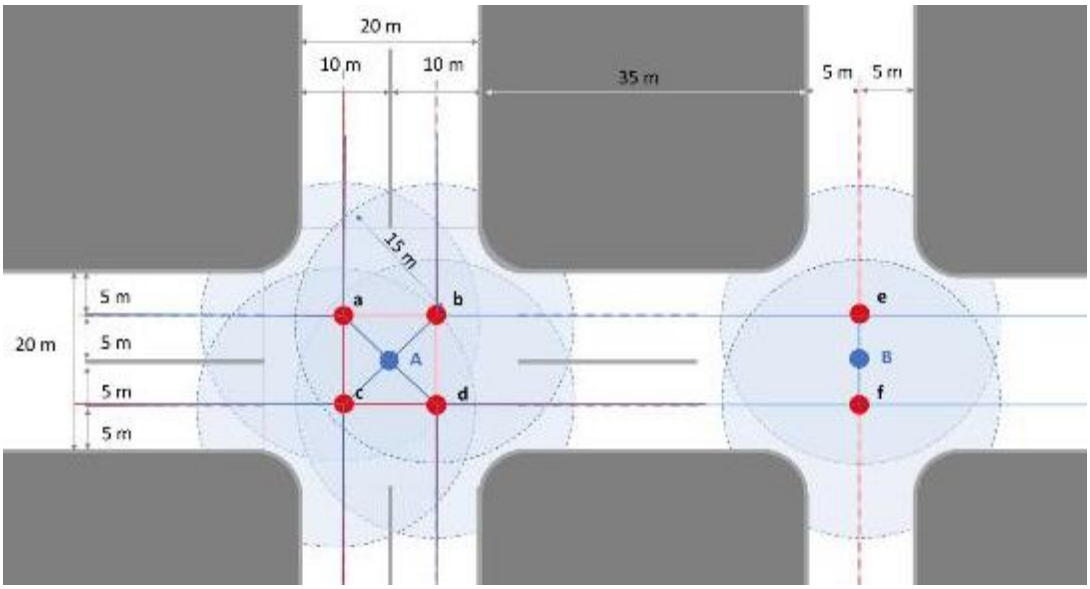


Figure 4. Example consolidated by a 15m buffer (true positive case)

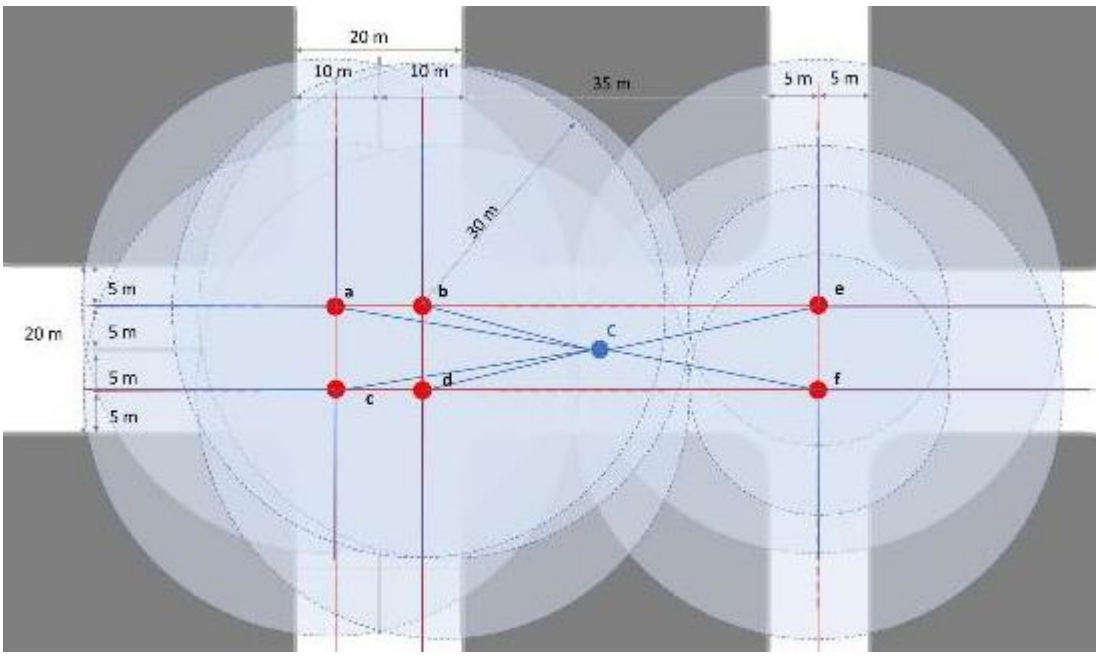


Figure 5. Example consolidated by a 30m buffer (false positive case)

We seek to investigate which parameterizations function well in different cities. The average street width is a useful determinant for parameterization. For example, we should set a larger tolerance for cities with wide streets and a coarse-grained network in order to merge nodes comprising complex intersections, but a smaller one for cities with narrow streets and a fine-grained network in order to prevent incorrectly merging distinct intersections.

The dataset from the Atlas of Urban Expansion specified the average road width by two geographical extents of 200 sampled cities (1) urban areas built before 1990; (2) urban areas built between around 1990 and 2015. In order to best portray the average road width of the complete urban extent today, this study pooled the above two indicators together and calculated the weighted average road width of each city using Equation 1.

Equation 1:

$$\omega = \alpha_{1990} * \gamma_{1990} / (\alpha_{1990} * \gamma_{1990} + \alpha_{1990-2015} * \gamma_{1990-2015}) * \omega_{1990} + \alpha_{1990-2015} * \gamma_{1990-2015} / (\alpha_{1990} * \gamma_{1990} + \alpha_{1990-2015} * \gamma_{1990-2015}) * \omega_{1990-2015}$$

Where: subscript 1990 and 1990-2015 refers to urban extent built before 1990, and built between 1990 - 2015, respectively. Where ω refers to average road width (m), α refers to urban extent area (ha) and γ refers to average density of all arterial roads (km/km^2).

The boxplot of the weighted road width of the 200 cities sampled in Atlas of Urban Expansion is shown in Figure 6. Cities in land-Rich developed countries have the highest median matched road width (10.45m). While the cities in South and Central Asia have the lowest median matched road width (6.85m).

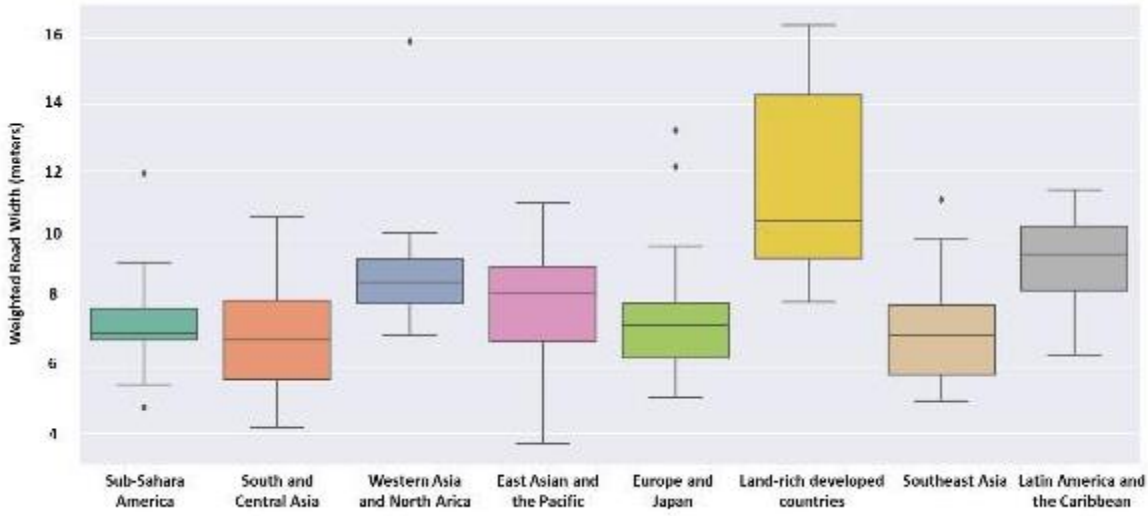


Figure 6. Distribution of cities with matched road width by world regions

As there are 8,910 cities with street network models that we are interested in examining, but only 200 cities have street width information, we need to assign each city its best proxy for street width information. After attaching the studied cities' urban centroid coordinates from GHSL data, we used spatial matching to identify the closest city with street width data within the same global region and used the matched city's weighted street width as a proxy of each city's average road width. We made this decision based on the premise that cities closer in

proximity should share a more similar urbanization background and pattern and, hence, more comparable street design.

As an additional, uniform, global parameter for network consolidation, we consolidated street networks using the median weighted road width of the 200 sampled cities (8.0 meters). In practice, this parameter would be useful for places with no matched road width information, as this provides information about "typical" street widths in the world.

Specifically, the following is the workflow we used to match all 8,910 cities with 200 sampled cities with street width information. First, we assigned 249 countries in the world (from ArcGIS hub, accessed 2022, July 13) to eight world regions to generate a list of countries matched to the world region. Second, we then assigned 8,410 cities with eight world regions via country ISO code. In particular, due to the mismatch between countries from ArcGIS Hub and countries listed in the graphml file, we manually assigned Taiwan (TWN), Kosovo (XKO), and Western Sahara (ESH) to East Asian and the Pacific, Europe and Japan, and Western Asia and North Africa, respectively. Third, we conducted a spatial join and matched each of the 8,410 cities to the nearest sampled cities with weighted average road width information within each world region.

3.2.3 Consolidation

Using the new algorithm, we consolidated the street networks of 8,910 cities using these two parameters: 1) matched street width and 2) world street width median (8.0m). We input the GraphML files of 8,910 cities street networks, consolidated by the above two parameters, and generated the output of 1) GraphML files of street networks consolidated by matched street width and world street width median (8.0m), and 2) a data set of how many nodes remained by each consolidation.

We analyzed the percentage of nodes consolidated by world region, city size, and intersection density of the studied cities. The city size and intersection density categories are defined in Table 2. The categorization of city size is based on the Organization for Economic Co-operation and Development's definition of urban regions (OECD, 2022), and the classification of intersection density is based on the intersection density quantiles of the studied cities.

Table 2. City size and intersection density category definition

City size category	Population	Compactness category	Intersection density count/km2
Small City	pop< =500,000	Low density (bottom quantile)	density<13.90
Medium City	200,000<=pop< 500,000	Lower median high density (0.25-0.5 quantile)	13.90<=density<46.49
Large City	500,000<=pop<1,500,000	Upper median high density (0.5-0.75 quantile)	46.49<=density<87.65
Megacity	1,500,000<=pop	High Density (top quantile)	87.65<= density

Figures 7 and 8 demonstrate that after consolidation, large cities have a lower median percentage of intersections remaining than small cities. A potential explanation is that small cities have simpler street networks that need minimal consolidation to correct, while complex street networks in large cities need substantial correction. This may also reflect the greater detail of digitalization in larger cities. To add, the street networks in small cities are likely to be smaller, and even a few node consolidations can shift a large percentage of node change, leading to extreme outliers.

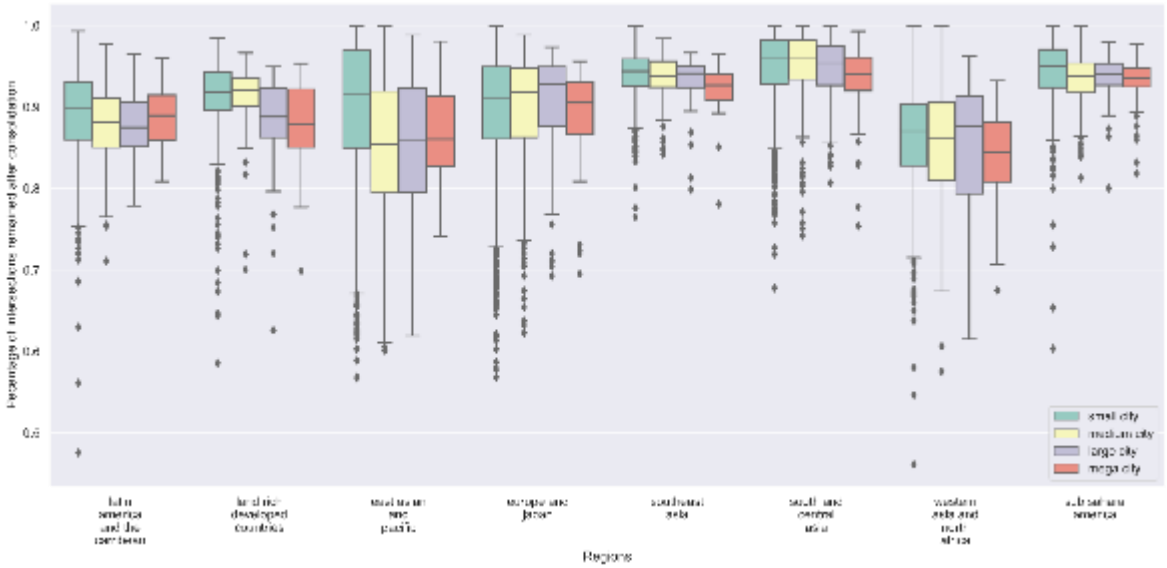


Figure 7. Percentage of intersections remaining after matched road width consolidation, grouping by world region and city size

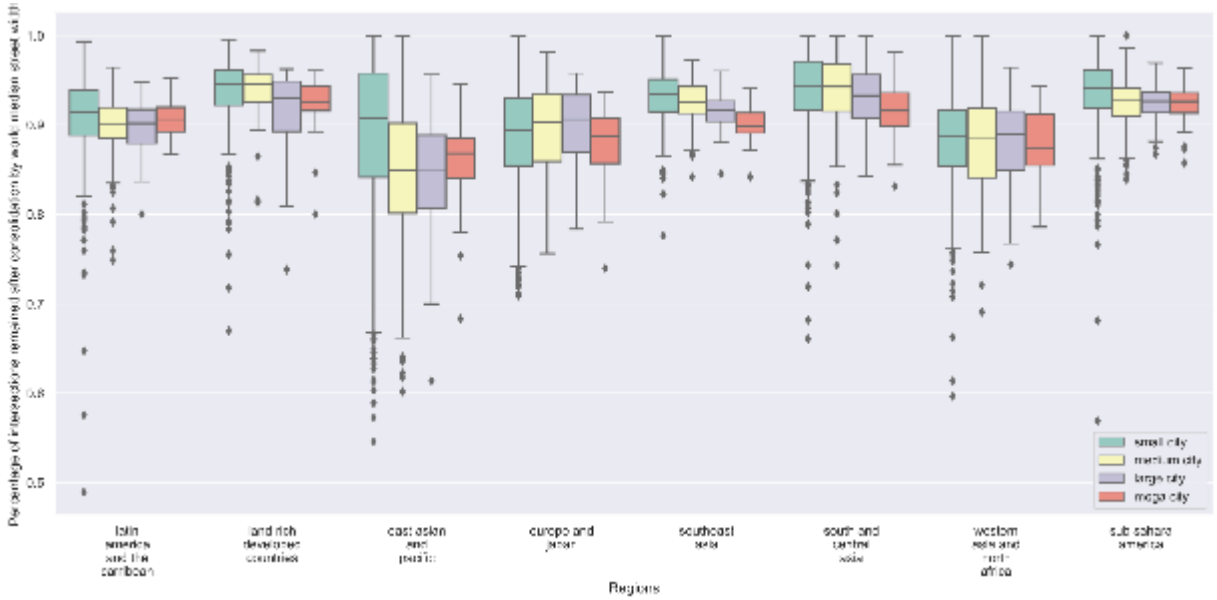


Figure 8. Percentage of intersections remaining after world median road width (8.0m) consolidation, grouped by world region and city size

Figures 9 through 10 show that a higher percentage of nodes get consolidated in cities with high intersection density compared with cities with low intersection density. This indicates that cities with high intersection density are more susceptible to the overcounts of redundant nodes.

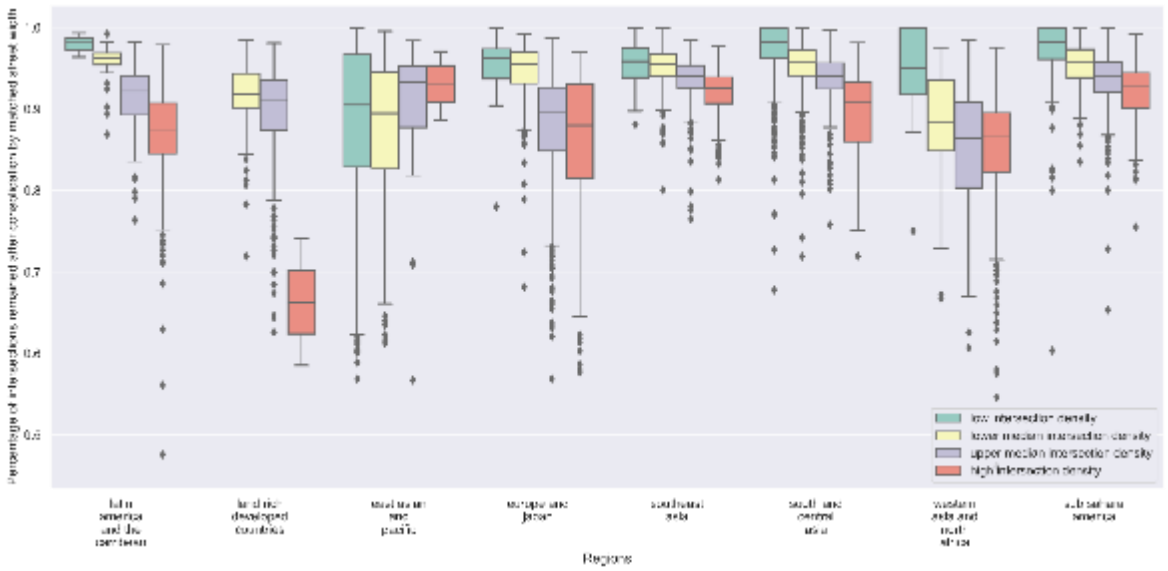


Figure 9. Percentage of intersection remained after matched road width consolidation – group by world region and intersection density

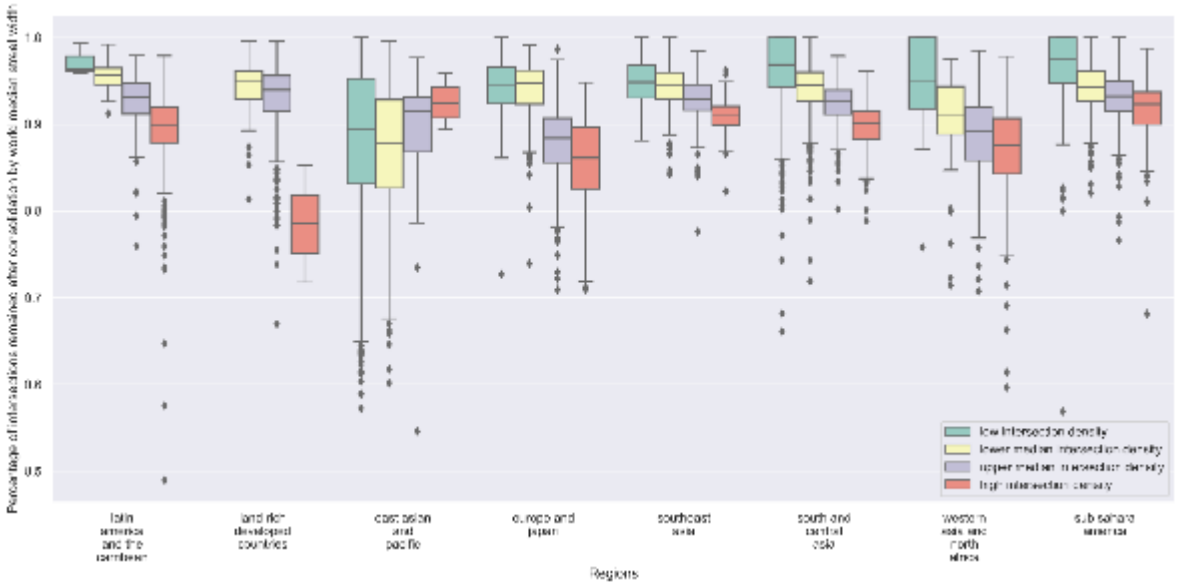


Figure 10. Percentage of intersection remained after world median road width (8.0m) consolidation – group by world region and intersection density

3.2.4 Validation

Stratified random selection of cities

As shown in Figures 7 through 10, the proportion of consolidated nodes is strongly correlated to a city's population (city size), intersection density (compactness), and the world region it belongs to. To get a diverse sample for validation and include a variety of scenarios, we stratified all the cities by world region, population, and intersection density, and then randomly selected 32 cities within each stratified subgroup. Specifically, we classified each city into one of eight global regions defined by Atlas of Urban Expansion. Within each world region, we divided cities into two population groups based on their respective median populations in 2015 (variable: resident_pop in Global Urban Street Networks Indicators file), and then further into two subgroups based on their respective median intersection density (calculated by dividing the number of original intersection count by variable: area in the Global Urban Street Networks Indicator file).

Excluding the unnamed cities, we randomly sampled one city from each subgroup (32 cities out of 8,712 cities). To preserve the sample's diversity, we ran the random selection algorithm three times and selected one of the outcomes where cities within each region were not all from the same country. The stratified random selection of cities for validation is shown in Table 3.

Table 3. Stratified random selection of cities

Region	Population	Original intersection density	City
East Asian and Pacific	High Population 111870.0	High intersection density (10.91)	Hohhot, China
		Low intersection density	Quangang, China
	Low Population	High intersection density (9.083)	Pyeongtaek Si, South Korea
		Low intersection density	Jiangyuan, China
Europe and Japan	High Population 108993.5	High intersection density (61.26)	Belfast, United Kingdom
		Low intersection density	Liege, Belgium
	Low Population	High intersection density (68.47)	Pisa, Italy
		Low intersection density	Uman, Ukraine
Land Rich Countries	High Population 109745.0	High intersection density (47.7)	Brisbane, Australia
		Low intersection density	Sherbrooke, Canada
	Low Population	High intersection density (52.1)	Boulder, United States
		Low intersection density	Abington, United States
Latin America and the Caribbean	High Population 106410.5	High intersection density (101.13)	Juliaca, Peru

		Low intersection density	Port of Spain, Trinidad and Tobago
	Low Population	High intersection density (103.06)	Huajuapan, Mexico
		Low intersection density	Chiquimula,Guatemala
South and central Asia	High Population 143412	High intersection density (21.72)	Chilakaluripet, India
		Low intersection density	Chakaria, Bangladesh
	Low Population	High intersection density (21.81)	Kokshetau, Kazakhstan
		Low intersection density	Jaspur, India
Southeast Asia	High Population 119587.0	High intersection density (63.52)	Palembang, Indonesia
		Low intersection density	Santa Cruz, Philippines
	Low Population	High intersection density (55.63)	Kamphaeng Phet, Thailand
		Low intersection density	Bagansiapiapi, Indonesia
Sub Sahara America	High Population 119100.0	High intersection density (63.96)	De Rust Caravan Park, South Africa
		Low intersection density	Ore, Nigeria
	Low Population	High intersection density (48.07)	Fatick, Senegal
		Low intersection density	Vila Nova do Seles, Angola
Western Asia and North Africa	High Population 113009.5	High intersection density (102.53)	M'Sila, Algeria

		Low intersection density	Dubai, United Arab Emirates
	Low Population	High intersection density (94.39)	Beja, Tunisia
		Low intersection density	Chelghoum Laid, Algeria

Stratified random selection of nodes

The street network model of each city contains three distinct types of nodes. The first category is 'not consolidated', which indicates that this node is not consolidated with other nodes by either of the two parameters – the matched road width and the median road width of the world (8.0m). The second category is 'one consolidated' meaning the node is consolidated by either matched road width or the worldwide median road width, whichever is larger. The third category is 'both consolidated', meaning the node is consolidated by both the matched road width and the worldwide median road width.

