ABSTRACT

Title of Thesis:	TOPOLOGICAL ANALYSIS OF DISTANCE WEIGHTED NORTH AMERICAN RAILROAD NETWORK: EFFICIENCY, ECCENTRICITY, AND RELATED ATTRIBUTES
	Sherief M. Elsibaie
	Master of Science 2023
Thesis Directed By:	Professor Bilal M. Ayyub,

The North American railroad system can be well represented by a network with 302,943 links (track segments) and 250,388 nodes (stations, junctions, and waypoints), and other points of interest based on publicly accessible geographical information obtained from the Bureau of Transportation Statistics (BTS) and the Federal Railroad Administration (FRA). From this large network a slightly more consolidated subnetwork representing the major freight railroads and Amtrak was selected for analysis. Recent improvements in network and graph theory and improvements in all-pairs shortest path algorithms make it more feasible to process certain characteristics on large networks with reduced computation time and resources. The characteristics of networks at issue to support network-level risk and resilience studies include node efficiency, node eccentricity, and other attributes derived from those measures, such as network arithmetic efficiency, network geometric central node, radius, and diameter, and some

distribution measures of the node characteristics. Rail distance weighting factors, representing the length of each rail line derived from BTS data, are mapped to corresponding links, and are used as link weights for the purpose of computing all pair shortest paths and subsequent characteristics. This study also compares the characteristics of North American railroad infrastructure subnetworks divided by Class I carriers, which are the largest railroad carriers classified by the Surface Transportation Board (STB) by annual operating revenue, and which together comprise most of the North American railroad network. These network characteristics can be used to inform placement of resources and plan for natural hazard and disaster scenarios. They relate to many practical applications such as network efficiency to distribute traffic and a network's ability to recover from disruptions. The primary contribution of this thesis is the novel characterization of a detailed network representation of the North American railroad network and Class I carrier subnetworks, with established as well novel network characteristics.

TOPOLOGICAL ANALYSIS OF DISTANCE WEIGHTED NORTH AMERICAN RAILROAD NETWORK: EFFICIENCY, ECCENTRICITY, AND RELATED ATTRIBUTES

by

Sherief M. Elsibaie

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2023

Advisory Committee:
Professor Bilal M. Ayyub Chair
Professor Paul Schonfeld
Professor Nii O Attoh-Okine

Acknowledgements

I would like to extend my immense gratitude and appreciation for my advisor, Professor Bilal Ayyub. Your insightful feedback, patience, and encouragement have not only fostered my intellectual growth but have also instilled in me a passion for excellence and a commitment to pushing the boundaries of knowledge. Your unwavering dedication to your students and your profound expertise in our field have been instrumental in shaping my education and research.

I would also like to extend my heartfelt thanks to my research colleague, Yujie Mao. Your collaborative spirit, tireless work ethic, and shared dedication to our research endeavors have made our educational journey together both enriching and rewarding. Your contributions have been invaluable, and I am grateful for the camaraderie and support you have provided.

To my family, whose unwavering belief in my abilities and endless reservoir of wisdom have been a guiding light in my life.

To my dearest wife, Alexandra, your support, love, and understanding have been my greatest source of strength. Your encouragement, especially during challenging times, has been the foundation of all my achievements.

Additionally, I would like to express my heartfelt thanks to my colleagues at the Federal Railroad Administration (FRA), Dr. Tarek Omar, Ms. Karen McClure, and Dr. Khaled E. Zaazaa and the National Institute of Standards and Technology (NIST), Dr. Juan Fung, Dr. Dustin Cook Dr. Jennifer F. Helgeson, Dr. Christina Gore and Dr. David Butry for their generous financial support and invaluable work opportunities. Their contributions have played a pivotal role in shaping my academic and professional endeavors. I wish to also thank my thesis committee members, Professor Nii O. Attoh-Okine and Professor Paul Schonfeld for volunteering their time and effort to look over my work.

Table of Contents

List of Figures

Chapter 1: Development of Network and Graph Theory on Eccentricity and Efficiency

1.1 Introduction

A network is any connected system where connections can be represented by links, and connection points can be represented by nodes. A link defines a connection between two nodes and can also carry properties such as distance or other characteristics such as condition, capacity, reliability, etc. Systems may be mapped or characterized as networks if they are structured appropriately, for example a railroad system of rail lines and junctions. This approach may have its roots when Leonhard Euler (1726) characterized the Seven Bridges of Königsberg as such a network, or a graph. The methodologies used in that problem, rather than the specifics of it, are often credited with the expansion of the field of graph theory and network analysis.

Ranking nodes on a network by centrality or another computational measure can help characterize a network's shape, structure and size in a more objective way and assist with planning, design, and operations. Alphabet Inc. (formerly Google) successfully used a form of network analysis to deliver more relevant search results using the network representation of the internet where webpages are nodes and hyperlinks are directional links. That network analysis has been expanding into other fields, wherever a system can be represented as a network as described above (Gleich, 2015). For a physical network, such as a railroad network, a central location is required for the effective placement of disaster response or distribution centers, or other centrally focused network infrastructure. Network extensions of different shapes

can also reduce the effectiveness of existing central infrastructure, by reducing their centrality. Past work in network and graph theory has focused on discretizing physical networks, social networks, and citation networks, among others, into mathematically defined networks (M. E. J. Newman, 2001). The objective is often to derive objective importance rankings, network properties, or more generally to enhance the language used to describe and characterize networks.