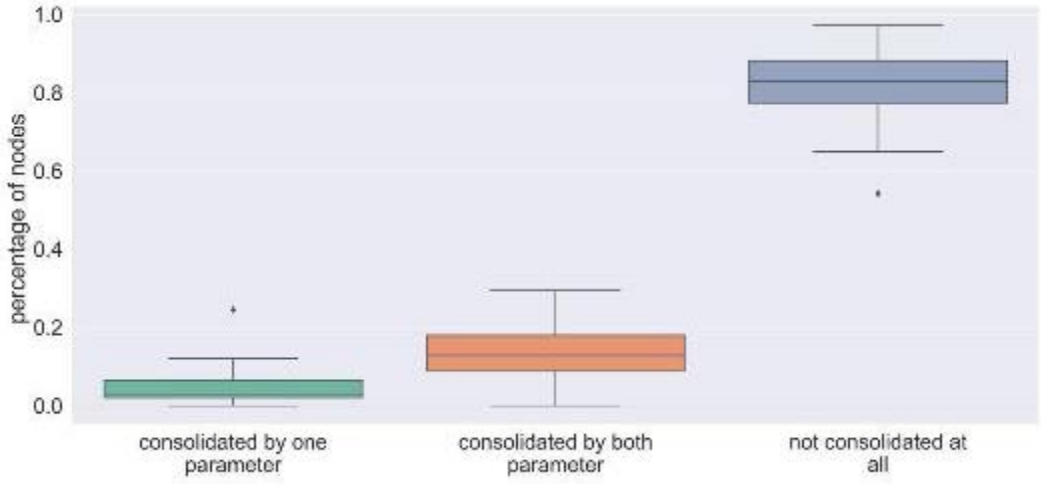


Figure 11. Proportion of each kind of node of the sampled cities

As Figure 11 shows, the median percentage of one-consolidated nodes, both-consolidated nodes, and not-consolidated nodes in sampled cities is about 3%, 13%, and 83%, respectively. If we sample each type of node proportionally, we may obtain an extremely small sample of nodes in the categories of one consolidated and both consolidated. In order to have a more balanced sample size for each category, we instead to randomly select 30, 30, and 60 nodes from each category for validation. If there are fewer nodes in a category, we will sample all of

them. Table 4 displays the number of sampled nodes in each category across all sampled cities, as well as their geospatially matched cities and corresponding road width.

Table 4. Stratified random selection of nodes

Sampled city	Matched road width (m)	Matched city	Sampled Not consolidated	Sampled One-consolidated	Sampled Both-consolidated
East Asia and Pacific					
Hohhot, China	8.5	Beijing, China	60	30	30
Jiangyuan, China	6.5	Pyongyang, North Korea	60	6	6
Quangang, China	6.5	Taipei, Taiwan (China)	60	21	29
Pyeongtaek Si, South Korea	6.8	Cheonan, South Korea	60	30	30
Europe and Japan					
Liege, Belgium	7.6	Antwerp, BEL	60	30	30
Pisa, Italy	6.1	Milan, Italy	60	30	30
Uman, Ukraine	8.0	Nikolaev, UKR	60	0	30
Belfast, United Kingdom	7.3	Manchester, GBR	60	30	30
Land rich developed countries					
Brisbane, Australia	14.6	Sydney, AUS	60	30	30
Sherbrooke, Canada	10.6	Montreal, CAN	60	30	30
Abington, United States	15.3	Philadelphia , USA	60	30	30
Boulder, United States	14.5	Killeen, USA	60	30	30
Latin America and the Caribbean					
Chiquimula,Guatemala	7.8	Guatemala City, GTM	60	3	30
Huajuapan, Mexico	11.4	Mexico City, MEX	60	30	30

Sampled city	Matched road width (m)	Matched city	Sampled Not consolidated	Sampled One-consolidated	Sampled Both-consolidated
Juliaca, Peru	9.5	Cochabamba, Bolivia	60	30	30
Port of Spain, Trinidad and Tobago	10.6	Caracas, Venezuela	60	30	30
South and Central Asia					
Chakaria, Bangladesh	5.5	Dhaka, Bangladesh	60	4	0
Chilakaluripet, India	6.6	Vijayawada, India	60	30	30
Jaspur, India	5.3	Sitapur, India	60	2	4
Kokshetau, Kazakhstan	8.0	Shymkent, Kazakhstan	60	0	30
Southeast Asia					
Bagansiapiapi, Indonesia	9.1	Rawang, Malaysia	60	10	30
Palembang, Indonesia	5.0	Palembang, Indonesia	60	30	30
Santa Cruz, Philippines	7.8	Manila, Philippines	60	4	30
Kamphaeng Phet, Thailand	8.0	Bangkok, Thailand	60	0	30
Sub-Sahara America					
Vila Nova do Seles, Angola	7.0	Luanda, Angola	60	4	19
Ore, Nigeria	4.8	Ibadan, Nigeria	60	30	14
Fatick, Senegal	7.0	Bamako, Mali	60	30	30
De Rust Caravan Park, South Africa	11.9	Johannesburg, South Africa	60	30	30
Western Asia and North Africa					
Chelghoum Laid, Algeria	7.0	Tebessa, DZA	60	27	30
M'Sila, Algeria	7.8	Algiers, DZA	60	16	30

Sampled city	Matched road width (m)	Matched city	Sampled Not consolidated	Sampled One-consolidated	Sampled Both-consolidated
Beja, Tunisia	7.1	Kairouan, TUN	60	30	30
Dubai, United Arab Emirates	15.9	Riyadh, Saudi Arabia	60	30	30

Qualitative validation of consolidation results

We validated the result of consolidation by cross-referencing with Google Earth and satellite imagery to determine qualitatively whether the consolidation is correct or not. There are four categories of validation result: true positive (the node is consolidated and should be consolidated); true negative (the node is not consolidated and should not be consolidated), false positive (the node is consolidated but should not be consolidated); and false negative (the node is not consolidated but should be consolidated). Within the true positive category, we created a subcategory called "true positive but not robust enough." For example, ideally we want five nodes in a roundabout to be consolidated together, but only four of the five nodes in this roundabout are consolidated together. This case belongs to the subcategory of true positive but not robust enough: in other words, the result is mostly right but not perfect.

3.3. Resilience Simulation

The methods can be explained in four parts: 1) measuring trip circuity of unperturbed networks, 2) perturbing the network, 3) estimating resilience of perturbed networks, and 4) identifying the determinants of street network resilience. For the resilience simulation, we first selected cities with at least 100 nodes in their original network leading us to remove 902 cities, and we had to additionally remove three small towns (Coloane, Gimmeizet Bilgai, Sithana) where the strongly connected networks from the original and consolidated networks were different. This resulted in strongly connected networks for 8005 cities out of 8910.

3.3.1 Measuring trip circuity of unperturbed networks

3.3.1.1 Creating base origin-destination pairs

Using the strongly connected component, we first generated 10,000 unique random OD (origin-destination) pairs per street network. We performed a number of sensitivity analyses and decided on the number of OD pairs after considering the level of precision and the amount of time of the computational calculations. For instance, even in the case of Amsterdam, a water-

based city relying extensively on bridges where the street network grows dramatically disconnected when removing nodes, the results for circuity had a narrow confidence interval.

For each city, we used three types of street networks: 1) original network from the Global Urban Street Networks GraphML dataset (Boeing, 2021); 2) consolidated network (tolerance = matched road width); 3) consolidated network (tolerance = world median road width).

3.3.1.2 Measuring trip circuity

By using the OD pairs generated for each city and network type, we measured the Euclidean distance and the shortest network path distance using the OSMnx package. We first measured circuity for each OD pair by dividing the network distance by Euclidean distance. Here, a larger value of circuity indicates that the network is more circuitous, in which the shortest path is less straight. Then, we calculated the average of the circuity values of the 10,000 pairs to measure circuity at the network level.

3.3.2 Perturbing the network

3.3.2.1 Generating node attributes

We perturbed the network using three approaches: 1) perturb nodes with high betweenness centrality, simulating the attacks on “important” intersections; 2) perturb nodes with low elevation, simulating perturbation due to flooding disasters; and 3) perturb nodes randomly, simulating random disruptions. Before we removed the nodes and disrupted the network, we created a table with node attributes. For each city, we generated three outputs according to network type, in which the output includes the latitude, longitude, betweenness centrality, and elevation.

Betweenness centrality is an indicator of centrality based on shortest paths; for each node, we compute the number of shortest paths of all network node pairs that pass through the node. We use the length of each edge as the weight, in which the shortest path is identified based on the network distance between nodes. We report the normalized value of betweenness centrality per city in our output table.

There are at least three packages that can be utilized to estimate betweenness centrality using the Python language: networkx, igraph, and graph-tool. These packages have tradeoffs between how easy it is to install (e.g., compilation requirements) and how fast it runs. For instance, the networkx does not require compilation, while the computing speed is significantly slower than others. On the other hand, the performance of the graph-tool package is the fastest; however, the disadvantage of using graph-tool is that it is difficult to run on Windows. In detail, installing graph-tool on Windows requires using Docker or the Ubuntu userspace. Thus, we used the igraph tool, which is easy to install regardless of the user’s platform and

shows good performance in terms of speed. There are also other packages, such as cugraph that provide fast betweenness centrality calculation via GPUs; however, cugraph does not support weights for each edge, which is inappropriate for our research project.

We also considered elevation. The Global Urban Street Networks dataset includes three elevation variables derived from different sources. We use the 'elevation' variable, which represents the meters above sea level from the ASTER or SRTM digital elevation model (DEM). One issue with the node elevation is that the elevation of consolidated nodes has a missing value. While it is possible to collect the elevation value from DEM sources, we simply calculated the average elevation of nodes that merged to form the consolidated node.

3.3.2.2 Measuring trip distance with network perturbation

Using the base OD pairs generated from the previous section, we re-estimated the network distance of OD pairs as we perturbed the network. We perturbed the network in three ways: 1) remove nodes that have high betweenness centrality values, 2) remove nodes with low elevation values, and 3) remove nodes randomly. We tested the results by removing 1 to 10 percent of the total nodes per city. For the random perturbation, we generated a random number for each node to ensure that our results were replicable. There were cases with small variations in node elevation; in these cases, we randomly selected the nodes to remove among the nodes that have equal elevation.

3.3.3 Estimating resilience of perturbed networks

We focus on two attributes to understand the resilience of street networks: 1) how quickly networks get disconnected, and 2) how network perturbation changes the trip circuity.

3.3.3.1 Percentage of solvable OD pairs

With network perturbation, a portion of the base OD pairs gets unsolvable, meaning that the origin could not reach the destination, mainly due to two reasons. First, either the origin or destination of the OD pair is removed from the network due to perturbation. In this case, we could have changed the removed origin (or destination) node to its closest part of the network; however, the nodes removed either with high betweenness centrality or low elevation are spatially clustered, in which we identified these cases as 'removed OD pairs'. For instance, if we remove 10% of the nodes in the network, it is likely that around 20% of the OD pairs are removed. Second, the network gets disconnected and the network distance of the OD pair is no longer solvable. A high percentage of unsolvable OD pairs signal that the network is vulnerable to perturbation. In Amsterdam, for instance, removing nodes with high betweenness centrality leads to a dramatic increase in the percentage of unsolvable OD pairs as nodes along the bridges connecting the city get removed.

3.3.3.2 Marginal and cumulative changes in trip distance

Estimating the changes in the trip distance is tricky. In theory, the circuity of OD pairs should increase as we remove nodes from the network. However, the average circuity value does not necessarily increase as there are OD pairs that got either unsolved or removed. To elaborate, as the street networks get perturbed, we are not necessarily comparing the circuity of a same set of OD pairs before and after perturbation. Thus, instead of estimating the changes in average circuity values from slightly different universes, we calculate the marginal changes in network distance within the same set of OD pairs.

For the marginal change in network distance between 'n% removed' and '(n+1)% removed' networks, we only focus on the OD pairs that are solved in both networks, thus using the same comparison universe. We calculate the average increase in network distance between the two networks. We then obtain the coefficient of the marginal change: for example, if the network distance increases by 10% when we remove 5% of the nodes compared to when we remove 4% of the nodes, the coefficient would be 1.10. Because we perturbed the network by removing 1 to 10% of the nodes, we obtained ten values of marginal changes. The cumulative changes are estimated by calculating the product of the ten values of marginal changes.

One limitation of this approach is that the cumulative change is sensitive to the number of marginal changes we estimate within a range. For instance, if we only estimate the marginal changes by removing 5% and 10% of the nodes, we obtain two values for the marginal changes. The first value would be dependent on the solved OD pairs after removing 5% of the nodes, and the second value would be dependent on the solved OD pairs after removing 10% of the nodes. If the base OD pairs were to be dramatically unsolvable with network perturbation, the marginal changes might underestimate the actual change. In other words, if we estimate the marginal changes every time we remove 0.1% of the nodes, we would get a larger number of marginal changes that would result in greater cumulative changes. Due to limited computing time and cost, we estimated the marginal changes for every 1% of node perturbation.

Another possible way to measure resilience is to only focus on OD pairs that are solvable after perturbing 10% of the nodes. In this approach, the average circuity value of OD pairs would increase as we sequentially perturb the network. However, one critical limitation of this approach is that we could be only addressing the OD pairs that are irrelevant to network perturbation; for instance, we would only use the OD pairs that are located within suburban areas or are too close to each other to be affected by the perturbation, in which the changes in circuity values may be biased and underestimated.

3.3.4 Identifying the determinants of street network resilience

Next, we estimate the relationships between street network features and street network resilience indicators. We have three dependent variables; (1) the percentage of solved OD pairs

after perturbing 10% of the network nodes, (2) the change in circuity of trip distances for solved OD pairs after perturbing 5% of the network nodes, and (3) change in circuity of trip distances for solved OD pairs after perturbing 10% of the network nodes (Table 5). For each type, we have three different measures based on the perturbation method (i.e., perturbation by betweenness centrality, elevation, and random). Accordingly, we present the results of nine simulations.

We utilize several indicators for street network design, in which we calculated the variables by using the OSMnx package. We use the strongly connected part of the original network to calculate variables such as the average node degree, the average circuity of streets, intersection density, and the total length per built-up area. To add, we use the standard deviation of node elevations as a proxy for the hilliness of the urban area. We also calculated the number of standard deviations between the maximum and mean value of betweenness centrality, which is used as a proxy for the presence of particularly important nodes within each network.

There are multiple variables used as statistical controls; percentage of open space, size of built-up area, population density, and the world region dummies. See Table 5 for the details.

Table 5. Description and sources of variables

Variable	Description	Source
Dependent variables		
% of solved OD pairs	The percentage of solved OD pairs after perturbing 10% of the network nodes by BC, elevation, and random	OpenStreetMap
Change in circuity of solved OD pairs	Change in circuity of trip distances for solved OD pairs after perturbing 5% or 10% of the network nodes by BC, elevation, and random	OpenStreetMap
Independent variables		
k average	Average node degree	OpenStreetMap
Circuity	Circuity of the network	OpenStreetMap
Intersection density	Number of intersections per square kilometer	OpenStreetMap
Elevation (std)	Standard deviation of node elevations	ASTER; SRTM
BC (max-mean)/std	Number of standard deviations between the maximum and mean value of betweenness centrality	OpenStreetMap
Length total per area	Total street segment length per square kilometer	OpenStreetMap
Open space	Percentage of open space	GHSL UCD
Built up area	Built up surface area (square kilometer)	GHSL UCD

Variable	Description	Source
Population density	Total residential population divided by built up area	GHSL UCD
World region	Major geographical region; Africa (reference), Asia, Europe, Latin America and the Caribbean (LAC), Oceania	GHSL UCD

Note: Variables are calculated by using the strongly connected part of the original networks

We estimate nine models using the ordinary least squares (OLS) approach. We checked the variance inflation factors (VIF) to avoid multicollinearity issues. We also present the beta coefficients to understand the effect size of each independent variable.

4. Results

4.1. Intersection Consolidation Results

4.1.1 Consolidation Result

Of all the 8,910 cities, the median percentage of nodes that remained after matched road width and world median road width consolidation is 92.67% and 92.00%, respectively. Figure 12 shows the percentage of remaining intersections by world region. For matched road width consolidation, the South and Central Asia region has the highest median percentage of nodes remained (95.79%), while Western Asia and North Africa has the lowest (86.89%). For the global parameter, world median road width consolidation (8.0m), the Land Rich Developed Countries region has the highest median percentage of nodes remained (94.43%), while the East Asian and the Pacific has the lowest (88.99%).

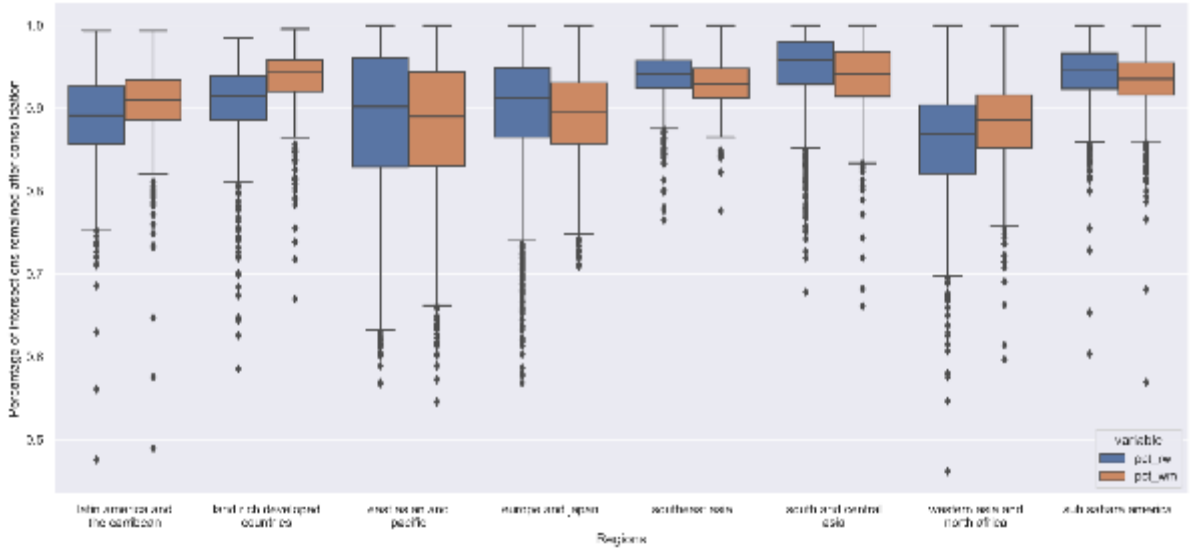


Figure 12. Percentage of intersection remained after matched road width (pct_rw) world median road width (8.7m, pct_wm) consolidation by world region

4.1.2 Validation Result

Global validation result

Over 95% of the time the algorithm is correctly consolidating the redundant nodes that should be practically treated as one (Table 6). The overall validation result reveals a true positive rate (the percentage of the consolidated nodes that are correctly consolidated) of 97.59% for matched road width consolidated nodes and 97.02% for the global parameter world median road width consolidated nodes. In addition, for unconsolidated nodes, the result shows a true negative rate (the percentage of unconsolidated nodes that shouldn't be consolidated) of about 90% for both parameterizations (Table 7).

Table 6. Validation result for sampled consolidated nodes

Validation Category	Consolidated by Matched road width	Consolidated by world median road width (8.0m)
True positive (including true positive and not robust enough)	97.59%	97.02%
False positive	2.41%	2.98%
True positive but not robust enough	5.77%	8.44%

**The percentage is out of 1,162 sampled nodes consolidated by matched road width, and 1,209 sampled nodes consolidated by world median road width (8.0m), respectively.*

Table 7. Validation result for sampled not consolidated nodes

Validation Category	Consolidated by Matched road width	Consolidated by world median road width (8.0m)
True negative	87.00%	89.33%
False negative	13.00%	10.67%

**The percentage is out of 2,277 sampled nodes consolidated by matched road width, and 2,230 sampled nodes consolidated by world median road width (8.0m), respectively.*

Validation result by world region

Table 8 shows the validation results by world regions. By both parameterizations, the true positive rates of all regions are above 90%. It indicates that the algorithm seldom falsely consolidates distinct intersections. As for unconsolidated nodes, validated cities in Land-Rich Developed Countries, Latin America and the Caribbean, South and Central Asia, Southeast Asia,

Sub-Saharan Africa, have at least one parameter with a true negative rate greater than 90%. In Land Rich Developed Countries and Latin America and the Caribbean, matched road width outperforms the global median road width in terms of correctly consolidating redundant nodes, whereas the pattern is inverted in Europe and Japan. However, for the regions of East Asia and the Pacific, Europe and Japan, and Western Asia and North Africa, neither of the two parameters has a true negative rate greater than 90%, indicating that the parameters are insufficiently robust.

Table 8. Validation result by world region

Region	Consolidated by matched road width – true positive rate	Consolidated by world median road width – true positive rate	Not consolidated by matched road width – true negative rate	Not consolidated by world median road width – true negative rate
East Asia and the Pacific	100.00%	100.00%	70.37%	75.93%
Europe and Japan	98.33%	98.10%	75.45%	87.50%
Land-Rich Developed Countries	97.50%	100%	98.33%	90.28%
Latin America and the Caribbean	94.76%	99.18%	95.88%	84.55%
South and Central Asia	100.00%	91.00%	95.02%	97.08%
Southeast Asia	97.69%	92.21%	96.72%	99.20%
Sub-Saharan Africa	99.19%	98.09%	88.49%	96.30%
Western Asia and North Africa	96.67%	96.37%	80.51%	85.93%

**The true positive rate include cases of true positive but not robust enough*

Land rich developed countries

In land-rich developed countries, the consolidation algorithm effectively corrected the street networks, particularly using the matched street width parameter. The urban texture of land-rich nations is characterized by grid-like street networks lined with single-family homes. The

majority of false positive cases in land-rich developed nations are defined by qualitative intersection judgments made by different researchers. For instance, in figure 13, it may appear to one researcher as one intersection and to another as two intersections (a) referencing the Google Earth view (b).

(a) Intersection consolidation (b) Google Earth View

Figure 13. Consolidation in Abington, United States

(red: original nodes; green dots: matched road width consolidated nodes, blue dots: world median road width consolidated nodes. The layer of consolidated nodes by the larger parameter is on the top, and the original nodes are at the bottom layer. This rule also applies to the following figures)

East Asia and the Pacific

The algorithm performs less effectively in East Asian and Pacific cities compared with cities in Land-Rich Countries. Three of the four sampled cities are in China, and their blocks are relatively large (around 300 to 500 m) and separated by large residential communities. As road density decreases, planners have constructed wide arterial roads (approximately 30m) to accommodate the traffic, resulting in numerous large intersections. China's major intersections in Hohhot and Quangang were not effectively consolidated by either of the two parameters (Figure 14 a & b). These extensive intersections need more aggressive parameters to be consolidated together. We didn't detect any false positive cases in the sampled nodes in East Asia and the Pacific region. However, an overly aggressive parameter may produce false positive cases in a specific urban texture in Chinese cities, namely dense urban villages and industrial zones. As for Pyeongtaek Si, besides its similar urban layout to Chinese cities, it has numerous large five-way intersections that are hard for the algorithm to completely consolidate.

(a) Hohhot, China (b) Quangang, China

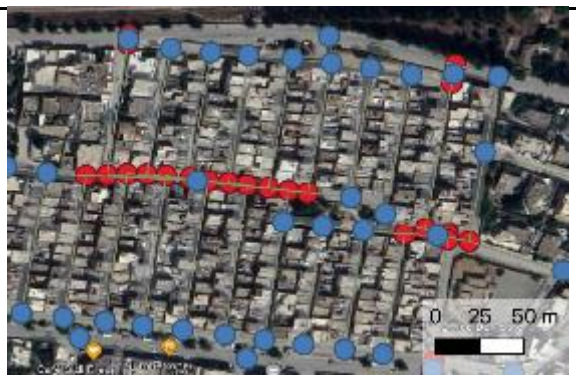
Figure 14. Large 2-by-2 divided roadway intersections in China

Europe and Japan

For street networks in Europe and Japan, the median world width parameter performs better than the matched road width parameter, as it is more aggressive. In European cities, block sizes vary, but are large enough to prevent false positive cases. Most 2-by-2 divided roadway intersections were effectively consolidated by the algorithm, but large roundabout intersections, triangle islands, and slip planes were not.

Western Asia and North Africa

Except for Dubai, a large and populous metropolitan area in the United Arab Emirates, cities sampled in western Asia and North Africa have mostly narrow local roads with one or two lanes per direction (10 meters wide). There are few wide roads or highways, and blocks are relatively small (around 30 x 60 m). Although the street network we used for consolidation is drivable, drivable roads in these cities may be defined differently as drivable for bikes or motorcycles. Figure 15 (a & b) shows a false case in Tunisia, Beja, where narrow roads are erroneously consolidated together.



(a) False intersection consolidation of narrow streets

(b) Corresponding Google Earth View

Figure 15. False positive case in Beja, Tunisia

Latin America and the Caribbean

The true positive rates of matched road width and world median road width consolidation, as well as the true negative rate of matched road width consolidation, are all greater than 90%. The algorithm correctly simplified the street network in Latin America and the Caribbean, which is predominantly grid-like with roadways circling around 70m*100m building blocks or winding streets around mountains. Contrary to the European street network, which is characterized by roundabouts, the street network in Latin America and the Caribbean has few roundabouts but numerous divided roadways on arterial roads. Therefore, the algorithm effectively corrected redundant nodes in divided roadways (Figure 16 a & b).

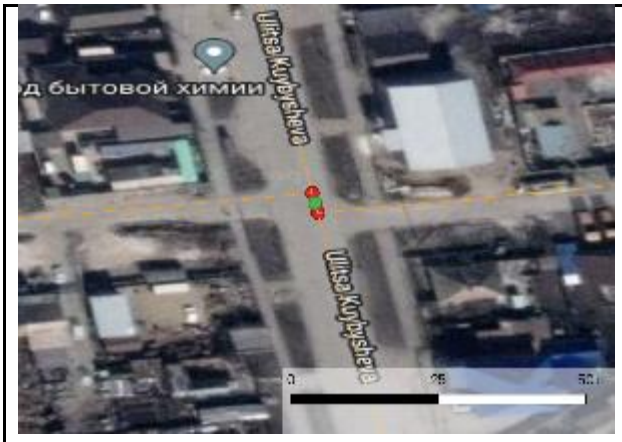
(a) Fresnos Road in Huajuapan, Mexico

(b) Western Main Road in Port of Spain, Trinidad and Tobago

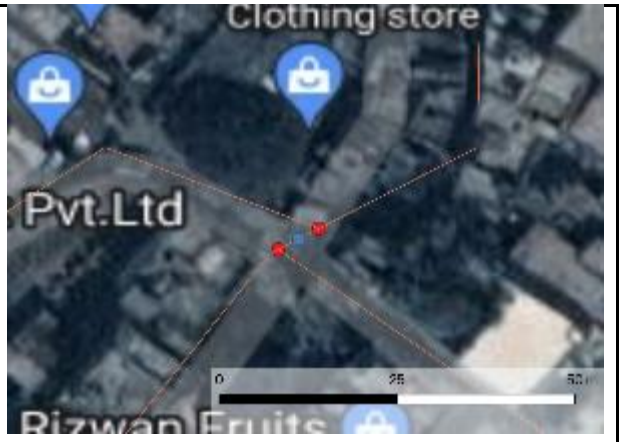
Figure 16. Consolidated divided roadways in cities in Latin America and the Caribbean

South and Central Asia

With a true positive and true negative rate greater than 90% for both parameters, the algorithm performed well in South and Central Asian cities. Two cities in the high intersection density group (Kokshetau, Kazakhstan, and Chilakaluripet, India) are grid-based. The grids in Kokshetau are approximately 120m* 120m, whereas the grids in Chilakaluripet are approximately 30m*120m. Two cities in the low intersection density group (Chakaria in Bangladesh and Jaspur in India) have circuitous and low-density street networks which need minimal correction. The algorithm consolidated the cases with redundant nodes, most likely, the face-to-face staggered intersections, in both cities with high (Figure 17, a) and low intersection density (Figure 17, b).



(a) Kokshetau, Kazakhstan

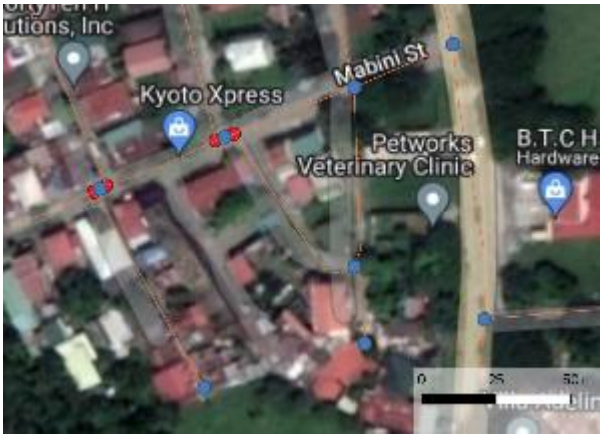


(b) Jaspur, India

Figure 17. consolidated nearby face-to-face staggered intersection in Jaspur, India (left), and in South and Central Asia

Southeast Asia

Both the matched and global median street widths effectively corrected Southeast Asian street networks. The sampled cities display a pattern of mixed high-density built-up area with large but narrow blocks and farmland, as well as non-gridded street networks. The algorithm merged redundant nodes from divided highways and nearby face-to-face intersections that are practically considered as one in the real world (Figure 18)



(a) Santa Cruz, Philippines

(b) Palembang, Indonesia

Figure 18. Consolidated redundant face-to-face nearby intersections in Southeast Asia

Sub Sahara America

Urban settlements in Sub Sahara America are crowded with low storage buildings. The street networks mostly consist of narrow (around 8m) local streets, with one or two major roads, such as highways (around 12m), passing through the city. The algorithm successfully consolidated the redundant nodes caused by the intersections between divided major roads and local streets (Figure 19, a), and multiple lane merging (Figure 19, b).

(a) Consolidated divided roadway in De Rust Caravan Park, South Africa

(b) Consolidated multiple lane merged intersections in Fatick, Senega

Figure 19. Consolidated intersections in Sub Sahara America

Validation result by intersection type

Divided roadways

This category of nodes should be merged because they belong to the same intersection, whereas in the original street network, two by two roadways often produced a four-node intersection. The consolidation algorithm consolidated redundant nodes (Figure 20 a & b).

(a) Hohhot, China

(b) Boulder, United States

Figure 20. Successful consolidation of redundant nodes resulted from divided roadways

Face-to-face nearby intersections

In the real world, two streets crossing can be considered one intersection, while in digitized street networks, it sometimes yields two face-to-face nearby nodes. The algorithm consolidates these face-to-face nearby intersections and simplifies the street network to be closer to the real life representation (Figure 21 a & b).

(a) Fatick, Senegal

(b) Ore, Nigeria

Figure 21. Consolidation of face-to-face nearby (staggered) intersection

Sliplanes

Despite the successful consolidation of the divided roadways, it is unlikely that the sliplanes at the intersection will be fully consolidated. Although it is a part of the intersection, it is located far from the actual intersection and is difficult to consolidate (Figure 22).

(a) Brisbane, Australia (b) Kamphaeng Phet, Thailand

Figure 22. Inadequate consolidation of Intersections with slip lane

Roundabout

The algorithm is capable of consolidating nodes in a three- to four-way roundabout. As large roundabouts have multiple entry lanes and slip lanes, it is difficult for the algorithm to completely consolidate all the nodes (Figure 23 a & b).

(a) Pisa, Italy (b) Dubai, Emirates

Figure 23. Consolidations in roundabouts

4.1.3 Intersection Density and Node Centrality

Intersection density

Figure 24 shows the boxplot distribution of intersection density by region and network type. The overall intersection density has decreased in the consolidated network as it combines redundant nodes; the median intersection density for all studied cities dropped from 150.8 to 136.2 nodes per square kilometer. The median intersection density is the lowest in Northern American cities, reflecting their grid-like, low-density, and extensive street networks. On the other hand, the highest median intersection density was observed in Africa and Latin America and the Caribbean (LAC) regions. The intersection density in almost half of the cities located in these two regions were greater than the upper quartile of all cities. It is also noteworthy that the variation in intersection density across cities was the largest in Asia and Africa; in contrast, it was relatively low for Oceania and Northern America.

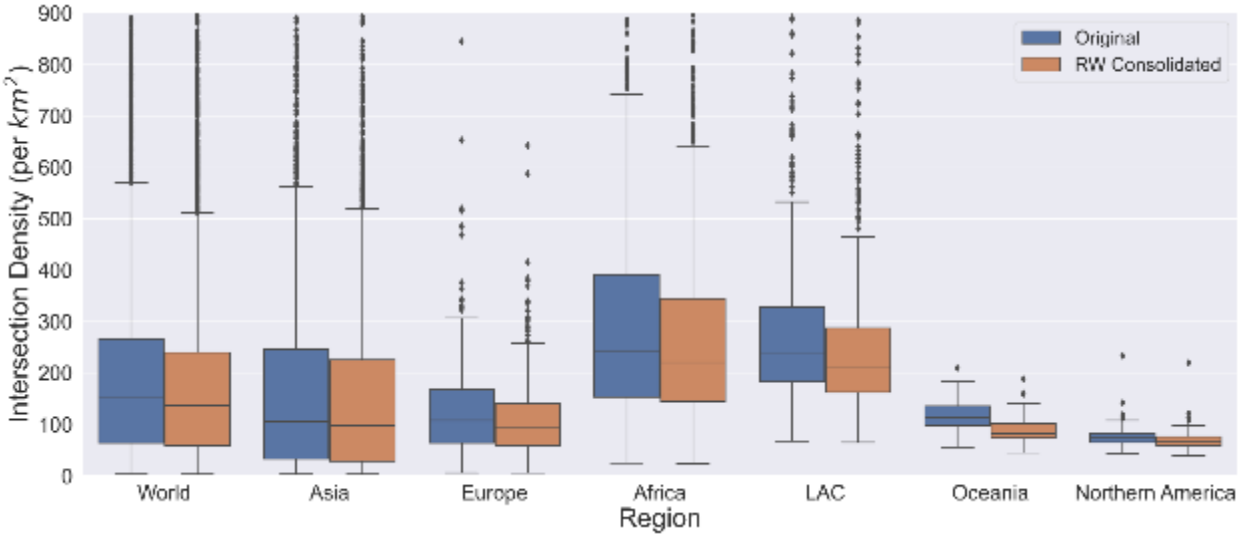


Figure 24. Boxplot of intersection density by region and network type
(Note: Cities with intersection density higher than 900 per km2 are not shown in the figure)

Considering the change in intersection density due to network consolidation, we may overestimate intersection density by around 5 to 15% (Figure 25). If we use the original street networks, we may overestimate the distribution of intersections which could be problematic for intra- and inter-city research. Cities with the most significant decrease in intersection density are most likely to be cities with small urban areas and few nodes; in small urban areas, a small number of node consolidations could result in a significant proportional change in intersections.

Among the six world regions, Oceania showed the most significant decrease in intersection density; for instance, the results for major cities in Oceania show that we need to address intersection density with caution (e.g., Sydney (77.6%); Melbourne (69.9%); Adelaide (72.0%); Perth (62.5%). On the other hand, the change in intersection density was less than 10% for most cities in Northern America. Among the cities with a large population, cities such as Tokyo

(93.2%), Dhaka (95.4%), and Seoul (93.4) showed less change, whereas the change was more remarkable in cities such as Cairo (70.6%), Tehran (77.6%), and Singapore (77.9%).

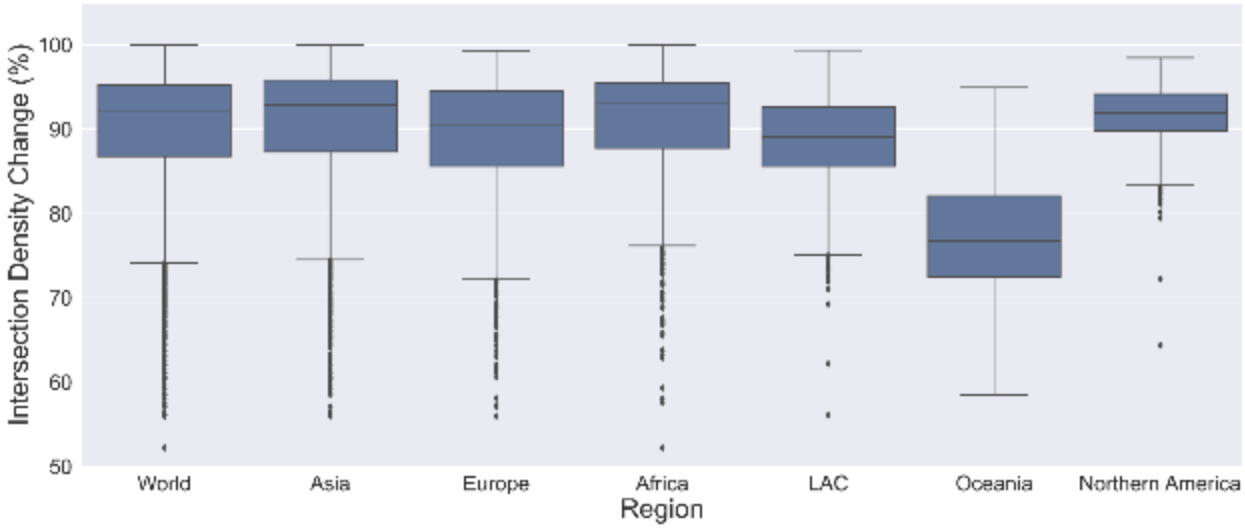


Figure 25. Boxplot of the percentage of intersections remained in the consolidated network (Note: Cities having less than 50% of the intersections remained after network consolidation are not shown in the figure)

Comparison of node betweenness centrality

We conducted a correlation analysis between the node betweenness centrality values calculated from the original and consolidated network. Here, we estimated the correlation coefficients by using two methods; Pearson and Spearman. While the betweenness centrality value provides significant implications, the rank order of the values is also important as they signal the spatial distribution of nodes with higher centrality. In addition, higher betweenness implies a higher possibility of passing traffic, further having an impact on the location of retail shops and rent. In this perspective, we argue that having a precise estimate of betweenness centrality is important for understanding the form of street networks. From our analysis in this section, we describe how betweenness centrality values can have bias without combining the redundant nodes with network consolidation.

The median coefficient of the Pearson correlation results is 0.96, mostly ranging from 0.90 to 1.00 (see Figure 26). For the Spearman correlation coefficients, the median is 0.98, which suggests that the change in betweenness centrality after network consolidation is not large. Among the six regions, the median coefficient of the Spearman correlation analysis was the smallest in LAC and Oceania. This result suggests that the interpretation of the street network may change by using the consolidated network; for instance, a significant node identified from the original network may be less significant in the consolidated network.

The results also suggest that cities where intersection density significantly changed due to consolidation may show gaps in betweenness centrality. For instance, the Spearman correlation coefficients in cities located in Oceania were relatively low (e.g., Melbourne (0.92), Adelaide (0.92), and Perth (0.91)). It should be noted that the interpretation of the results in this section is made based on the assumption that our consolidation algorithm is correct and valid.

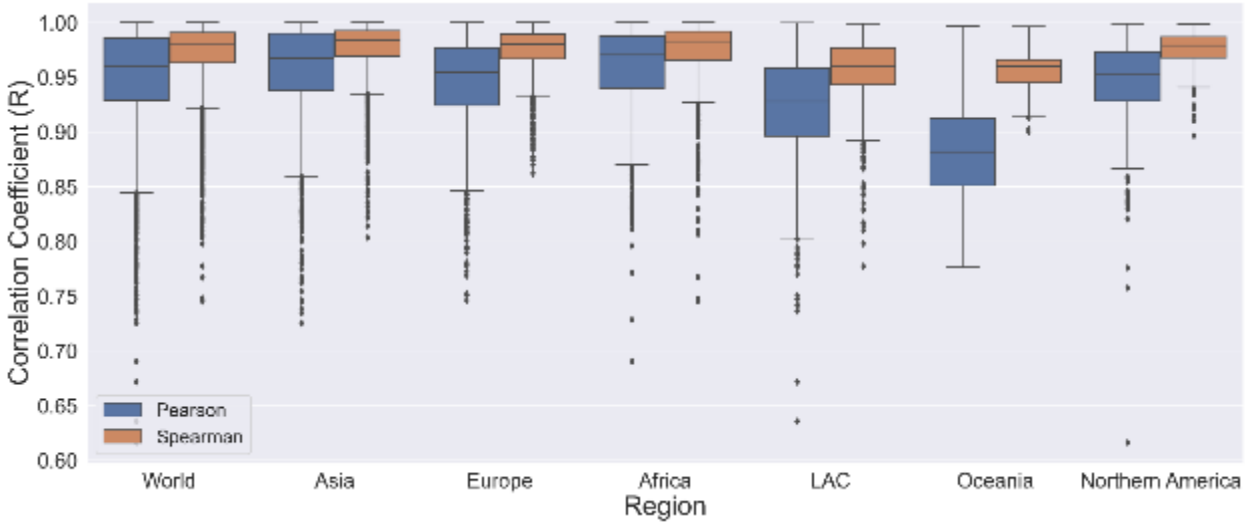


Figure 26. Correlation coefficients of betweenness centrality between the original and the consolidated street network (Note: Cities showing coefficients less than 0.60 are omitted)

4.2. Resilience Simulation Results

4.2.1 Trip Circuity

Figure 27 shows the boxplot distribution of trip-level circuity of the originally sampled 10,000 OD pairs by region and network type. After consolidation, the trip-level circuity slightly increased, but the result is practically the same. As we reconnected the consolidated nodes, there will be a trivial difference in the edge length, contributing to the circuity difference. In addition, the characteristics of OD pairs created for each city and network type may differ, leading to gaps between network types.

As shown in Figure 27, the median is around 1.4, which indicates that the network trip distance is around 40% longer than the Euclidean distance. The median value was low in Africa and Northern America, whereas regions such as Europe and Oceania showed slightly greater values. Among populated cities, Hong Kong (original network=1.65; consolidated network=1.69) and Mumbai (original network=1.46; consolidated network=1.48) showed higher trip-level circuity.

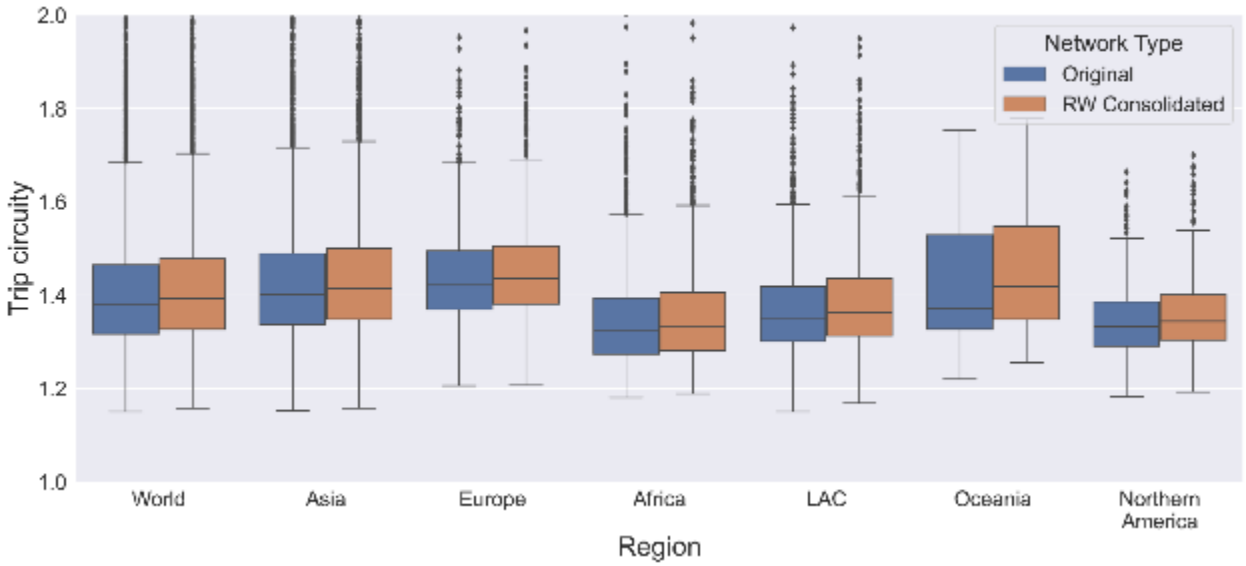


Figure 27. Trip-level circuity by region and network type (Note: Cities with trip-level circuity higher than 2.0 are not shown in the figure)

4.2.2 Network Perturbation

Here, we describe the results of network perturbation. We perturb the network in three ways; (1) attack nodes with high betweenness centrality, (2) attack nodes with low elevation, and (3) attack nodes randomly. We remove the nodes from 1% to 10% of the total nodes per city. We limited our network perturbation up to 10% since attacking more than 10% of the nodes would result in significantly unconnected networks. On the other hand, we have multiple indicators to understand the impact of network perturbation.

First, we examine the percentage of solved OD pairs. If there are fewer solved OD pairs as we perturb the street network, it indicates that the street network gets disconnected. Here, it should be noted that around 20% of the OD pairs should be disconnected as we remove 10% of the nodes from the network; if we remove 10% of the total nodes, we should lose 10% of the nodes designated as origins and another 10% of the nodes designated as destinations in the OD pairs.

Second, we calculate the marginal change in trip distances as we remove nodes. For instance, the marginal change between the cases when we attack 7% and 8% of the nodes is calculated based on the solvable OD pairs after perturbing 8% of the nodes. To add, we calculate the cumulative change based on the product of the ten numerical values of marginal changes.

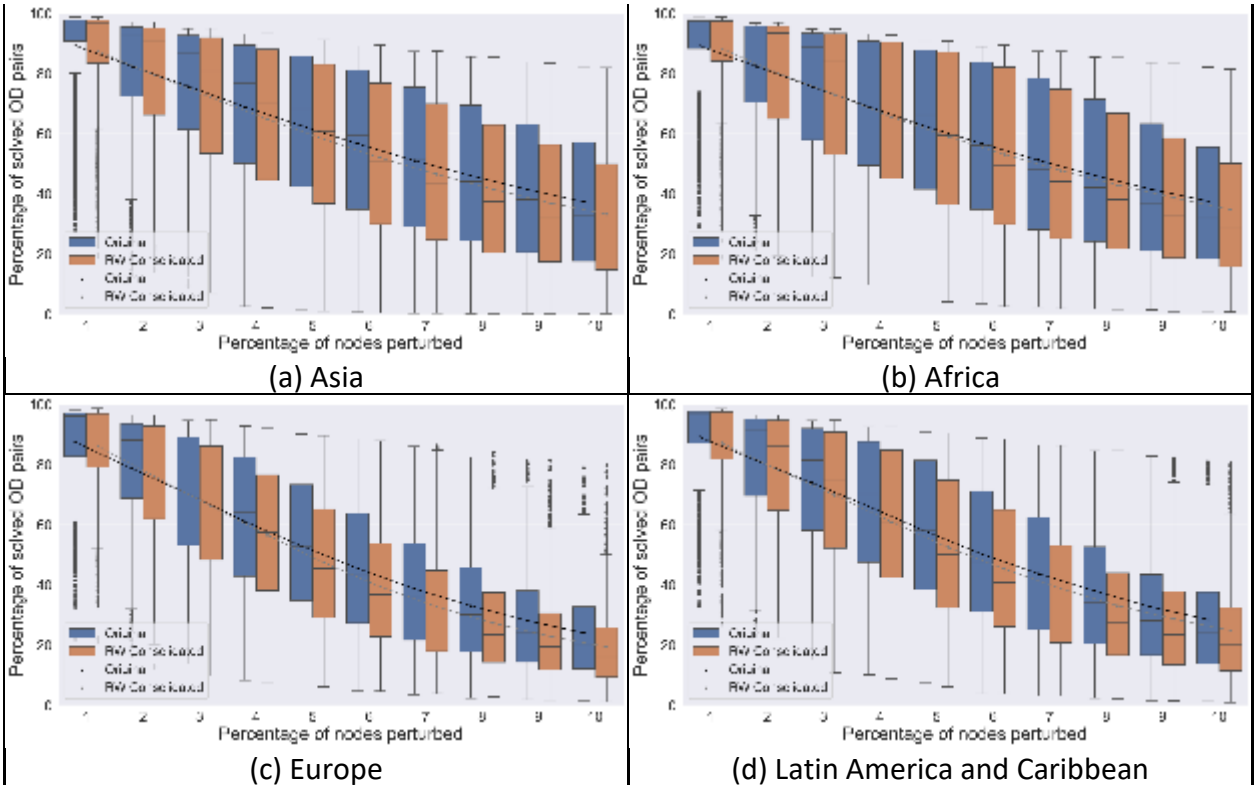
Lastly, we calculate the change in trip circuity when we remove 5% or 10% of the nodes by using the solved OD pairs. For instance, if we have only 40% of the OD pairs solved after removing 5% of the nodes, we compute how trip-level circuity have changed for these solved

OD pairs. This approach is designed to overcome the limitations of estimating cumulative changes, where we cannot address the circuity changes for unsolved OD pairs. However, it should be noted that some cities have a minimal number of OD pairs to compare after attacking 10% of the nodes. To add, the cumulative change might be overestimated as it accounts for the changes in trip circuity for disconnected OD pairs.

Perturb nodes with high betweenness centrality

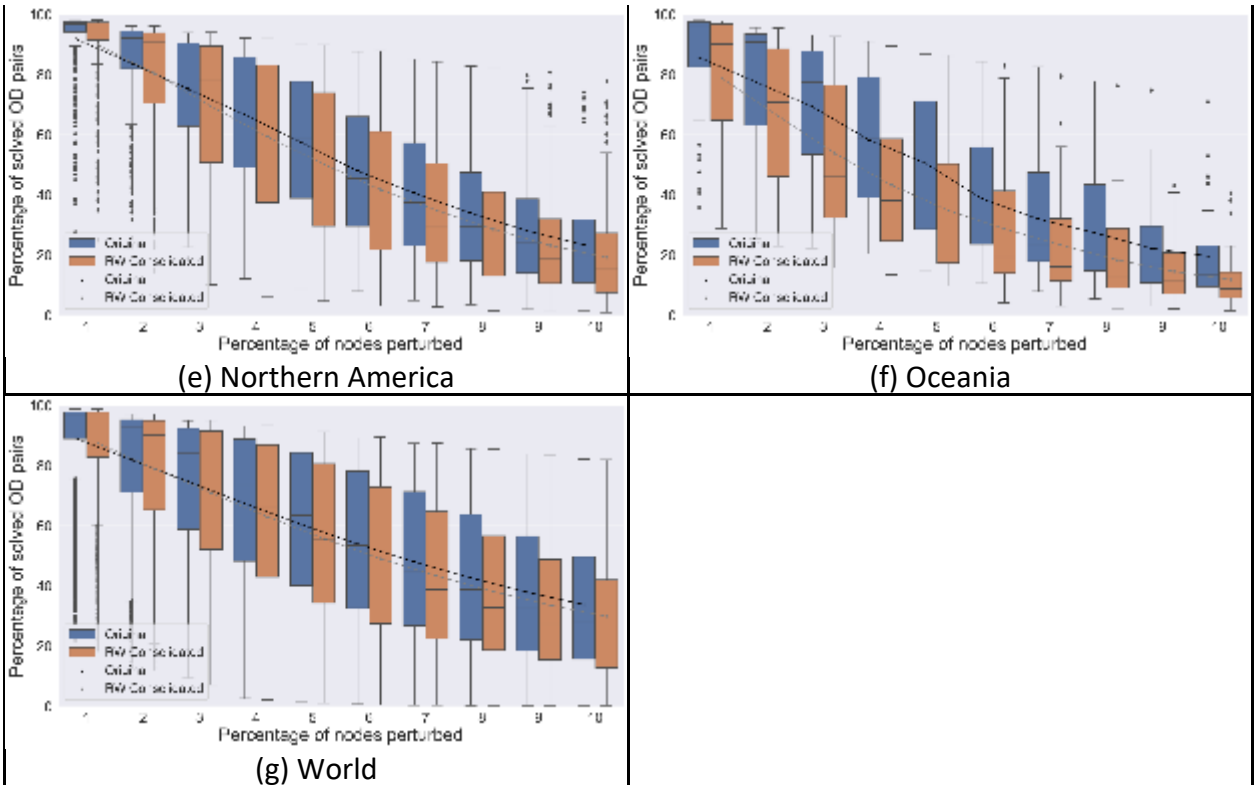
When we disrupted 10% of the nodes with high betweenness centrality in both the original and consolidated networks, more than 50% of the OD pairings became disconnected and unsolvable (see Figure 28). Asia and Africa experienced the slowest decline in the percentage of solved OD pairs as we increasingly perturb them, showing resilient street networks that do not rely excessively on major intersections.

On the other hand, the street networks get disconnected the fastest in Oceania and Europe with network perturbation by high betweenness centrality. This result demonstrates their reliance on high-centrality intersections and consequently their susceptibility to attacks when important intersections are targeted. Another significant finding is that the percentage of solved OD pairs fastly drops for the consolidated networks; this result suggests that using unconsolidated networks may introduce bias in understanding street network resilience.



(a) Asia (b) Africa

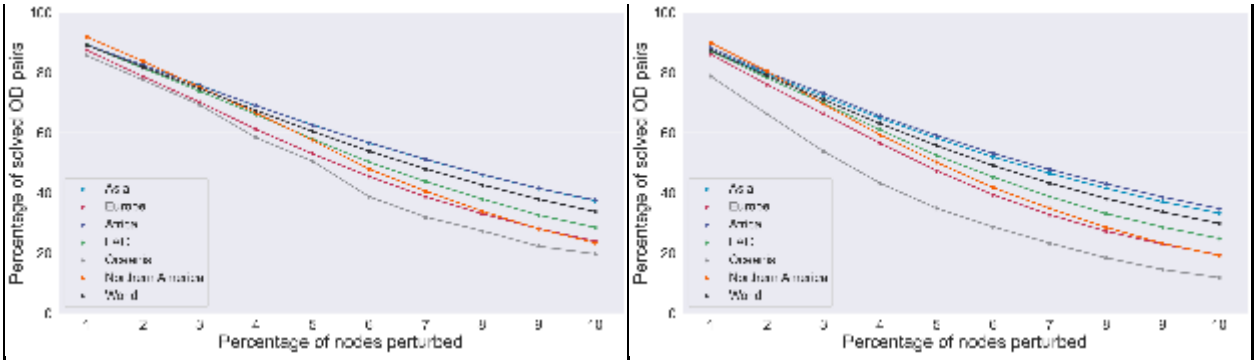
(c) Europe (d) Latin America and Caribbean



(e) Northern America (f) Oceania

(g) World

Figure 28. Boxplot of the percentage of solved OD pairs by region and network type based on network perturbation by betweenness centrality (Note: The trend lines are based on the mean value)



(a) Original (b) RW Consolidated

Figure 29. Trend lines of the percentage of solved OD pairs by region and network type based on network perturbation by betweenness centrality

Figure 29 and Table 9 show the trend lines and regression estimates of the percentage of solved OD pairs based on network perturbation. As shown in Table 9, around 6.6% of the OD pairs get unsolvable as we perturb 1% of the nodes. Among the six regions, the street network in cities in the Oceania region is mostly likely to be rapidly disconnected due to network perturbation,

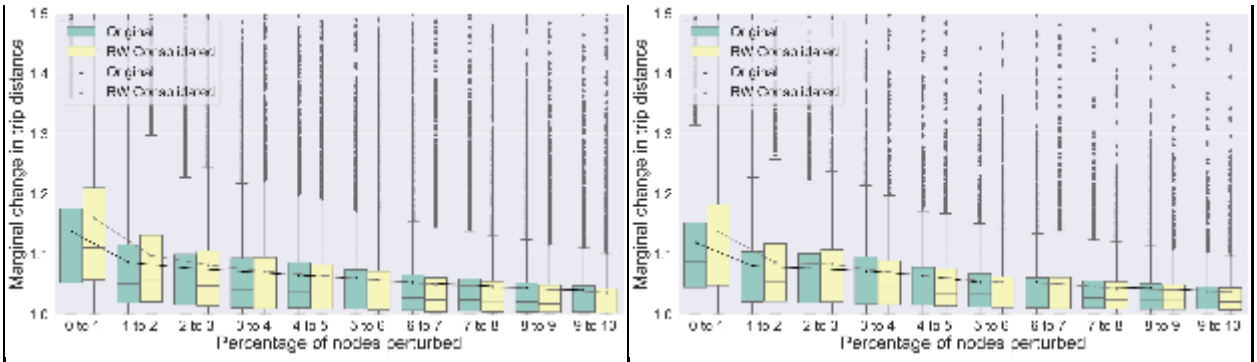
whereas cities in Africa and Asia show relatively higher resilience. For instance, when we attack 10% of the nodes, only 10-20% of the OD pairs remain solved in Oceania cities, while almost 40% remain solved in cities in Asia and Africa.

Table 9. Regression estimates on the trend lines of the percentage of solved OD pairs (network perturbation by betweenness centrality)

Region	Original network			RW Consolidated Network		
	R-square	Intercept	Coef.	R-square	Intercept	Coef.
World	0.488	95.485	-6.550	0.498	93.576	-6.889
Asia	0.449	95.421	-6.138	0.462	93.580	-6.485
Africa	0.438	95.344	-6.116	0.439	93.920	-6.352
Europe	0.614	94.227	-7.570	0.657	92.494	-8.056
LAC	0.569	96.092	-7.183	0.575	93.739	-7.481
Northern America	0.657	98.820	-7.962	0.642	95.806	-8.302
Oceania	0.642	93.796	-8.181	0.634	84.764	-8.333

Figure 30 shows the distribution of marginal changes in the trip distance by region and network type. The y-axis in Figure 30 represents the coefficient of marginal change; for instance, 1.2 indicates that the trip distance increases by 20% when we perturb the nodes by 1%. For the original network, the trip distance will increase the most (around 12.5%) when we perturb 1% of the nodes with the highest betweenness centrality. The increase gets more flattened as we continue to perturb the network by targeting high betweenness centrality nodes.

A similar trend applies to consolidated networks, but the increase in trip distance is more aggressive as we removed redundant nodes that appear to have high betweenness centrality in the original networks. For consolidated networks, trip distance increased by 15% when the highest 1% betweenness centrality nodes were perturbed. Oceania shows large gaps according to network type, which may be attributed to the small number of cities in the region.



(a) Asia (b) Africa

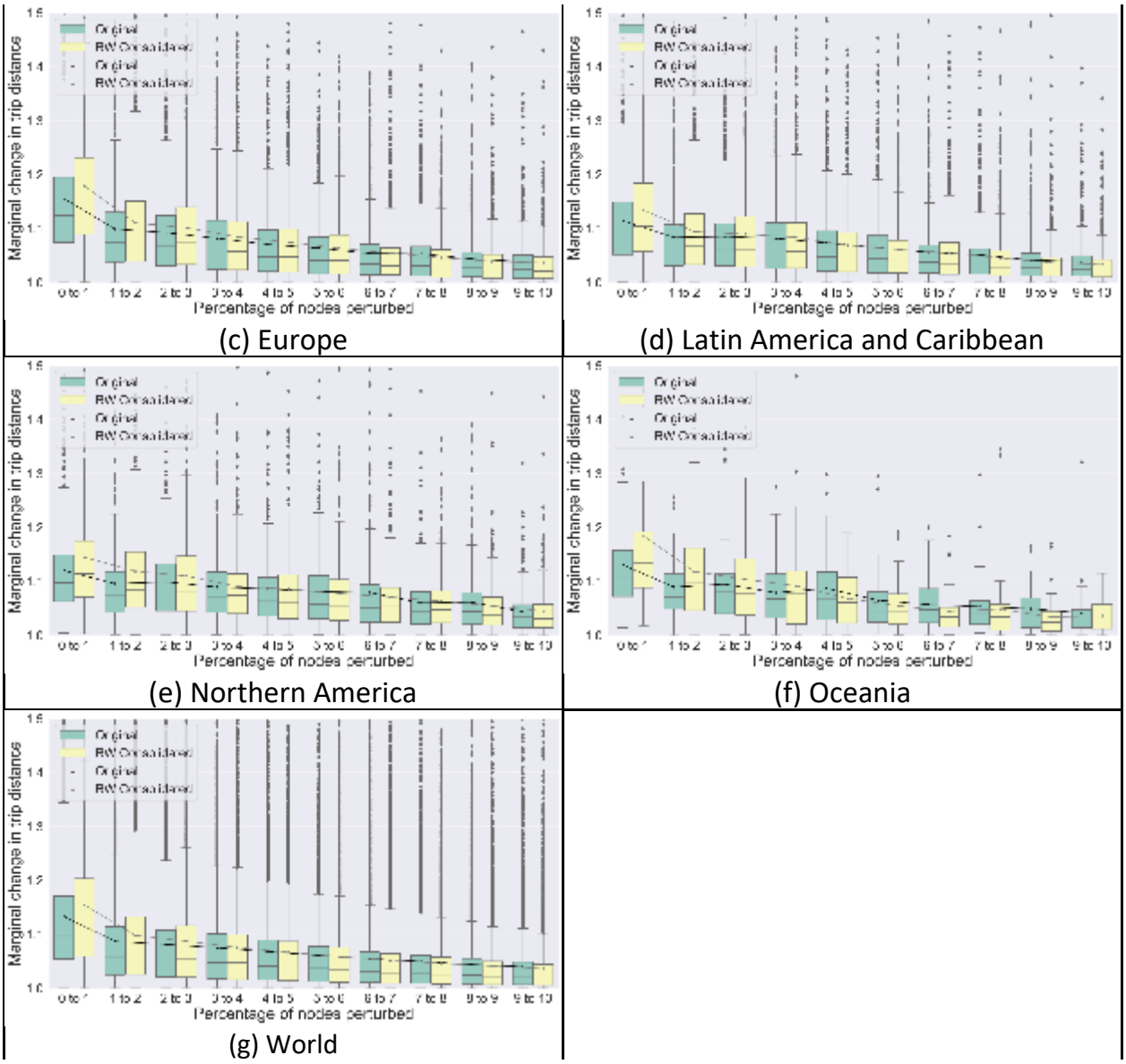
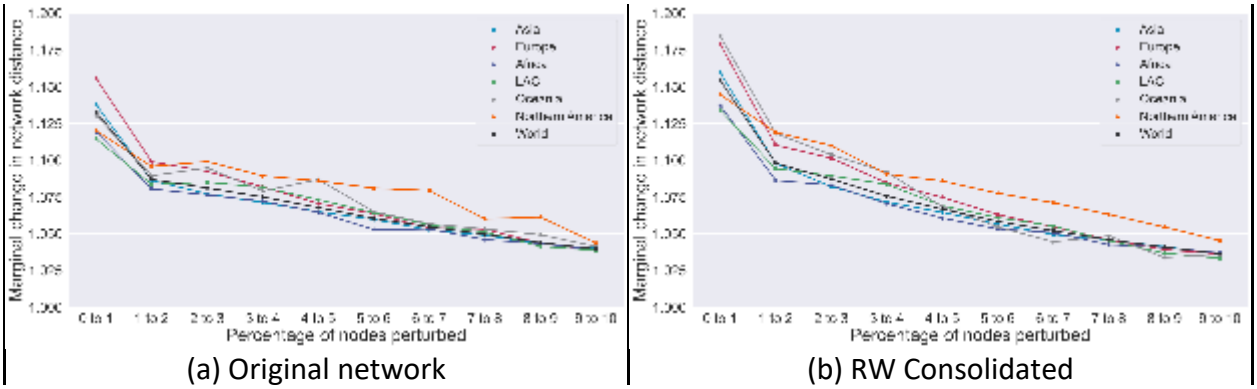


Figure 30. Boxplot of the marginal change in trip distance by region and network type based on network perturbation by betweenness centrality (Note: The trend lines are based on the mean value)

From Figure 31 and Table 10, we can compare how the marginal change in trip distance varies across the six regions based on the mean trend line. As shown in the figure 30, attacking the network focusing on the nodes with high betweenness centrality initially leads to greater marginal changes, while the impact diminishes as we continue perturbing the network. The results imply two findings; 1) protecting the network focusing on nodes with high betweenness centrality during potential disruption is critical, and 2) the diminishing impact on marginal change in trip distances may be due to the network getting disconnected from each other.

Thus, the results on marginal change should be interpreted with caution, considering how network perturbation impacts the percentage of solved OD pairs.



(a) Original network (b) RW Consolidated

Figure 31. Trend lines of the marginal change in trip distance by region and network type based on network perturbation by betweenness centrality

The street network in Amsterdam, for instance, includes several bridges connecting areas across rivers; in this case, the nodes associated with bridges are likely to have high betweenness centrality, in which perturbing the nodes with high betweenness centrality suddenly reduces the percentage of solvable OD pairs rather than increasing the trip distance.

Table 10. Regression estimates on the trend lines of the marginal change in trip distance (network perturbation by betweenness centrality)

Region	Original network			RW Consolidated Network		
	R-square	Intercept	Coef.	R-square	Intercept	Coef.
World	0.070	1.115	-0.008	0.098	1.130	-0.011
Asia	0.063	1.114	-0.008	0.087	1.129	-0.011
Africa	0.060	1.105	-0.007	0.082	1.116	-0.009
Europe	0.122	1.133	-0.010	0.158	1.151	-0.013
LAC	0.080	1.110	-0.007	0.121	1.124	-0.010
Northern America	0.066	1.120	-0.007	0.108	1.141	-0.010
Oceania	0.118	1.121	-0.008	0.197	1.157	-0.014

The percentage of trip distance in Europe increased the most when we perturbed the nodes with the highest 2% of betweenness centrality, demonstrating the crucial role this 2% of the nodes plays. In contrast, the travel distance increase in North America cities is steady across each percentage of nodes with high betweenness getting disrupted. This result demonstrates that cities in North America are not excessively dependent on the top percentage of nodes with the highest betweenness centrality. In contrast, the importance of the nodes is more dispersed compared to European cities.

Figure 32 shows the boxplot distribution of cumulative change in trip distance. The cumulative change is computed by the product of the marginal changes. The median value of all cities is 1.85 for the original network, suggesting that the trip distance is expected to increase by 85% when 10% of the nodes with high betweenness centrality are removed. Europe showed the biggest cumulative increase in trip distance when high centrality nodes are disrupted, indicating that the key function of European street networks is highly centralized in nodes with high betweenness centrality, making them more susceptible to targeted attacks. While Africa and Asia have the lowest travel distance increase, showing their important nodes are more decentralized and hence more resilient when facing focused attack compared with Europe.

The change was more prominent when using the consolidated network; this result shows that perturbing the network without combining the redundant nodes may result in underestimating the impact of network perturbation.

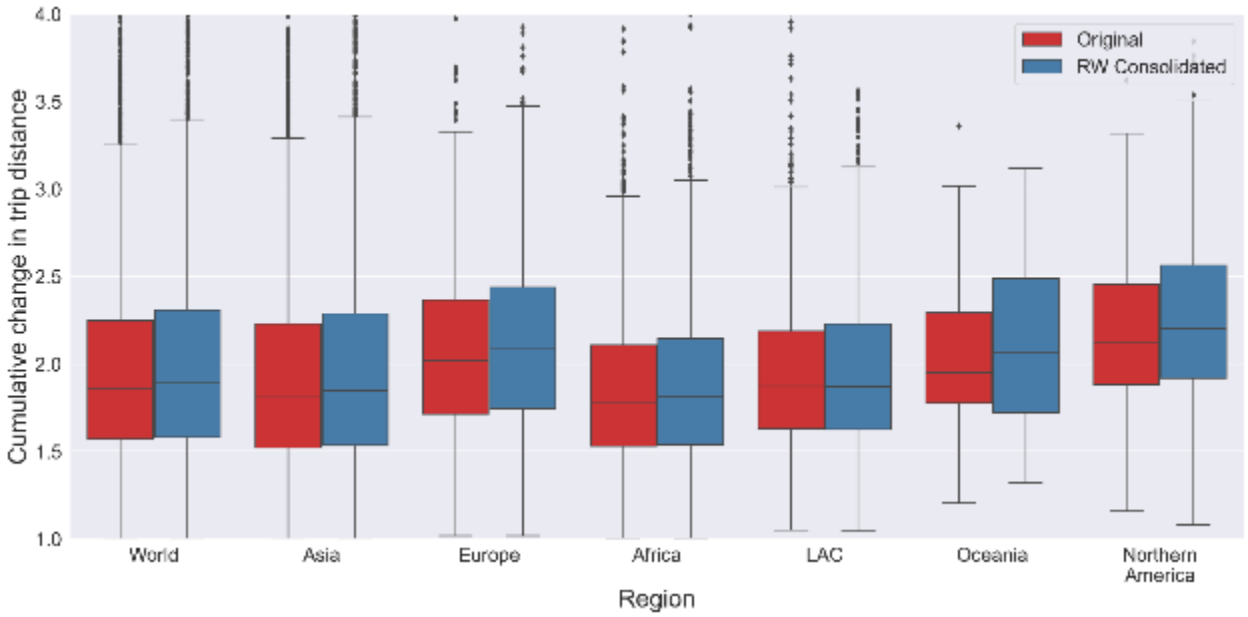


Figure 32. Boxplot distribution of cumulative change in trip distance by region and network type (network perturbation by betweenness centrality) (Note: Cities having cumulative change more than 4.0 are not shown in the figure)

Figure 33 shows the boxplot distribution of change in trip distance after 5% of node perturbation only using the OD pairs that were solvable after the perturbation. In other words, we are not using the OD pairs that are disconnected after the perturbation. We see a similar pattern across the regions compared to the cumulative change, while the value is smaller. For instance, the figure still shows that the change in trip level circuity is greater for regions such as Europe and Northern America. This is mainly because a large percentage of OD pairs gets unsolvable when attacking the nodes with high betweenness centrality.

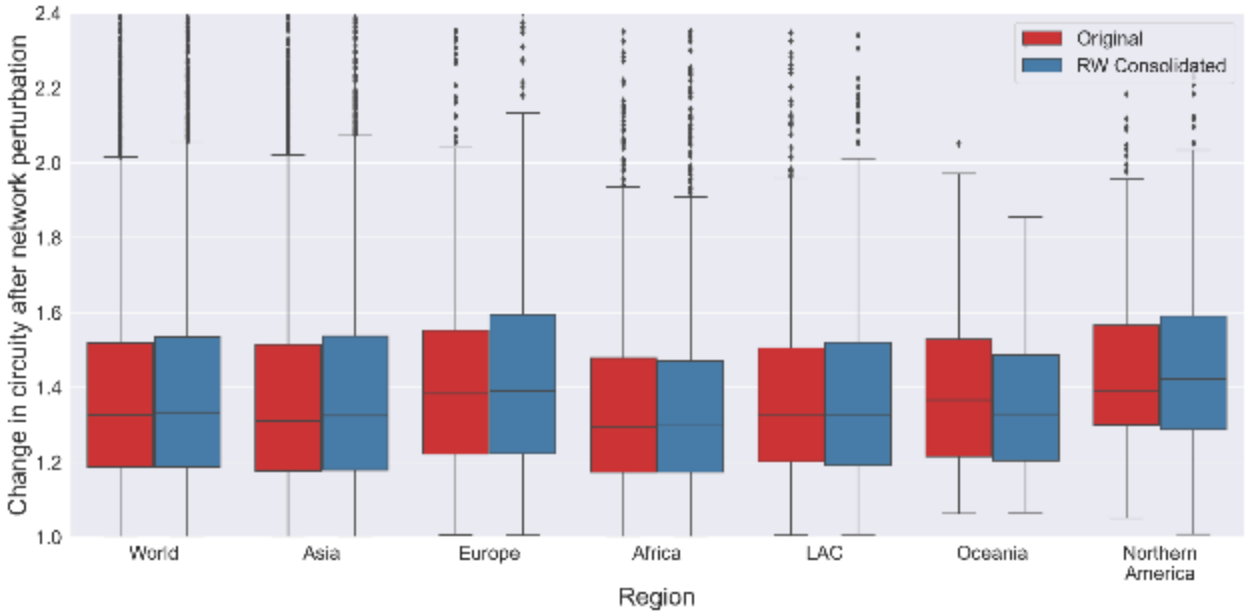


Figure 33. Boxplot distribution of change in trip distance by region and network type based on OD pairs solved when 5% of the nodes are perturbed (network perturbation by betweenness centrality) (Note: Cities having cumulative change more than 2.4 are not shown in the figure)

Similarly, Figure 34 shows the distribution of change in trip distance after 10% of node perturbation. In this case, we see a different pattern compared to the figures addressed above. As shown earlier, there are generally less than 40% of the OD pairs solvable after removing 10% of the nodes based on betweenness centrality. This indicates that most of the OD pairs are disconnected after the perturbation, and since the OD pairs that are most likely connected even after the perturbation are short in terms of their Euclidean distance, we would not expect a large change in circuity. Relatedly, the cities in Europe no longer stand out mainly because they have a smaller number of OD pairs connected after the perturbation. On the other hand, the change in circuity for Northern American cities is still great, which is partially due to the fact that a majority of the OD pairs are still connected to each other.

In terms of understanding the change in trip distance due to network perturbation, we have multiple variables that we address in this section. To elaborate, 1) cumulative change is a variable that comes from the marginal change, which is limited for addressing the OD pairs that are no longer solvable due to network attack, 2) change in trip distance could be identified by looking at the OD pairs that are solvable after the network perturbation by either 5% or 10%; however, this method is also limited because we are no longer addressing the OD pairs that are unconnected due to network perturbation. Relatedly, see Figure 35 for the scatter plot between the cumulative change and the change in trip length after 5% node attack. They have a

positive and linear relationship, indicating that these indicators can be a valid proxy of one another.

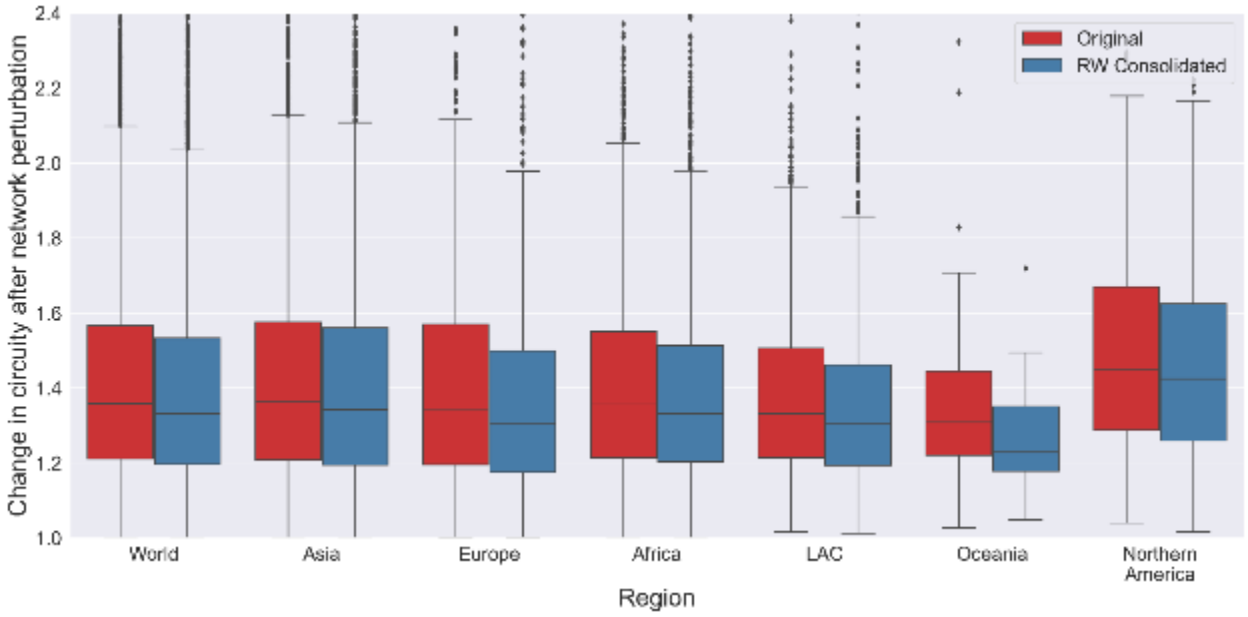
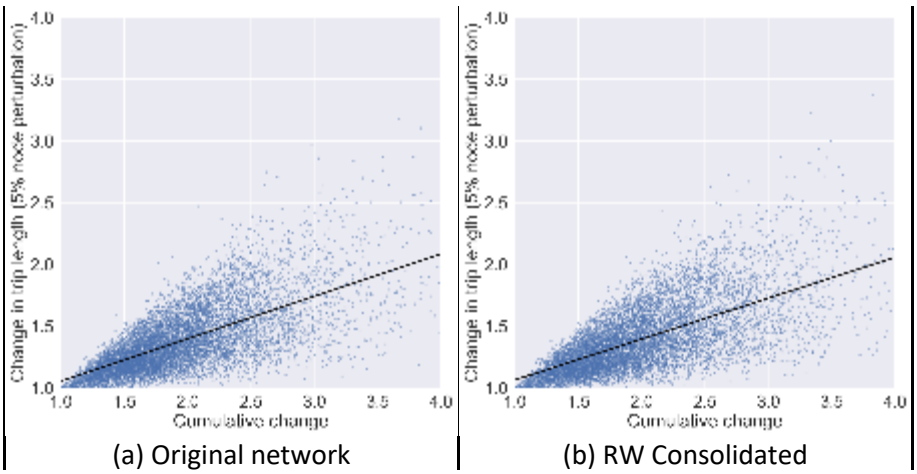


Figure 34. Boxplot distribution of change in trip distance by region and network type based on OD pairs solved when 10% of the nodes are perturbed (network perturbation by betweenness centrality) (Note: Cities having cumulative change more than 2.4 are not shown in the figure)

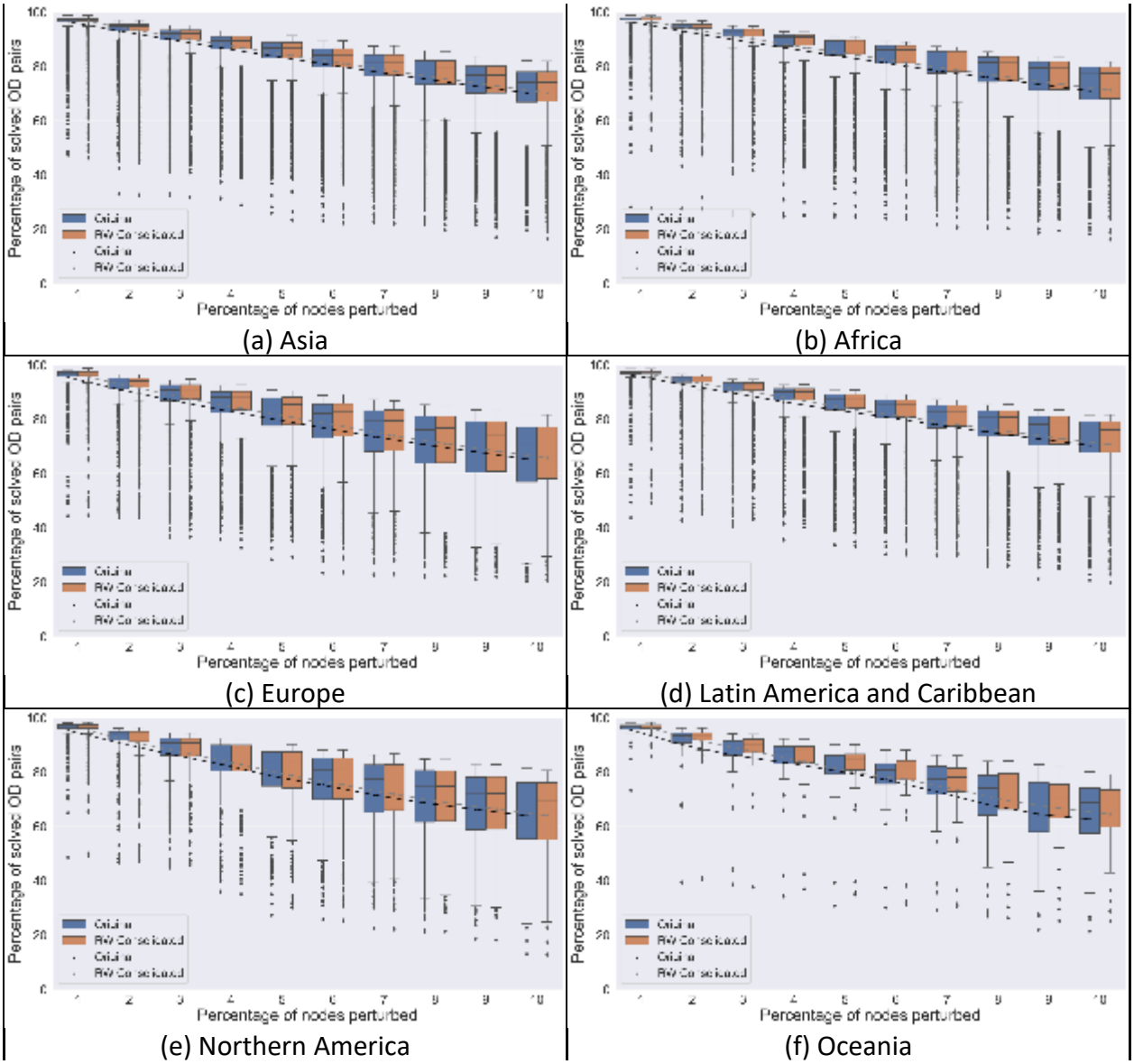


(a) Original network (b) RW Consolidated

Figure 35. Scatter plot between the cumulative change and change in trip distance by network type (network perturbation by betweenness centrality) (Note: outliers are not shown in the figure)

Perturb nodes with low elevation

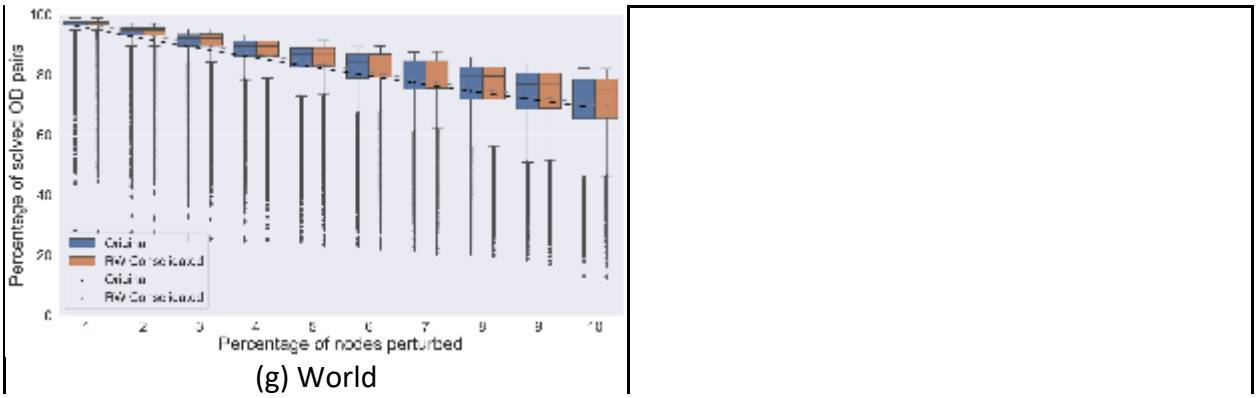
Next, we perturb the nodes with low elevation, which are likely more susceptible to disruptions due to flood. Considering the fact that around 20% of the OD pairs get unsolved when we remove 10% of the nodes, the change in the percentage of solved OD pairs is not dramatic compared to the network attack based on betweenness centrality (Figure 36). Among the six regions, Europe, Northern America, and Oceania turned out to be the most susceptible regions when we attacked low-elevation nodes. The results imply that the street network of cities in the three regions is vulnerable to flood disruptions, whereas cities in Asia, Africa, and LAC are more resistant to flood.



(a) Asia (b) Africa

(c) Europe (d) Latin America and Caribbean

(e) Northern America (f) Oceania

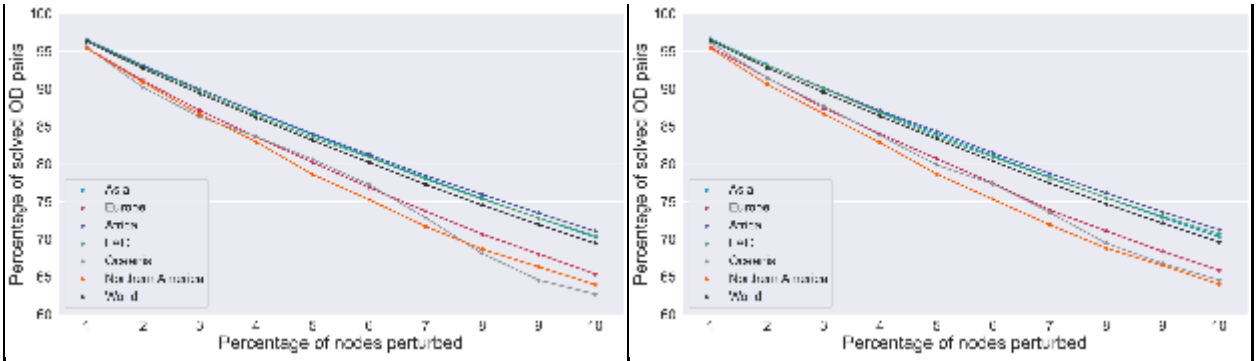


(g) World

Figure 36. Boxplot of the percentage of solved OD pairs by region and network type based on network perturbation by elevation (Note: The trend lines are based on the mean value)

There are a couple of things to be noted; 1) low elevation nodes within cities do not always indicate that they are vulnerable to floods; for example, we would have to consider how low they are compared to the other nodes and also address the local climate of each city, 2) the cities that have a low percentage of OD pairs solved due to elevation perturbation may imply that the low-lying nodes are likely to have high betweenness centrality.

Figure 37 and Table 11 show the trend lines of the percentage of solved OD pairs due to elevation perturbation. We would expect a decrease of around 3% in the percentage of solved OD pairs as we perturb 1% of the nodes. We can also see that the regions such as Europe, Oceania, and Northern America are the most susceptible to network perturbation by elevation.



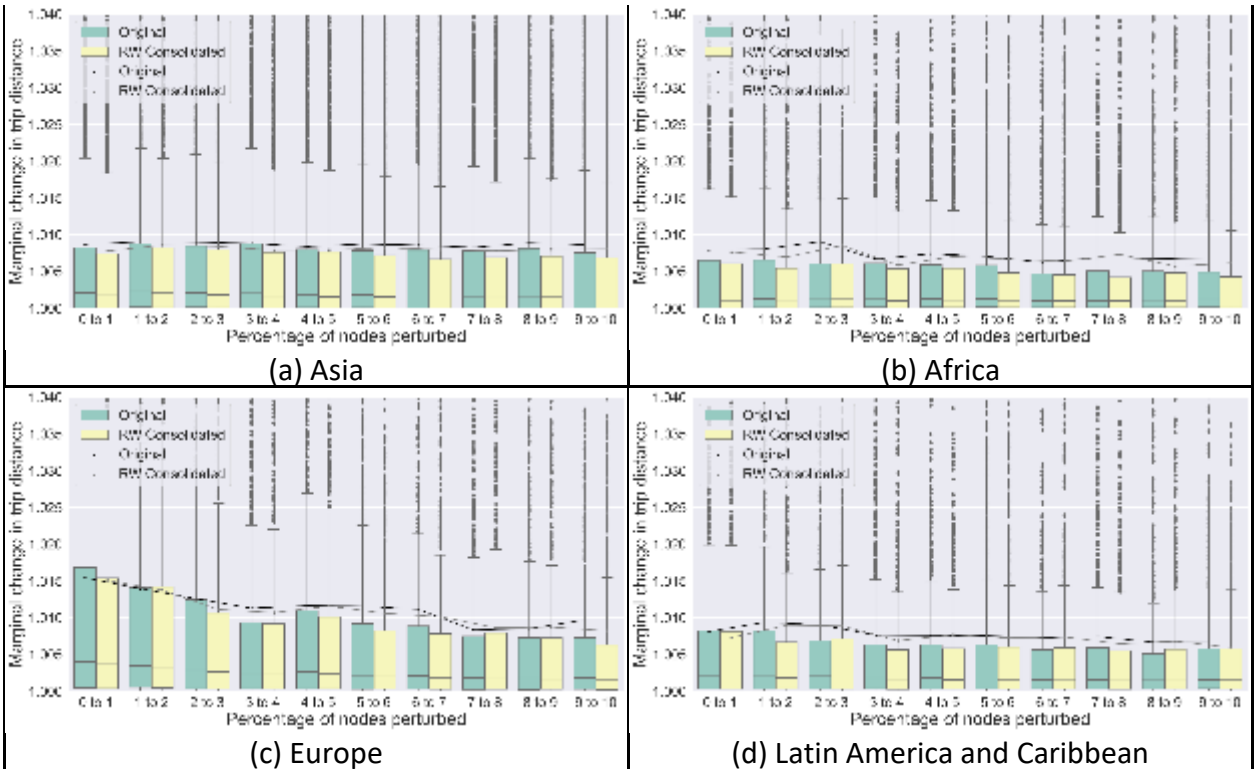
(a) Original network (b) RW Consolidated

Figure 37. Trend lines of the percentage of solved OD pairs by region and network type based on network perturbation by elevation

Table 11. Regression estimates on the trend lines of the percentage of solved OD pairs (network perturbation by elevation)

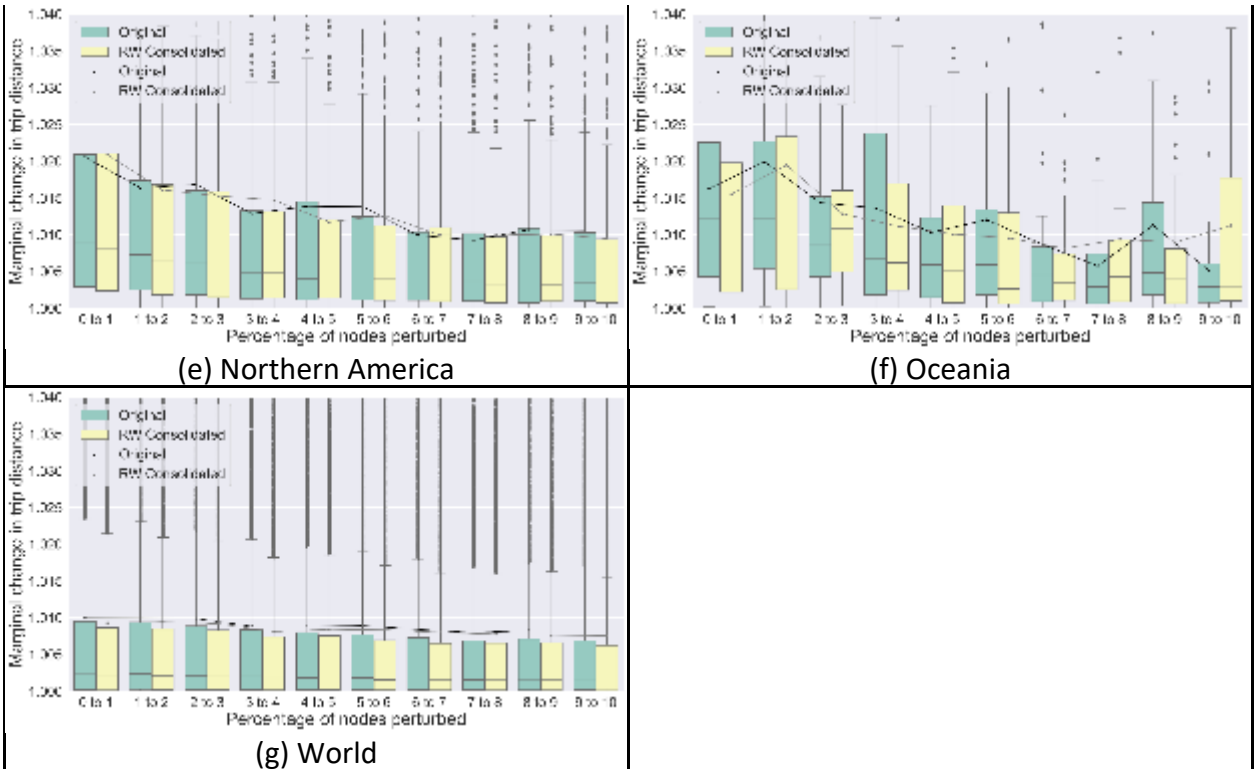
Region	Original network			RW Consolidated Network		
	R-square	Intercept	Coef.	R-square	Intercept	Coef.
World	0.496	98.953	-3.045	0.501	99.033	-3.033
Asia	0.552	99.189	-2.967	0.554	99.243	-2.962
Africa	0.445	98.990	-2.880	0.454	99.065	-2.861
Europe	0.481	98.264	-3.434	0.485	98.454	-3.405
LAC	0.470	98.929	-2.944	0.476	99.072	-2.935
Northern America	0.473	98.238	-3.643	0.471	98.151	-3.619
Oceania	0.489	98.791	-3.728	0.494	98.934	-3.595

The marginal change was the largest in North American cities, showing that low-elevation node disruption will result in longer travel distances (Figure 38). On the contrary, travel distance increased the least in Africa, demonstrating that flood perturbations will minorly affect street network efficiency in African cities. In short, African networks are more resilient to flood disruptions, while North American networks are vulnerable to such perturbation. Compared to the results when we perturbed nodes with high betweenness centrality, the impact of attacking nodes with low elevation is marginal. Another difference is that the impact of network perturbation on trip distance change is more negligible for the consolidated network than the original network, which differs from our findings on node perturbation by centrality.



(a) Asia (b) Africa

(c) Europe (d) Latin America and Caribbean

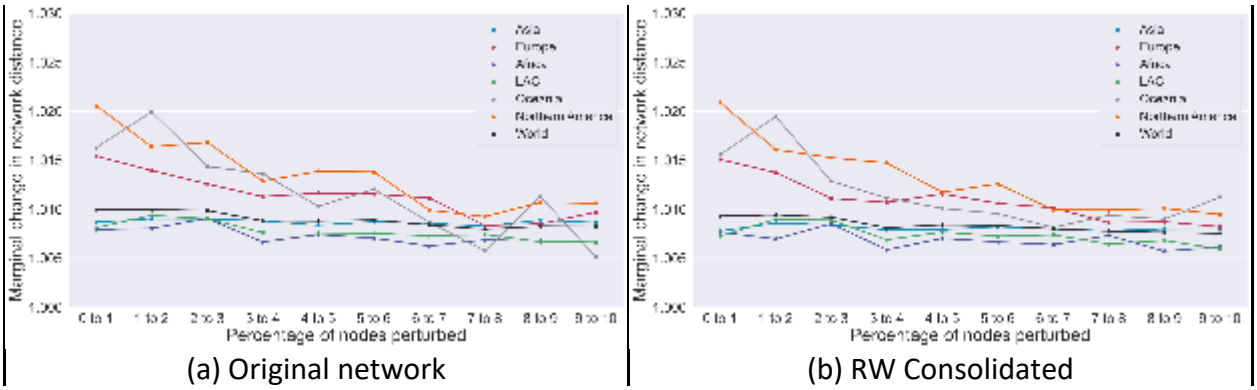


(e) Northern America (f) Oceania

(g) World

Figure 38. Boxplot of the marginal change in trip distance by region and network type based on network perturbation by elevation (Note: The trend lines are based on the mean value)

From Figure 39 and Table 12, we can see the difference across regions and identify the regions where the increase in trip distance is expected due to network attack by elevation. The marginal change in trip distance was relatively significant for regions such as Europe, Oceania, and Northern America. For other regions, the mean marginal change was less than 1%.



(a) Original network (b) RW Consolidated

Figure 39. Trend lines of the marginal change in trip distance by region and network type based on network perturbation by elevation

Table 12. Regression estimates on the trend lines of the marginal change in trip distance (network perturbation by elevation)

Region	Original network			RW Consolidated Network		
	R-square	Intercept	Coef.	R-square	Intercept	Coef.
World	0.00082	1.01018	-0.00023	0.00077	1.00957	-0.00022
Asia	0.00002	1.00888	-0.00003	0.00001	1.00820	-0.00003
Africa	0.00081	1.00837	-0.00020	0.00057	1.00773	-0.00017
Europe	0.00490	1.01513	-0.00068	0.00528	1.01459	-0.00068
LAC	0.00121	1.00910	-0.00025	0.00112	1.00862	-0.00023
Northern America	0.01159	1.01935	-0.00107	0.01246	1.01923	-0.00112
Oceania	0.04998	1.01897	-0.00132	0.01778	1.01628	-0.00084

As expected, the cumulative change is also high in regions such as Europe, Oceania, and Northern America (see Figure 40). For the other regions, the median value of the cumulative change was smaller than 1.05, which indicates that the trip distance only increases by 5% even after removing 10% of the nodes with low elevation. However, we can see from the figures that some regions are more susceptible to network attack by elevation. This implies that, unlike the attack on nodes with high betweenness centrality, there is great heterogeneity in the impact of elevation perturbation.

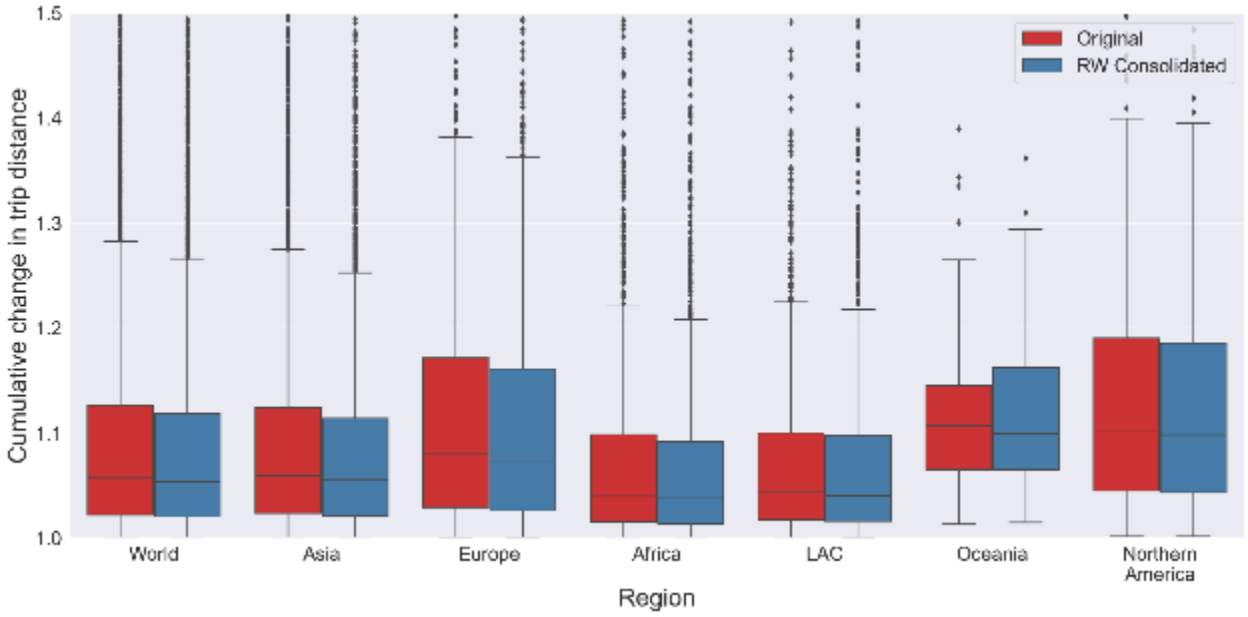


Figure 40. Boxplot distribution of cumulative change in trip distance by region and network type (network perturbation by elevation) (Note: Cities having cumulative change more than 4.0 are not shown in the figure)

Among the most populated cities, Sao Paulo showed a large cumulative change (original network=2.36; consolidated network=2.40). Since Sao Paulo is located near the coast, it would be essential to address the street network's vulnerability to flooding. Other cities such as Rome (original network=1.92; consolidated network=1.75), Seattle (original network=1.56; consolidated network=1.60), and Boston (original network=1.37; consolidated network=1.36) showed relatively large cumulative change values. On the other hand, cities such as Beijing (original network=1.01; consolidated network=1.01) and Tehran (original network=1.00; consolidated network=1.00) showed minimal cumulative changes.

Figures 41 and 42 show the change in the trip distance when 5% and 10% of the nodes are perturbed by elevation. The patterns are similar to the cumulative change, mainly because we have a large percentage of OD pairs solved after network perturbation in this case. As shown in Figure 40, however, the cumulative change is slightly overestimated when comparing only the OD pairs that are solved.

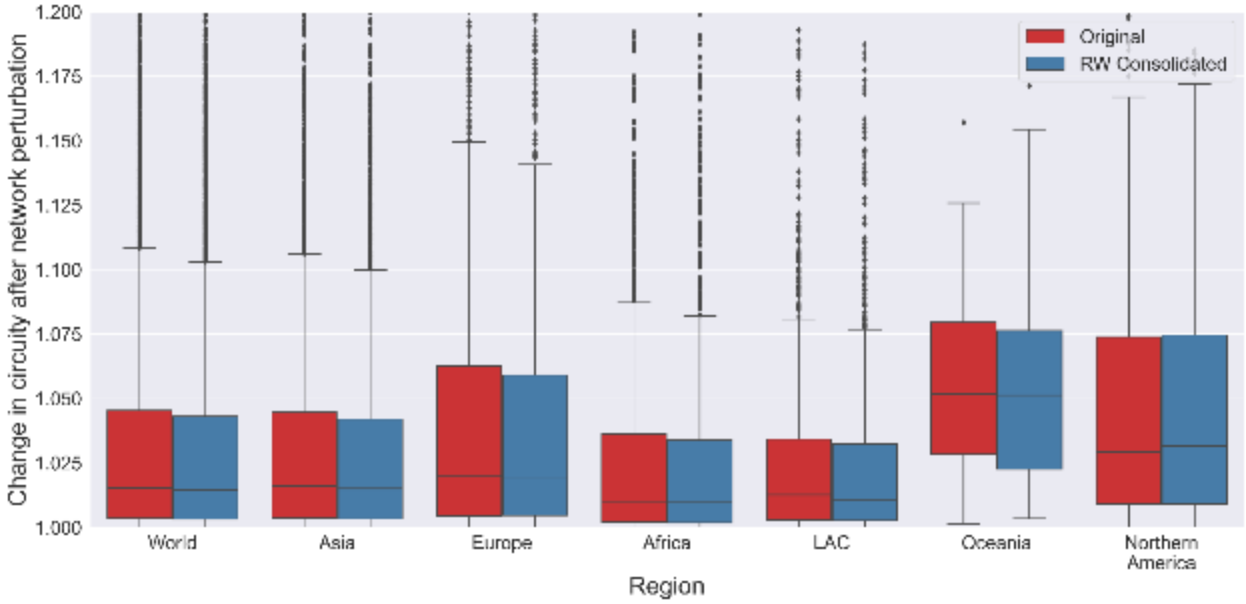


Figure 41. Boxplot distribution of change in trip distance by region and network type based on OD pairs solved when 5% of the nodes are perturbed (network perturbation by elevation) (Note: Cities having cumulative change more than 1.2 are not shown in the figure)

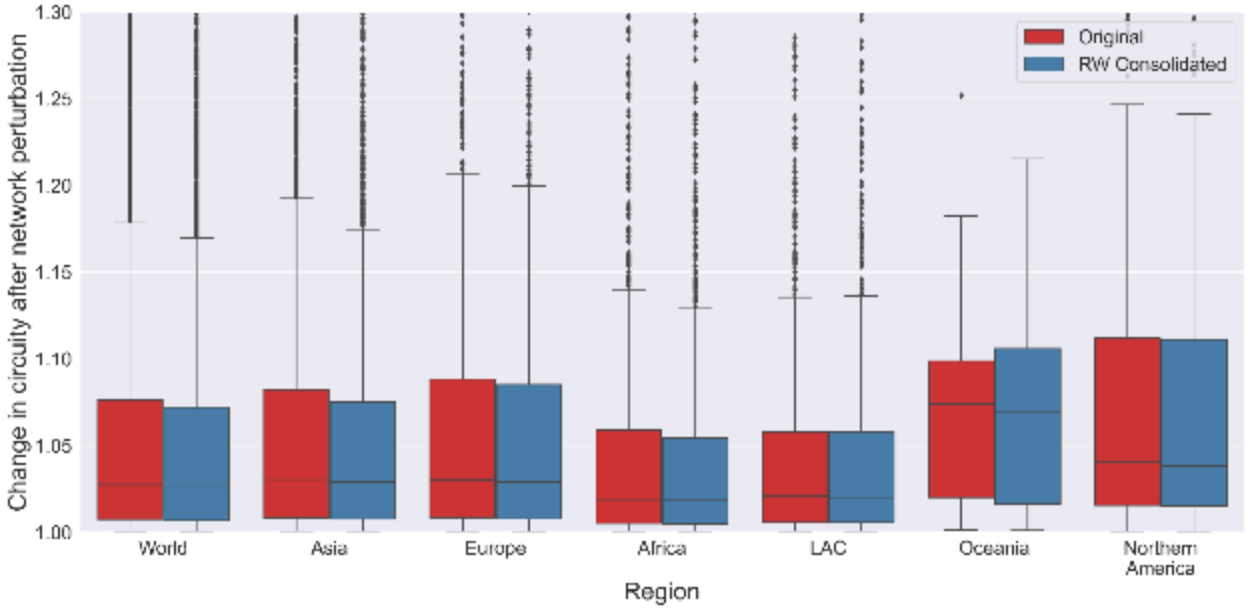
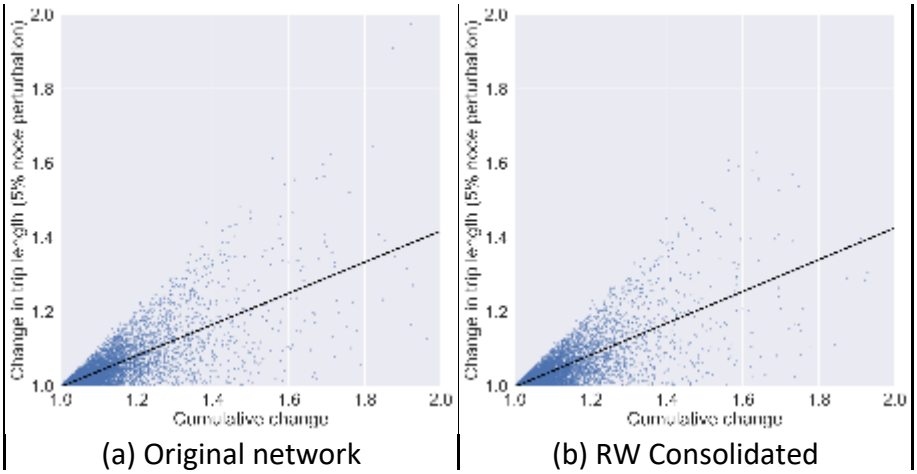


Figure 42. Boxplot distribution of change in trip distance by region and network type based on OD pairs solved when 10% of the nodes are perturbed (network perturbation by elevation) (Note: Cities having cumulative change more than 1.3 are not shown in the figure)



(a) Original network (b) RW Consolidated

Figure 43. Scatter plot between the cumulative change and change in trip distance by network type (network perturbation by betweenness centrality) (Note: outliers are not shown)

Perturb nodes randomly

We assimilated these random disruptions and examined how street networks perform in different regions. The percentage of solved OD pairs steadily decreases as we continue to randomly perturb street intersections. Around 30% of the OD pairs get unsolvable after removing 10% of the nodes randomly (see Figure 44).

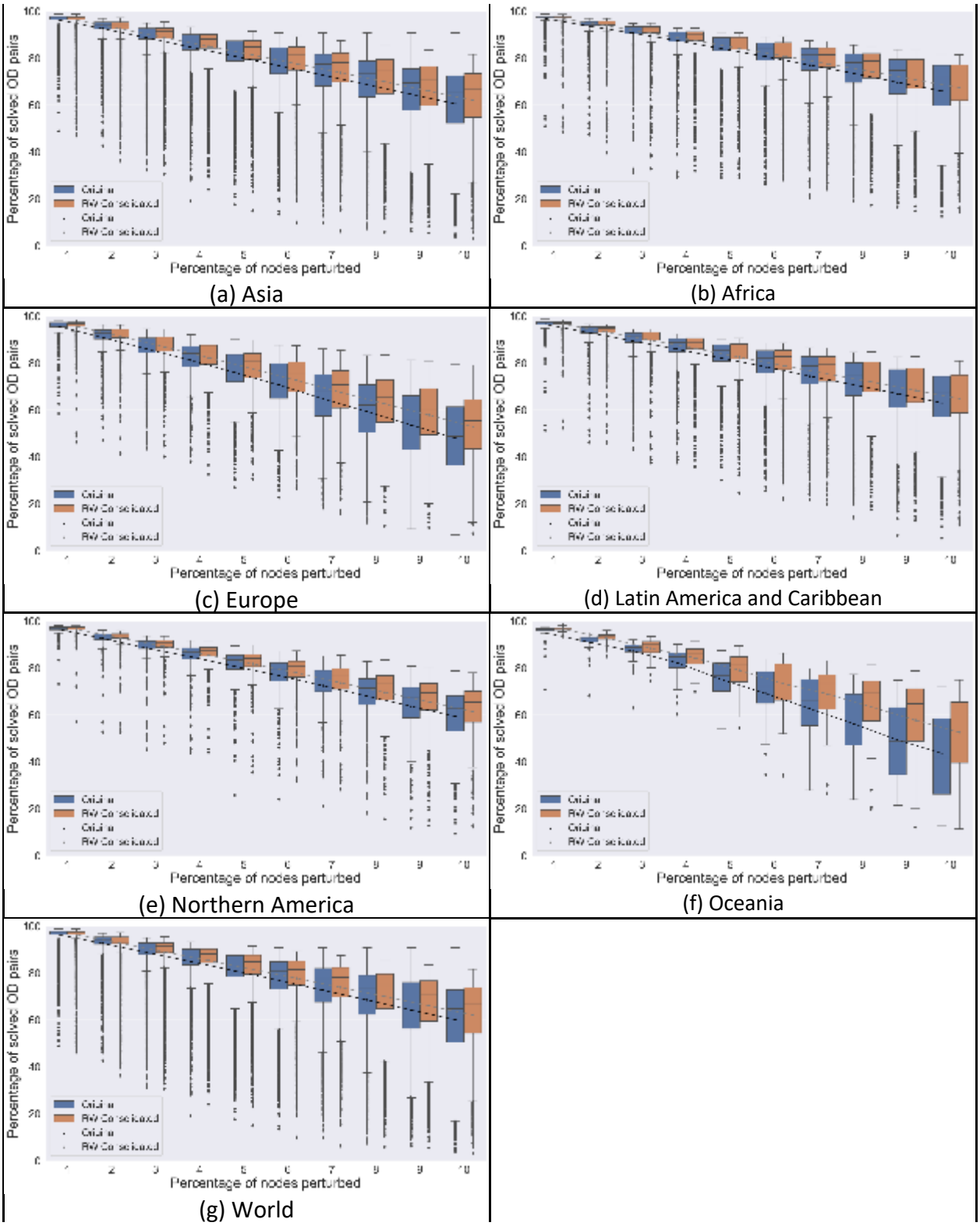
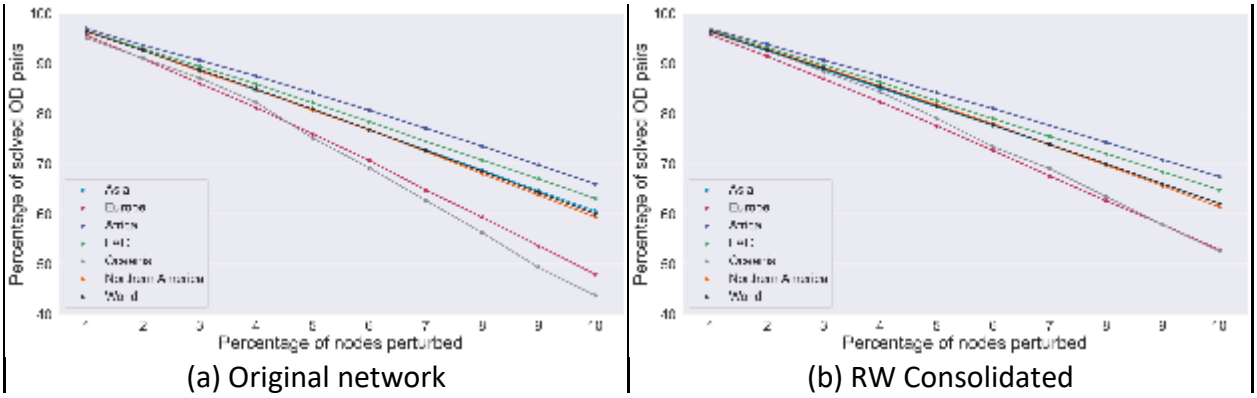


Figure 44. Boxplot of the percentage of solved OD pairs by region and network type based on network perturbation by random (Note: The trend lines are based on the mean value)

Cities in Oceania and Europe have the steepest decrease in solvable OD pairs, indicating that their street networks are vulnerable to random disruptions (Figure 45). Most OD pairs in Africa stay solvable, showing its street network resilience towards random attacks. From Table 13, around 3% to 5% of the nodes get unsolvable as we remove 1% of the nodes. Another notable difference is that the impact of network perturbation is greater for the original network than for the consolidated network. One possibility is that the random perturbation is more likely to remove significant nodes on the original network as they have more redundant nodes; in the consolidated network, the possibility of removing significant nodes by chance would be low.



(a) Original network (b) RW Consolidated

Figure 45. Trend lines of the percentage of solved OD pairs by region and network type based on network perturbation by random

Table 13. Regression estimates on the trend lines of the percentage of solved OD pairs (network perturbation by random)

Region	Original network			RW Consolidated Network		
	R-square	Intercept	Coef.	R-square	Intercept	Coef.
World	0.581	100.584	-4.006	0.577	100.311	-3.791
Asia	0.580	100.412	-3.958	0.570	100.156	-3.781
Africa	0.567	100.604	-3.392	0.554	100.327	-3.256
Europe	0.681	101.303	-5.235	0.684	100.795	-4.743
LAC	0.583	100.388	-3.697	0.576	100.234	-3.531
Northern America	0.696	100.690	-4.070	0.687	100.604	-3.857
Oceania	0.737	102.483	-5.732	0.671	102.056	-4.816

Figure 46 shows the marginal change in trip distance. The marginal change is generally more remarkable in regions such as Oceania and Europe. In these two regions, the trip length gets longer by around 4% as we remove 1% of the nodes. Compared to the other perturbation methods, the marginal change slightly increases as we perturb the network. For instance, the mean value of the marginal change for all cities is around 2% when we remove 1% of the nodes from the complete network. In contrast, the mean value increases to approximately 3% when we remove 1% of the nodes after removing 9% of the nodes in the network.

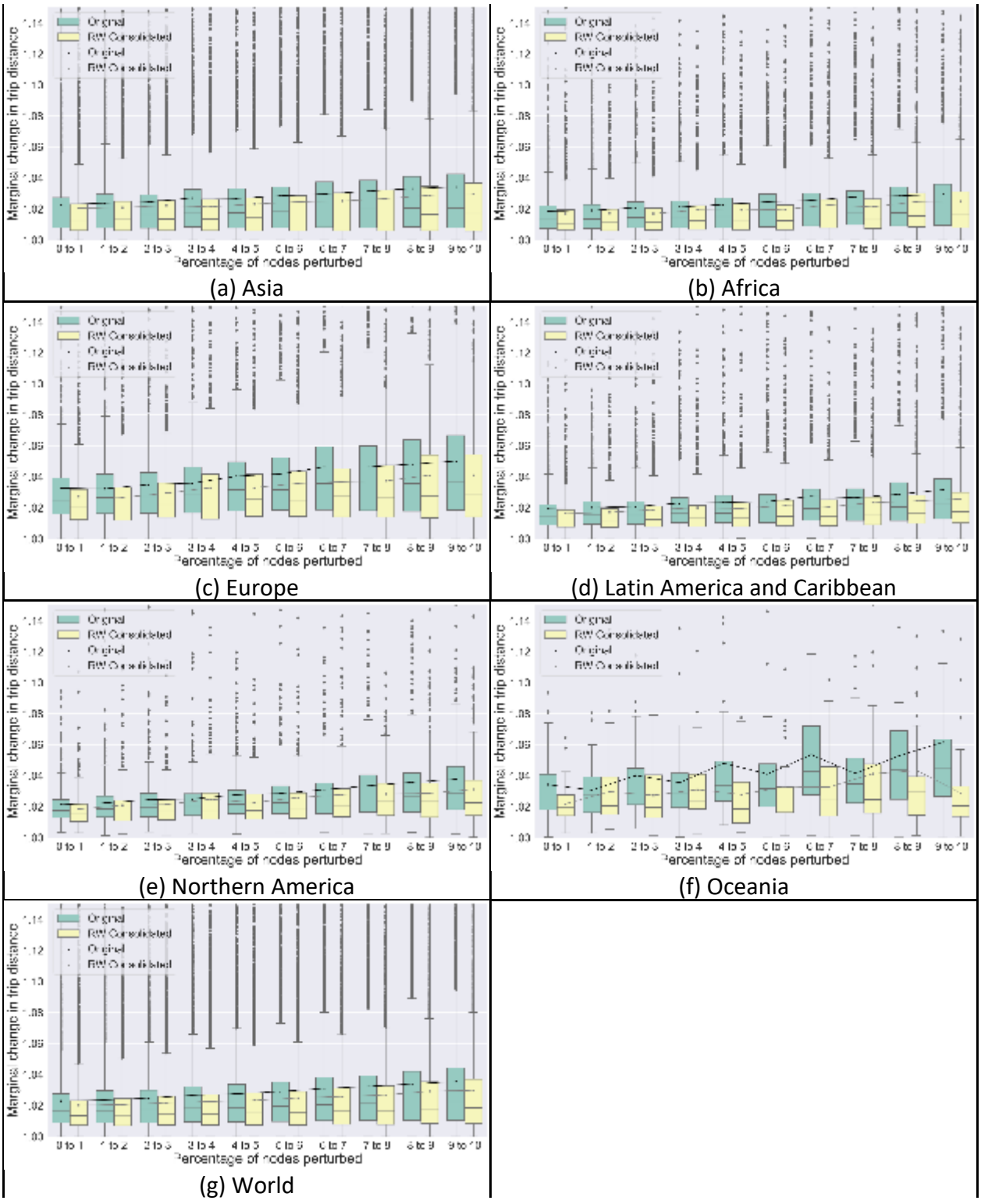
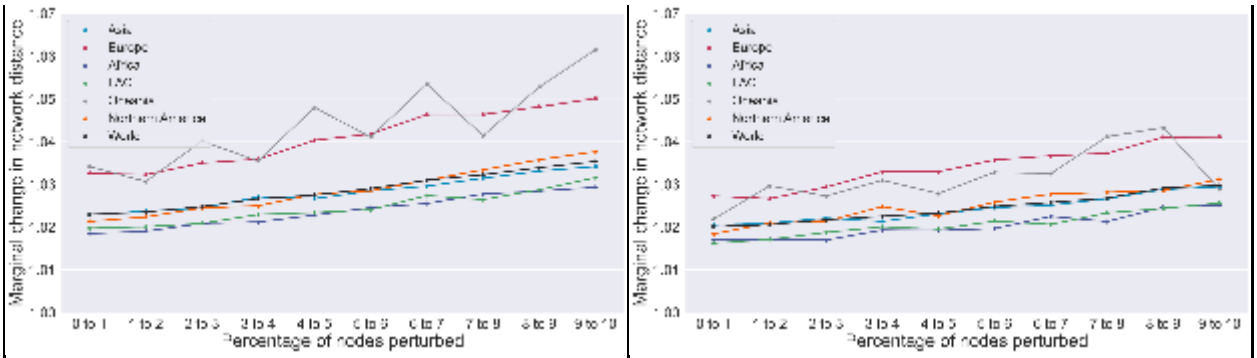


Figure 46. Boxplot of the marginal change in trip distance by region and network type based on network perturbation by random (Note: The trend lines are based on the mean value)



(a) Original network (b) RW Consolidated

Figure 47. Trend lines of the marginal change in trip distance by region and network type based on network perturbation by random

See Figure 47 and Table 14 for the trend line on the marginal change. In particular, the trend line for Oceania is not smooth, which may be due to the small number of cities located in Oceania.

Table 14. Regression estimates on the trend lines of the marginal change in trip distance (network perturbation by elevation)

Region	Original network			RW Consolidated Network		
	R-square	Intercept	Coef.	R-square	Intercept	Coef.
World	0.013	1.021	0.001	0.009	1.018	0.001
Asia	0.009	1.021	0.001	0.007	1.019	0.001
Africa	0.015	1.017	0.001	0.010	1.015	0.001
Europe	0.025	1.029	0.002	0.019	1.025	0.002
LAC	0.020	1.018	0.001	0.011	1.015	0.001
Northern America	0.045	1.019	0.002	0.026	1.018	0.001
Oceania	0.043	1.029	0.003	0.018	1.024	0.001

Using the marginal changes computed above, we calculated the cumulative change (see Figure 48 for the differences across regions). As expected, the cumulative change is relatively large in Europe and Oceania, suggesting that the street networks of cities in these regions are susceptible to random perturbation. The median value of the cumulative change for these two regions is around 1.5, indicating that the trip distance increases by 50% when 10% of the nodes are removed randomly. On the other hand, the cumulative change in Africa and LAC was around 1.2-1.3, and the results suggest that the cities in the two regions are relatively resilient.

Figure 48 shows the distribution of the cumulative change, and Figures 49 and 50 show the distribution of trip distance by using the OD pairs solvable after removing either 5% or 10% of the nodes from the network. The pattern across regions is similar, mainly because we have a large percentage of solved OD pairs even after removing 10% of the nodes.

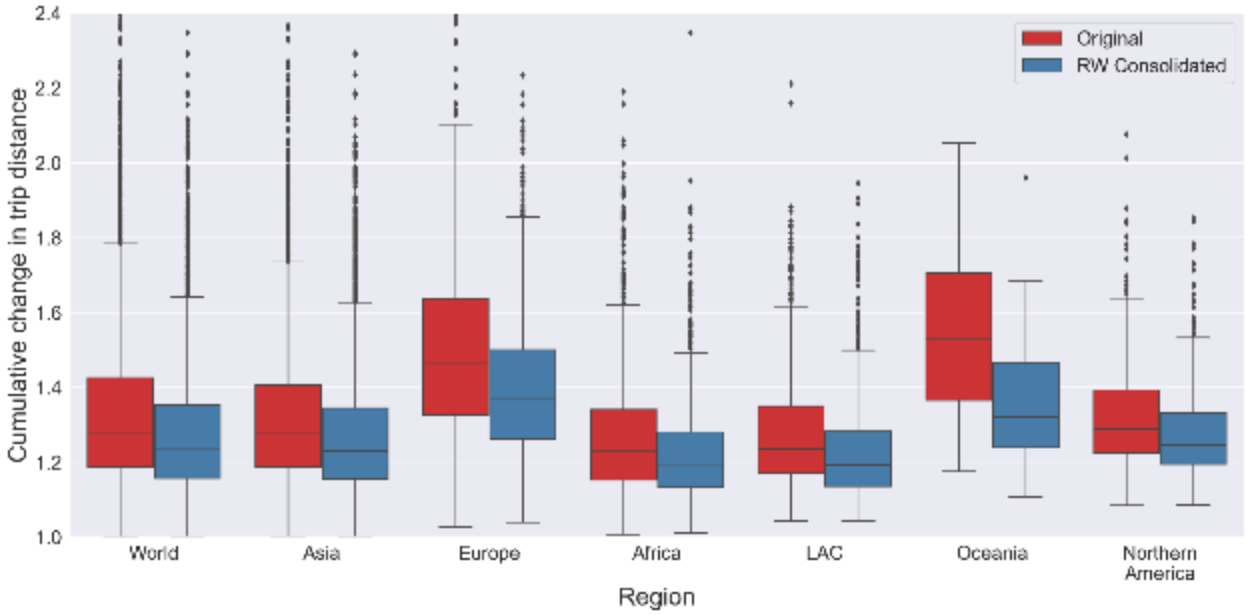


Figure 48. Boxplot distribution of cumulative change in trip distance by region and network type (network perturbation by random) (Note: Cities having cumulative change more than 4.0 are not shown in the figure)

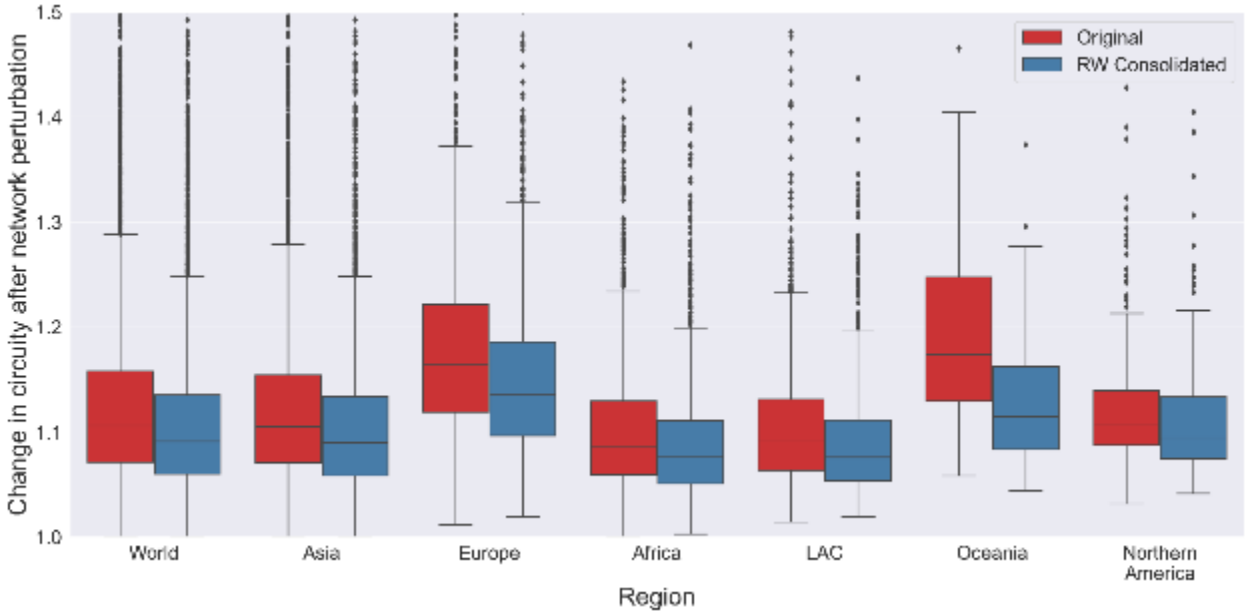


Figure 49. Boxplot distribution of change in trip distance by region and network type based on OD pairs solved when 5% of the nodes are perturbed (network perturbation by random) (Note: Cities having cumulative change more than 1.2 are not shown in the figure)

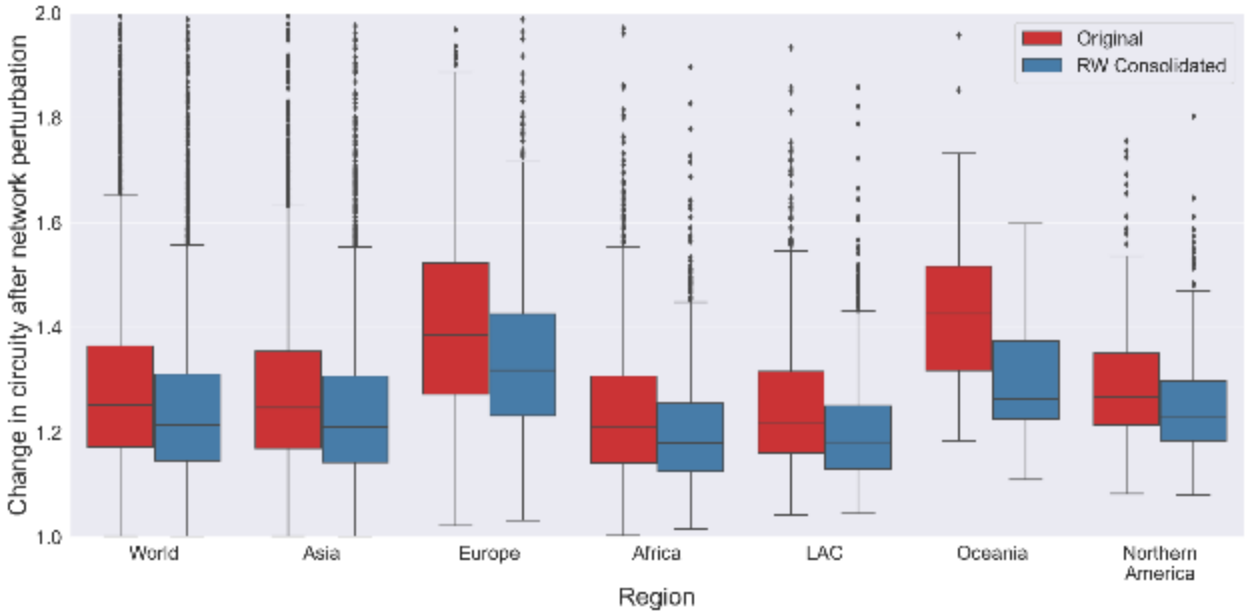


Figure 50. Boxplot distribution of change in trip distance by region and network type based on OD pairs solved when 10% of the nodes are perturbed (network perturbation by random) (Note: Cities having cumulative change more than 1.2 are not shown in the figure)

Figure 51 shows the scatter plot between the variables calculated in Figures 49 and 50. Similar to previous results, the cumulative change is overestimated than the change in trip length with solved OD pairs. This is because the cumulative change considers the change in trip distance for OD pairs that are unsolved when the nodes are removed. If most of the OD pairs are solved even after perturbing the network, the scatter plot should locate close to the x=y line.

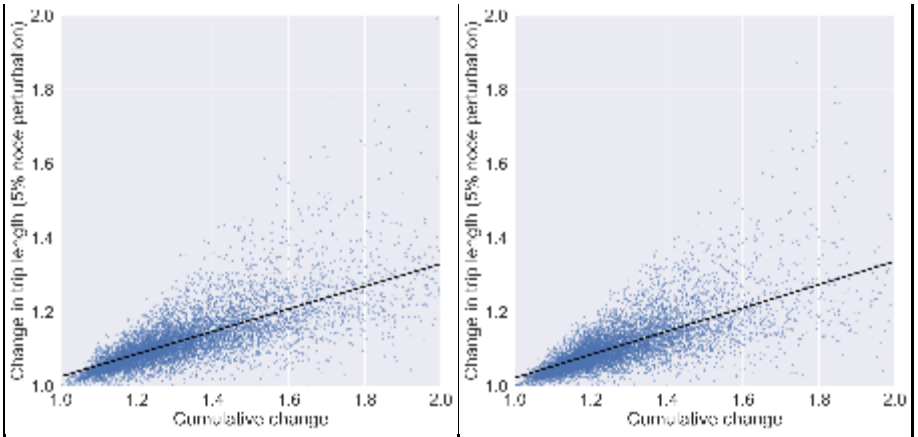


Figure 51. Scatter plot between the cumulative change and change in trip distance by network type (network perturbation by betweenness centrality) (Note: outliers are not shown in the figure)

4.2.3 Determinants of Network Resilience

Descriptive statistics

Table 15 shows the descriptive statistics of the dependent and independent variables. The percentage of solved OD pairs after perturbing the network varied across perturbation methods. The mean value of percentage of solved OD pairs after perturbing 10% of the nodes is 33.90 for the BC perturbation, while it is around 60-70% for the other perturbation methods. This clearly shows that the network gets rapidly disconnected when the nodes with high importance are attacked. The change in circuity of solved OD pairs slightly increased between 5% and 10% network node perturbation. Among the perturbation methods, BC perturbation greatly changed circuity compared to the other two methods.

We also see a large variance in street network features. For instance, the average intersection density is 197.29 per square kilometer, varying from 2.43 to 5484.73. Another important measure was the number of standard deviations between the maximum and mean value of betweenness centrality; the maximum value is above 30, which implies that the network has particularly important nodes. For the world region dummy variables, it should be noted that we have only 41 cities from Oceania, whereas more than half of the cities are located in Asia.

Table 15. Descriptive statistics

Variable	Mean (%)	Min	Max	SD
% solved OD pairs after perturbing 10% of nodes				
Perturbation by betweenness centrality (BC)	33.90	0.08	82.12	22.22
Perturbation by elevation	69.42	13.06	82.06	12.74
Perturbation by random	60.10	3.79	90.87	16.22
Change in circuity of solved OD pairs (5% perturb)				
Perturbation by betweenness centrality	1.39	1.00	3.83	0.28
Perturbation by elevation	1.04	1.00	2.08	0.06
Perturbation by random	1.13	1.00	2.31	0.09
Change in circuity of solved OD pairs (10% perturb)				
Perturbation by betweenness centrality	1.43	1.00	4.11	0.30
Perturbation by elevation	1.06	1.00	2.28	0.08
Perturbation by random	1.29	1.00	2.48	0.17
k average	2.82	1.57	3.66	0.25
circuity	1.05	1.00	1.43	0.03
intersection density	197.29	2.43	5484.73	211.10
elevation (std)	17.16	1.41	255.39	18.36
BC (max-mean)/std	6.70	2.03	37.13	2.98
length total per area	30809.60	3364.24	678800.49	25040.46

Variable	Mean (%)	Min	Max	SD
pct open space	67.25	17.70	100.00	17.31
built up area	35.91	1.00	4632.77	143.20
pop density	22995.97	973.89	538314.10	29916.67
World region				
Africa	1423	17.8%		
Asia	4113	51.4%		
Europe	1049	13.1%		
Latin American and the Carribean	1007	12.6%		
Northern America	372	4.6%		
Oceania	41	0.5%		

OLS results - Network disconnection (10% network perturbation)

Table 16 reports the OLS regression results. The dependent variable is the percentage of solved OD pairs after 10% of network perturbation; in other words, a lower value indicates that the network gets more disconnected due to perturbation. We would expect to see positive coefficient signs for *k average*, *intersection density*, and *length total per area*; higher node degree and intersection densities may allow more possible trip routes after network perturbation. On the other hand, we would expect negative signs for *circuity* and *BC (max-mean)/std*. If the network is circuitous or has particularly important nodes, it is more likely that the network gets rapidly disconnected with perturbation.

Among the indicators associated with street network design, k average, circuity, and the number of standard deviations between the maximum and mean of betweenness centrality showed results as expected. That is, street networks having larger node degrees, less circuity, and fewer nodes with particular importance are more likely to be resilient to network perturbation.

On the other hand, the results for intersection density and the standard deviation of elevation showed mixed signs. While higher intersection density is hypothesized to have a positive impact on network resilience, it was not true for the BC and random perturbations. Since we removed 10% of the nodes, we may have perturbed a large number of intersections in cities with high intersection density; this could have resulted in severe network disconnection by taking out nodes in certain areas. The standard deviation of node elevation also showed mixed results; it showed a positive impact on network resilience when the low-lying nodes are attacked. This result implies that cities having less variation in node elevations may be more vulnerable to attacks such as floods; however, it showed a negative sign for random perturbation, while the effect was insignificant for BC perturbation.

Considering the effect size, k average and the number of standard deviations between the maximum and mean value of betweenness centrality was the largest compared to other variables. This suggests that the average node degree and the presence of important nodes are

intimately associated with street network resilience. The R-squared value was around 0.4-0.5 for the models based on BC and random perturbation; the R-squared for the model based on elevation perturbation was 0.128, implying that our independent variables are limited for explaining the variation in the percentage of solved OD pairs due to elevation perturbation.

Cities with a higher percentage of open space also showed less resilience, in which border vacuums can increase the probability of network disconnection. Per our control variables, cities with large built-up area and high population density were less likely to be vulnerable to network perturbation. Our region dummy variables show that cities in Africa are likely the most resilient; the percentage of solved OD pairs were significantly lower for cities in other regions.

Table 16. Regression model parameter estimates for three network perturbation types (DV: percentage of solved OD pairs after perturbing 10% of the network nodes)

	BC perturbation		Elevation perturbation		Random perturbation	
	coef.	beta	coef.	beta	coef.	beta
(constant)	-24.915***	0.111	104.744***	0.078	51.816***	0.263
k average	41.396***	0.459	8.534***	0.165	31.559***	0.480
circuity	-29.772***	-0.040	-48.145***	-0.112	-66.799***	-0.122
intersection density	-0.012***	-0.116	0.006***	0.096	-0.007***	-0.098
elevation (std)	-0.001	-0.001	0.120***	0.173	-0.020**	-0.023
BC (max-mean)/std	-3.654***	-0.490	-0.827***	-0.194	-0.633***	-0.116
length total per area	-0.000	-0.020	-0.000**	-0.086	0.000	0.046
pct open space	0.007	0.006	-0.079***	-0.108	-0.036***	-0.038
built up area	0.017***	0.112	0.002**	0.026	0.004***	0.035
pop density	0.000***	0.135	0.000***	0.118	0.000***	0.098
Africa (ref)						
Asia	-0.439	-0.020	-0.012	-0.001	-3.216***	-0.198
Europe	-5.148***	-0.232	-3.608***	-0.283	-11.819***	-0.729
LAC	-9.868***	-0.444	-2.362***	-0.185	-6.560***	-0.405
Northern America	-6.441***	-0.290	-4.080***	-0.320	-3.503***	-0.216
Oceania	-5.934**	-0.267	-5.818***	-0.457	-13.950***	-0.860
R-squared	0.492		0.128		0.408	

N=8005; *** p<0.01, ** p<0.05, * p<0.10

OLS results - Change in circuity of solved OD pairs (5% network perturbation)

Table 17 shows the regression estimates for our models on change in circuity. The dependent variables are the changes in the circuity of solved OD pairs after removing 5% of network nodes using three different methods. We hypothesize that k average and intersection density would show a negative coefficient; we expect that cities with high node degrees and more intersections would experience fewer changes in circuity from network perturbation. On the

other hand, we hypothesize a positive coefficient for circuity and the number of standard deviations between the maximum and mean value of betweenness centrality. Street networks that are circuitous or have particularly important nodes would have more changes in trip distances due to network attacks.

Table 17. Regression model parameter estimates for three network perturbation types (DV: change in circuity of solved OD pairs after perturbing 5% of the network nodes)

	BC perturbation		Elevation perturbation		Random perturbation	
	coef.	beta	coef.	beta	coef.	beta
(constant)	1.151***	-0.078	0.745***	-0.077	1.043***	-0.194
k average	-0.030*	-0.027	-0.003	-0.011	-0.099***	-0.277
circuity	0.244*	0.026	0.268***	0.128	0.324***	0.109
intersection density	0.000***	0.145	-0.000***	-0.098	0.000***	0.116
elevation (std)	0.000	0.016	-0.000***	-0.134	-0.000	-0.008
BC (max-mean)/std	0.026***	0.276	0.002***	0.104	0.003***	0.102
length total per area	-0.000*	-0.061	0.000**	0.088	-0.000	-0.039
pct open space	-0.002***	-0.139	0.000***	0.043	-0.000***	-0.039
built up area	0.000***	-0.080	0.000**	0.028	-0.000***	-0.062
pop density	0.000	0.013	-0.000***	-0.071	-0.000***	-0.080
Africa (ref)						
Asia	0.039***	0.142	0.003	0.044	0.016***	0.179
Europe	0.025**	0.090	0.014***	0.219	0.052***	0.591
LAC	-0.024**	-0.087	0.006**	0.093	0.013***	0.152
Northern America	0.029*	0.104	0.017***	0.267	0.005	0.056
Oceania	-0.024	-0.089	0.024**	0.394	0.056***	0.632
R-squared	0.104		0.050		0.179	

N=8005; *** p<0.01, ** p<0.05, * p<0.10

Across the three models, circuity showed a consistent and significant positive sign. This result suggests that circuitous street networks are more vulnerable to network perturbation. In addition, the number of standard deviations between the maximum and mean value of betweenness centrality also showed significant and positive signs. If there are particularly important nodes within the network, it is highly likely that trip distances increase due to network perturbation.

Other variables showed mixed results. The k average showed a negative sign as expected, but the result was not significant in the model based on elevation perturbation. The results for intersection density matched with our hypothesis only when we perturbed the low-elevation nodes; interestingly, cities with more intersections may be vulnerable to network perturbation when the attack is conducted in nodes either with high betweenness centrality or by random. In addition, the length of street segments per area showed a negative sign for BC perturbation,

while it showed a positive sign for elevation perturbation. Here, we would expect a negative sign because cities having greater street segment density is likely to provide other alternative routes after perturbation.

The beta coefficient of variables also showed different patterns across the three models. For instance, intersection density and the presence of particularly important nodes stood out when the perturbation was conducted in nodes with high centrality. On the other hand, circuity and the standard deviation of elevations showed large beta coefficients for the elevation perturbation model. When we perturbed the network by random, the k average showed the largest beta coefficient. Overall, the R-squared values were relatively low, ranging from around 0.05 to 0.18.

OLS results - Change in circuity of solved OD pairs (10% network perturbation)

Lastly, Table 18 shows the regression results on the change in circuity of solved OD pairs after 10% network perturbation. We would expect the results to be similar to the ones shown in the previous section. However, a couple of results were different since we had a lower percentage of OD pairs solved when we perturbed the network by removing 10% of the nodes.

Table 18. Regression model parameter estimates for three network perturbation types (DV: change in circuity of solved OD pairs after perturbing 10% of the network nodes)

	BC perturbation		Elevation perturbation		Random perturbation	
	coef.	beta	coef.	beta	coef.	beta
(constant)	0.693***	-0.052	0.648***	-0.105	1.016***	-0.198
k average	0.235***	0.192	-0.004	-0.012	-0.160***	-0.235
circuity	0.093	0.009	0.380***	0.133	0.640***	0.112
intersection density	0.000*	0.059	-0.000***	-0.154	0.000***	0.104
elevation (std)	0.000	0.010	-0.001***	-0.163	0.000***	0.032
BC (max-mean)/std	0.003**	0.034	0.003***	0.096	0.011***	0.189
length total per area	-0.000	-0.049	0.000***	0.145	-0.000	-0.013
pct open space	-0.001***	-0.073	0.000**	0.030	-0.001***	-0.085
built up area	0.000***	0.036	0.000***	0.035	-0.000***	-0.071
pop density	0.000***	0.058	-0.000***	-0.076	-0.000***	-0.092
Africa (ref)						
Asia	0.039***	0.128	0.010***	0.113	0.032***	0.193
Europe	0.019	0.063	0.016***	0.186	0.113***	0.674
LAC	-0.073***	-0.240	0.009**	0.102	0.006	0.035
Northern America	0.059***	0.194	0.016***	0.184	0.007	0.044
Oceania	-0.019	-0.062	0.010	0.124	0.129***	0.766
R-squared	0.055		0.051		0.212	

N=8005; *** p<0.01, ** p<0.05, * p<0.10

The coefficient of the k average flipped to a positive sign for the BC perturbation. We argue that the result is less reliable since we may not be capturing the changes in circuity well due to disconnected networks across cities. The coefficient of the standard deviation of node elevation showed a positive and significant sign; the change in circuity is greater for cities that have a hilly topography when the network is perturbed randomly. Except for these two differences, the results were basically identical to the ones explained in the previous section.

5. Discussion

5.1. Intersection Consolidation Discussion

In practice, the most prevalent method for counting street junctions using planar centerlines can overcount complex intersections such as interchanges, roundabouts, and slip lanes. This overcount can bias our understanding of street networks that are less connected and resilient than they are in the real world. Methods like clustering, machine learning, and topological conversion have been attempted to solve the overcounting problem. However, the aforementioned approaches have the limitation of generalization due to individualized parameterization, method complexity, and labor-intensive preprocessing stages. In this study, we introduce a novel method to address this overcounting issue by applying empirical localized and global parameterizations to topologically consolidate spatially adjacent redundant nodes. Our street intersection consolidation in 8,910 cities in the world using customized parameterization (spatially matched average road width) and global parameterization (world median average road width) and our validation show that our new method achieves overall good results (with true positive rate and negative rate around 95% and 90% respectively) and is broadly generalizable.

However, the consolidation parameterization of this study is based on the average road width of 200 sampled cities from the Atlas of Urban Expansion dataset and the assumption cities located in close proximity to one another have comparable street network features. Future studies should acquire more detailed data on street width and block size in each city and adjust the parameters accordingly to attain even better correction results.

Geographically, the proposed algorithm performs best in cities in land rich developed countries where streets are wide with gridded networks. For example, as the largest land rich developed country in the sample, the US is known for its wide streets, which are historically rooted in a federal standard that required a minimum 15.2 meter right-of-way for residential streets in the 1930s (Southworth and Ben-Joseph, 2013; Millard-Ball 2022). In addition, urban sprawl of single-family detached dwellings, fueled by vehicle dependency in the early 20th century,

influenced the current low-density urban development in land-rich nations. At the start of the 21st century, there is evidence that street networks in the US have returned to somewhat more gridlike configurations (Boeing, 2020). The proposed algorithm efficiently corrects wide-street, grid-like, and low-density street networks in land-rich developed nations with little false consolidation.

However, the consolidation algorithm performs less effectively in cities in East Asia and the Pacific, with the majority of cities in China. Despite the fact that the algorithm seldom consolidates the nodes that should not be considered one intersection, approximately 20% of the time, it fails to consolidate the nodes that should be considered one intersection. Street networks in Chinese cities are categorized as a mix of major roads in urbanized modern urban areas and local roads in industrial zones and urban villages. On the one hand, major roads in China are wide, leading to large and complex intersections. The gated residential developments that obstruct public roadway access may be a contributing factor to the existence of incredibly big crossroads (Mao, 2017). As a result, to accommodate the traffic, major roads are wide and the nodes in large intersections are too far away from one another to be successfully merged together. On the other hand, streets in urban villages are narrow and dense. If we use a greater parameter to consolidate large intersections, distinct intersections in urban villages may get falsely consolidated together in the meantime. In addition, studies have revealed that OpenStreetMap street networks in Chinese cities are incomplete and of doubtful quality compared to those in western cities. Zheng and Zheng (2014) found that though on average 66% of OSM data was accurate, 71% of the OSM data was less detailed than the Baidu map, a commercial navigation map can be seen as a proxy of real-world street networks . Less complete and accurate data may also have a negative impact on the algorithm performance. Furthermore, both the OSM data quality and urban development in China are undergoing rapid change and we anticipate a better consolidation performance in Chinese cities in the future research.

As for consolidation of different types of intersections, the algorithm efficiently solved the overcounting problems caused by multiple roadways and nearby face-to-face junctions (staggered junctions). Although the algorithm could partially address overcount problems caused by large complex interchanges, roundabouts, slip lanes, it can hardly completely consolidate all the nodes of the above junctions as one due to their large spatial extent.

With a false positive rate of less than 3%, the algorithm consolidates the nodes that should be treated as one junction more than 95% of the time, making the street network model more real-world-alike. The consolidation result reveals that our previous model overestimates the real-world intersection count by 10% worldwide. The consolidated networks are more accurate and closer to the real-world street networks' representation. Using the corrected consolidated

networks may improve our understanding of how street design connects to resilience, travel behavior, and public health.

5.2. Resilience Simulation Discussion

From the resilience simulation, we identified how the results could change by correcting the street network by node consolidation. As shown in Table 19, the estimations on network disconnection may be either over or underestimated when using the street networks without consolidation. In detail, we might underestimate the impact of network attacks on nodes with high betweenness centrality, while it is also likely to overestimate the result of random attacks. Similarly, using street networks without correction may lead to either over- or under-estimation of changes in circuity due to network perturbation (Table 20). As some cities are likely to have more changes in the distribution of intersections due to consolidation, researchers focusing on those cities should be cautious when dealing with street network files for resilience analysis.

Table 19. Comparison of percentage of solved OD pairs after perturbing 10% of network nodes across region and network type

Region	Network type	Perturbation type		
		BC	Elevation	Random
World	**Original**	33.9%	69.4%	60.1%
	RW consolidated	30.0%	69.6%	62.1%
Asia	**Original**	37.5%	70.2%	60.6%
	RW consolidated	33.5%	70.3%	62.1%
Africa	**Original**	37.6%	71.0%	66.0%
	RW consolidated	34.8%	71.3%	67.5%
Europe	**Original**	24.0%	65.4%	48.0%
	RW consolidated	19.4%	65.8%	52.9%
LAC	**Original**	28.5%	70.4%	63.1%
	RW consolidated	24.9%	70.7%	64.8%
Northern America	**Original**	23.4%	63.9%	59.4%
	RW consolidated	19.3%	64.0%	61.4%
Oceania	**Original**	19.7%	62.7%	43.7%
	RW consolidated	11.9%	64.5%	52.6%

Moreover, we examined the impact of network perturbation by comparing different perturbation methods across world regions. Among the three perturbation types, attacking nodes with high betweenness centrality led to severe network disconnection and large changes in trip circuity. Random perturbation showed a moderate impact, while attacking low-lying nodes showed the least impact. Here, we note that attacking low-elevation nodes may not be an issue in some cities if there are extremely low risks of flood disasters. Across the six world regions, cities located in Europe, Northern America, and Oceania were most vulnerable to

network perturbation, whereas cities in Asia, Africa, and LAC showed relatively better resilience.

Table 20. Comparison of change in circuity of solved OD pairs after perturbing 5% of network nodes across region and network type

Region	Network type	Perturbation type		
		BC	Elevation	Random
World	**Original**	1.385	1.038	1.127
	RW consolidated	1.395	1.036	1.109
Asia	**Original**	1.380	1.036	1.126
	RW consolidated	1.395	1.034	1.108
Africa	**Original**	1.363	1.030	1.104
	RW consolidated	1.360	1.029	1.091
Europe	**Original**	1.419	1.049	1.179
	RW consolidated	1.438	1.048	1.151
LAC	**Original**	1.380	1.033	1.109
	RW consolidated	1.380	1.031	1.093
Northern America	**Original**	1.449	1.062	1.122
	RW consolidated	1.462	1.061	1.110
Oceania	**Original**	1.403	1.064	1.193
	RW consolidated	1.390	1.054	1.139

From the regression analyses, we identified features that show larger effects by perturbation methods. Figures 52 and 53 display the beta coefficients from our regression models. Most interestingly, circuity and having particularly important nodes (i.e., the number of standard deviations between the maximum and mean value of betweenness centrality) showed large effects. Both variables showed a negative impact on street network resilience. The results suggest that cities having circuitous networks or particularly important nodes are more likely to have their street network disconnected and experience an increase in trip distances when the network is attacked, regardless of the perturbation method.

On the other hand, a larger node degree turned out to be conducive to street network resilience, leading to fewer disconnected trips. Street network features such as intersection density, the standard deviation of node elevation, and the street length per area showed mixed results.

Furthermore, we acknowledge multiple limitations. First, we provide a limited understanding of the differences across regions. For instance, 51.4% of the cities are located in Asia, in which it is important to identify the differences across subregions. To add, only 0.5% of the cities are located in Oceania. Second, we conducted regression analyses based on the results from the original network. Future works may test whether the regression results change when using the

consolidated network. Lastly, there may be additional street network features that are associated with resilience. Since we have limited access to data sources from cities around the world, we only addressed the variables that are readily available.

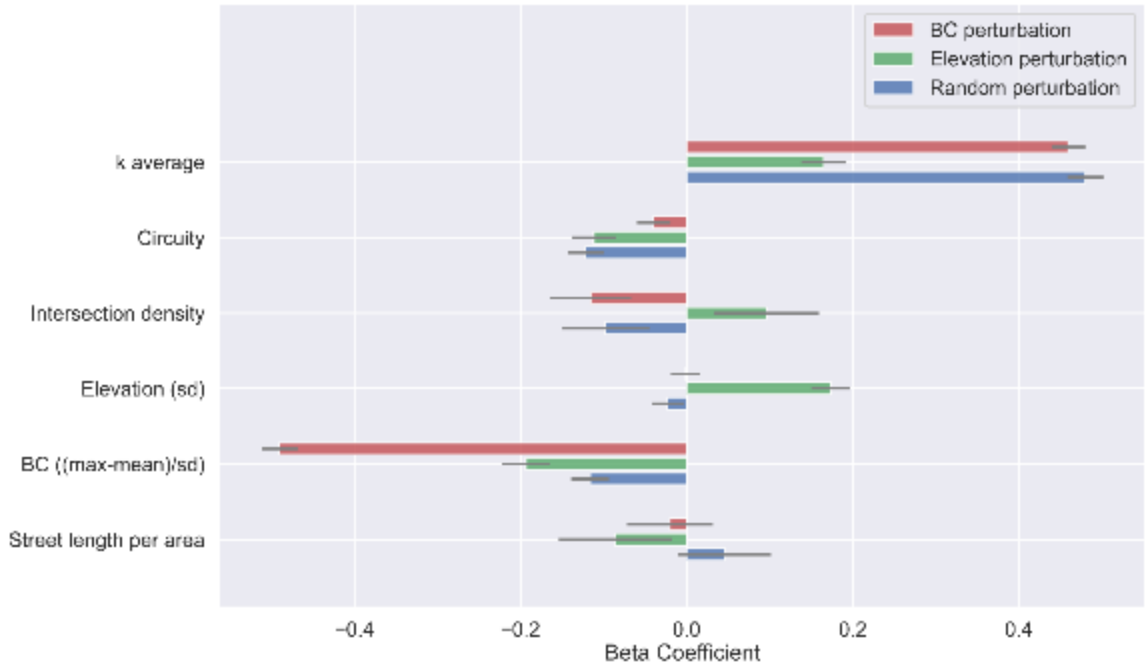


Figure 52. Comparison of beta coefficients across perturbation type (DV: percentage of solved OD pairs after perturbing 10% of the network nodes) (Note: The black line refers to the range of 95% confidence intervals)

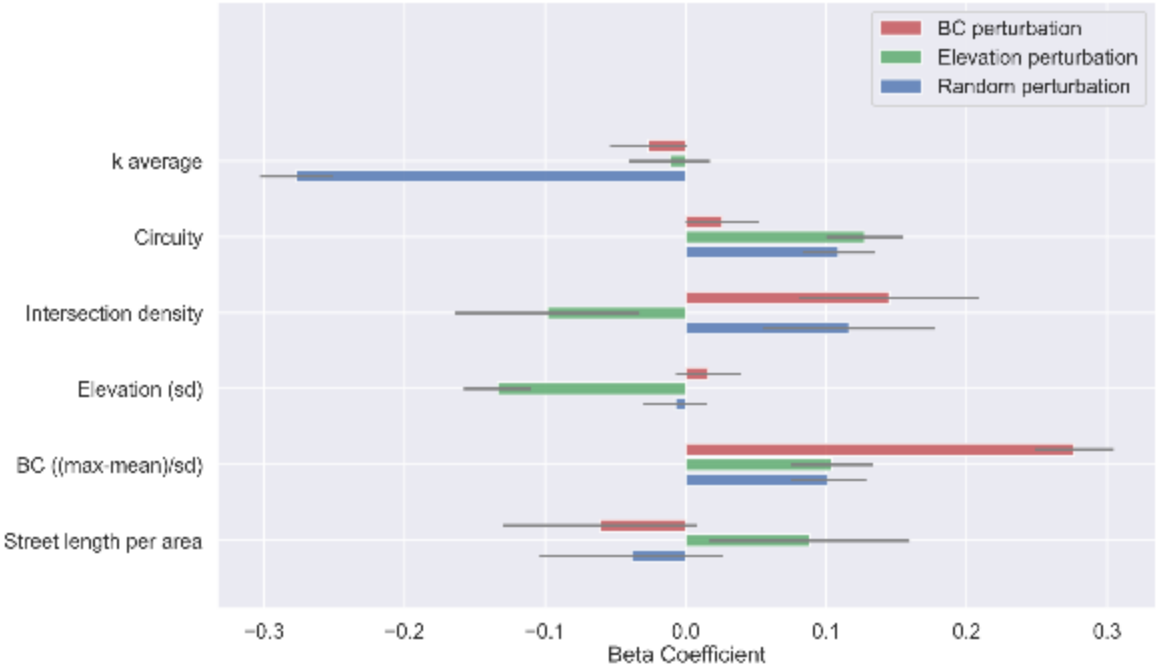


Figure 53. Comparison of beta coefficients across perturbation type (DV: change in circuity of solved OD pairs after perturbing 5% of the network nodes) (Note: The black line refers to the range of 95% confidence intervals)

6. Conclusion

This study investigated the measurement of intersection counts using spatial and topological data. It addressed the problem of overcounting intersections when using GIS centerline data, particularly at points of divided roads' intersections, roundabouts, and sliplanes. It developed an algorithm to topologically consolidate complex intersections into a single intersection. This network consolidation algorithm was then run on graph models of every urban area in the world for an empirical analysis of the phenomenon's effects, and then a sample of the resulting models was manually validated. Next, we assessed the impacts of intersection overcounting in simulations of resilience. We considered resilience from three perspectives: a random attack on the network, an elevation-based attack on the network (mimicking a flooding event), and a targeted attack on the network's most important nodes. Important, if small, differences in the results arose between the original network models versus the corrected consolidated models. In particular, these methods set up future research investigating the characteristics of networks associated with resilience to perturbation, be it random, elevation-based, or targeted to inflict maximum disruption. In sum, this study demonstrated the importance of proper representations of intersections in transportation models, which is necessary for data-driven, evidence-based urban planning to understand and support urban transportation.

References

Angel et al. (2016a), Atlas of Urban Expansion—2016 Edition, Volume 1: Areas and Densities, New York: New York University, Nairobi: UN-Habitat, and Cambridge, MA: Lincoln Institute of Land Policy.

Angel et al. (2016b), Atlas of Urban Expansion—2016 Edition, Volume 2: Block and Roads, New York: New York University, Nairobi: UN-Habitat, and Cambridge, MA: Lincoln Institute of Land Policy.

Barrington-Leigh, C., & Millard-Ball, A. (2015). A century of sprawl in the United States. *Proceedings of the National Academy of Sciences*, *112*(27), 8244-8249.

Barrington-Leigh, C., & Millard-Ball, A. (2017). More connected urban roads reduce US GHG emissions. *Environmental Research Letters*, *12*(4), 044008.

Barrington-Leigh, C., & Millard-Ball, A. (2020). Global trends toward urban street-network sprawl. *Proceedings of the National Academy of Sciences*, *117*(4), 1941-1950.

Boarnet, M. G., Nesamani, K. S., & Smith, S. (2003). Comparing the Influence of Land Use on Nonwork Trip Generation and Vehicle Distance Traveled: An Analysis using Travel Diary Data.

Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, *65*, 126-139.

Boeing, G. (2020). Planarity and street network representation in urban form analysis. *Environment and Planning B: Urban Analytics and City Science*, *47*(5), 855-869.

Boeing, G. (2021). Off the grid… and back again? The recent evolution of American street network planning and design. *Journal of the American Planning Association*, *87*(1), 123-137.

Boeing, G. (2022). Street network models and indicators for every urban area in the world. *Geographical Analysis*, *54*(3), 519-535.

Bruyns, G. J., Higgins, C. D., & Nel, D. H. (2021). Urban volumetrics: From vertical to volumetric urbanisation and its extensions to empirical morphological analysis. *Urban Studies*, *58*(5), 922-940.

Carleo, G., & Troyer, M. (2017). Solving the quantum many-body problem with artificial neural networks. *Science*, *355*(6325), 602-606.

Ewing, R., & Cervero, R. (2010). Travel and the built environment: A meta-analysis. *Journal of the American planning association*, *76*(3), 265-294.

Frizzelle, B. G., Evenson, K. R., Rodriguez, D. A., & Laraia, B. A. (2009). The importance of accurate road data for spatial applications in public health: customizing a road network. *International journal of health geographics*, *8*(1), 1-11.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

Li, C., Zhang, H., Wu, P., Yin, Y., & Liu, S. (2020). A complex junction recognition method based on GoogLeNet model. *Transactions in GIS*, *24*(6), 1756-1778.

Li, H., Hu, M., & Huang, Y. (2019). Automatic identification of overpass structures: a method of deep learning. *ISPRS International Journal of Geo-Information*, *8*(9), 421.

Mackaness, W. A., & Mackechnie, G. A. (1999). Automating the detection and simplification of junctions in road networks. *GeoInformatica*, *3*(2), 185-200.

Mattsson, L. G., & Jenelius, E. (2015). Vulnerability and resilience of transport systems–A discussion of recent research. *Transportation research part A: policy and practice*, *81*, 16-34.

Mao, X. (2017). Opening up Gated Communities to Public Road System and Urban Traffic Microcirculation: A Case Study of Donghu District in Nanchang. Journal of Landscape Research, 9(4).

Millard-Ball, A. (2022). The width and value of residential streets. *Journal of the American Planning Association*, *88*(1), 30-43.

OECD (2022), Urban population by city size (indicator). doi: 10.1787/b4332f92-en (Accessed on 26 December 2022)

Salat, S., Vialan, D., & Nowacki, C. (2010). A common metrics and set of indicators for assessing buildings and urban fabric sustainability performance. *CESB: Prague, Czech Republic*, 1-12.

Sharifi, A. (2019). Resilient urban forms: A review of literature on streets and street networks. *Building and Environment*, *147*, 171-187.

Southworth, M., & Ben-Joseph, E. (2013). *Streets and the Shaping of Towns and Cities*. Island Press.

Thomson, R. C., & Richardson, D. E. (1999). The 'good continuation' principle of perceptual organization applied to the generalization of road networks.

Touya, G. (2010). A road network selection process based on data enrichment and structure detection. *Transactions in GIS*, *14*(5), 595-614.

United Nations Population Division. 2014. World urbanization prospects: The 2014 revision. New York:

United Nations Department of Economic and Social Affairs. Online at: https://esa.un.org/unpd/wup/.

Wang, J. (2015). Resilience of self-organised and top-down planned cities—a case study on London and Beijing street networks. *PloS one*, *10*(12), e0141736.

Wang, S., Yu, D., Kwan, M. P., Zhou, H., Li, Y., & Miao, H. (2019). The evolution and growth patterns of the road network in a medium-sized developing city: A historical investigation of Changchun, China, from 1912 to 2017. *Sustainability*, *11*(19), 5307.

Xue, H., Cheng, X., Jia, P., & Wang, Y. (2020). Road network intersection density and childhood obesity risk in the US: a national longitudinal study. *Public health*, *178*, 31-37.

Yang, J., Zhao, K., Li, M., Xu, Z., & Li, Z. (2020). Identifying complex junctions in a road network. *ISPRS International Journal of Geo-Information*, *10*(1), 4.

Yu, W., Zhang, Y., Ai, T., Guan, Q., Chen, Z., & Li, H. (2020). Road network generalization considering traffic flow patterns. *International Journal of Geographical Information Science*, *34*(1), 119-149.

Zhou, Q., & Li, Z. (2015). Experimental analysis of various types of road intersections for interchange detection. *Transactions in GIS*, *19*(1), 19-41.

Data Management Plan

Products of Research

No primary data were collected for this study, only publicly accessible secondary data were used. These data are documented here, alongside persistent links to their permanent locations.

Data Format and Content

1. Urban Street Networks

1.1 Global Urban Street Network Model

DATA STORAGE:
- Data Format: GraphML
- Find data at: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/KA5HJ3

DATA DESCRIPTION:
- Pre-existing data collected by a third-party
- Data description: Topologically corrected 8,914 urban street networks across 178 countries collected using OSMnx based on OpenStreetMap (OSM) data with elevation information attached to nodes. This is our primary data source. We used a proposed algorithm to consolidate nodes in the street network models, and further used the original and consolidated network models for resilience analysis.

METADATA INFORMATION
- Title: Street Network Models and Indicators for Every Urban Area in the World.
- Year: 2021
- Author: Geoff Boeing, boeing@usc.edu, University of Southern California
- Link to the original metadata information:
https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/WMPPF9
where metadata-graph-edges.tab and metadata-graph-nodes.tab contain the metadata information for edges and nodes in the street network models respectively

1.2 GHS-UCDB R2019A - GHS Urban Centre Database 2015, multitemporal and multidimensional attributes

DATA STORAGE:
- Data Format: excel spreadsheet (xls)
- Find data at: https://data.jrc.ec.europa.eu/dataset/53473144-b88c-44bc-b4a3-4583ed1f547e

DATA DESCRIPTION:
- Pre-existing data collected by a third-party
- Data description: The data include geometric centroid of cities around the world. We matched the centroid coordinates to the corresponding studied 8,914 cities via unique ID.

METADATA INFORMATION
- Title: GHS Urban Centre Database 2015, multitemporal and multidimensional attributes
- Year: 2019
- Authors: Florczyk, Aneta; Corbane, Christina; Schiavina, Marcello; Pesaresi, Martino; Maffenini, Luca; Melchiorri, Michele; Politis, Panagiotis; Sabo, Filip; Freire, Sergio; Ehrlich, Daniele; Kemper, Thomas; Tommasi, Pierpaolo; Airaghi, Donato; Zanchetta, Luigi, jrc-ghsl-data@ec.europa.eu
- Original metadata information:
embedded in the index tab of the data GHS_STAT_UCDB2015MT_GLOBE_R2019A_V1_0.xls file

1.3 Global Street Network Indicator

DATA STORAGE:
- Data Format: CSV
- Find data at: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ZTFPTB

DATA DESCRIPTION:
- Pre-existing data collected by a prior study
- Data description: Street network and socio-economic indicators of 8,914 urban street networks in the world. The indicators include: intersection counts, node counts, average node degree, total resident population, areas within urban center boundary polygons, built-up surface area, and etc.

METADATA INFORMATION
- Title: Global Urban Street Networks Indicators
- Year: 2021
- Author: Geoff Boeing, boeing@usc.edu, University of Southern California
- Link to the original metadata information:
https://dataverse.harvard.edu/file.xhtml?fileId=4288605&version=2.0

2. Urban Atlas of Expansion

2.1 Atlas of Urban Expansion—2016 Edition, Volume 1: Areas and Densities

DATA STORAGE:
- Data Format: XLSX (Areas_and_Densities_Table_1.xlsx)
- Find data at: http://atlasofurbanexpansion.org/data

DATA DESCRIPTION:
- Pre-existing data collected by a third-party
- Data description: The Atlas of Urban Expansion collects and analyzes data on the quantity and quality of urban expansion in a stratified global sample of 200 cities. This data includes information on the area of urban extent, urban built up area, urbanized open space, and etc. in around 1990, 2000, and 2015 for the sampled 200 cities.

METADATA INFORMATION
- Title: Angel et al.,Atlas of Urban Expansion—2016 Edition, Volume 1: Areas and Densities, New York: New York University, Nairobi: UN-Habitat, and Cambridge, MA: Lincoln Institute of Land Policy, 2016.
- Year: 2016
- Author: Shlomo Angel, Alejandro M. Blei, Jason Parent,Patrick Lamson-Hall, and Nicolás Galarza Sánchez with Daniel L. Civco, Rachel Qian Lei and Kevin Thom

2.2 Atlas of Urban Expansion—2016 Edition, Volume 2: Block and Roads

DATA STORAGE:
- Data Format: XLSX (Blocks_and_Roads_Table_1.xlsx)
- Find data at: http://atlasofurbanexpansion.org/data

DATA DESCRIPTION:
- Pre-existing data collected by a third-party
- Data description: The Atlas of Urban Expansion collects and analyzes data on the quantity and quality of urban expansion on a stratified global sample of 200 cities. This dataset includes information on average road width, density of all arterial roads (km/km2), and other blocks and roads indicators for urban areas built pre-1990 and urban areas built between the 1990s and 2015 in a sample of 200 cities.

METADATA INFORMATION:
- Title: Angel et al., Atlas of Urban Expansion—2016 Edition, Volume 2: Block and Roads, New York: New York University, Nairobi: UN-Habitat, and Cambridge, MA: Lincoln Institute of Land Policy, 2016.
- Year: 2016
- Author: Shlomo Angel, Patrick Lamson-Hall, Manuel Madrid, Alejandro M. Blei, and Jason Parent, with Nicolás Galarza Sánchez and Kevin Thom

3. Countries shapefiles

DATA STORAGE:
- Data Format: Shapefiles and csv
- Find data at: https://hub.arcgis.com/datasets/esri::world-countries-generalized/explore?location=-0.112555%2C0.000000%2C2.64

DATA DESCRIPTION:
- Pre-existing data collected by a third-party
- Data description: This dataset is provided by ArcGIS Hub and represents generalized boundaries for the countries of the world. Each country has a two digit ISO code embedded in the attribute table. We accessed the shapefiles on July 13rd, 2022, and there were 249 countries listed in the shapefile. Then we manually added the world regions (defined by the Atlas of Urban Expansion) the country belongs to in the attribute table for the following analysis.

METADATA INFORMATION

- Title:World Countries (Generalized)
- Year: 2022
- Author: Esri Data and Maps
- License: licensed under the Esri Master License Agreement.
- Link to the original metadata and license information:
https://services.arcgis.com/P3ePLMYs2RVChkJx/arcgis/rest/services/World_Countries_(Generalized)/FeatureServer/0
https://www.arcgis.com/sharing/rest/content/items/2b93b06dc0dc4e809d3c8db5cb96ba69/info/metadata/metadata.xml?format=default&output=html

Data Access and Sharing

These data are available as described above from their original creators.

Reuse and Redistribution

Restrictions on reuse and redistribution are defined by each dataset's original creator as defined above.