The network characteristics of interest in this thesis that are utilized in the analysis of Railroad networks are those based on a network property called the "all-pairs shortest path matrix." This matrix is determined by repeated executions of a search algorithm designed to find the shortest path and its associated length between two nodes on a network. Once computed, the matrix contains the length of the shortest path between every pair of nodes of the network. The algorithm used is mainly Dijkstra's algorithm, efficiently implemented by the "igraph" code library (West, 2000). Performance of these algorithms has been greatly improved over time by making tradeoffs between memory usage and compute time and in the implementation of other advanced coding techniques beyond the scope of this thesis (Peixoto, 2020).

1.2. Eccentricity

The mathematician Camille Jordan formalized one of the measures of network centrality used in this thesis, now called "eccentricity" (Jordan, 1869). This measure is computed for each node as "node eccentricity" and is, informally, the length of the shortest path required starting from that node to reach the furthest possible node from

it on the network. The most central node or nodes are those with a minimum eccentricity as defined in subsequent sections.

Applications of eccentricity were explored in modern graph and network theory, including applications on network reductive problems. Harary (1969) codified eccentricity as a centrality measure in his paper "Graph Theory". Later, Hage and Harary (1995) demonstrated some applications to anthropologic networks. They projected trade routes on chains of the Marshall Islands in Micronesia as networks and found that during a period in their history in the early 20th century, only the most central nodes identified by Jordan's methods "[predicted] the politically and symbolically most important islands in each network".

As eccentricity is a node property with potential variation, there has been interest in statistical measures of the distribution of node eccentricity across a network, hence termed network eccentricity. Takes and Kosters (2013) provide evidence that network eccentricity is "a relevant descriptive property" of networks, and they compute and compare the eccentricity distribution across a few different human social networks. Network shape and size description parameters "radius" and "diameter" were defined as the minimum eccentricity and maximum eccentricity across a network, respectively (M. Newman, 2010). Dankelmann, et al. (2004) explored the average eccentricity of a network, characterized various bounds on average eccentricity for different standard network shapes, and explored the best-case and worst-case effect of removing an edge on the average eccentricity of a network.

1.3. Efficiency

Another property of network nodes of interest herein has been termed "node efficiency" (Latora & Marchiori, 2001). Node efficiency, and the related distribution of all node efficiencies on a network, "network efficiency," are very closely related to node eccentricity and network eccentricity as they are also based on the shortest path length between nodes, but there are key differences. Node efficiency is the average of the reciprocal (normalized) of the shortest paths required starting from that node to reach all other node from it on the network. These terms will be defined more formally later. Additionally, A novel ranking measure based on network efficiency, without normalization, is proposed and formalized in this thesis, to address scaling problems associated with traditional network efficiency and the use of network links with a weight or length of less than one, which appear quite frequently in railroad data.

Efficiency and eccentricity describe different connectivity characteristics and shape characteristics for a network and can help communicate differences between networks, and nodes within a network. For example, to characterize nodes of a network and rank them to determine the location with the best average response time to all locations on the network, node efficiency should be used. To characterize nodes of a network and rank them to determine the location with the best worst-case response time to all possible locations on the network, node eccentricity should be used. Networks can also be ranked based on their network efficiency. Additionally, minimum and maximum network eccentricity measures, also known as network

radius and diameter, can help describe a network's geometric size and enhance typical size descriptions which may be based on the number of links and nodes. The minimum and maximum efficiency of a network and even the related skew of the distribution of node efficiencies can describe the structure of a network, for example the relative dominance of long, sparse chains or dense grids in the network's structure.

1.4. Applications to Railroads

Since 1995, The U.S. Department of Transportation (DOT) Bureau of Transportation Statistics (BTS) together with the Federal Railroad Administration (FRA) has published data on the North American Railroad network links and nodes (*Geospatial at the Bureau of Transportation Statistics*, n.d.). The data in full is comprised of 250,321 nodes and 302,865 links as of October 2022 and represents survey and submitted data by railroad operators and their regulators. This thesis explores network eccentricity, network efficiency, and both of their distributions' parameters for several networks. Several papers laid the groundwork for this exploration, including an analysis by Saadat et al. (2019) of the application of network analysis to determine criticality or centrality of nodes and links of the Washington D.C. Metro rail transit network. Mao (2021) furthered the application of network analysis to a simplified network representation of the North American Railroad network for a similar purpose.

This thesis presents simple networks and associated computations for illustration of concepts, and network analysis for the North American Railroad networks using data

referenced above from BTS. Reports on several large Railroad operators are compiled, illustrating for each the most efficient node and its location, the least eccentric node and its location, the network radius and diameter, average node efficiency and several other related properties which may help to gain insights for each network severally and collectively as the entire network covering the US, Canada, and Mexico.

Chapter 2: Network Characteristic Definitions

2.1. Foundations

Eccentricity and other related network characteristics are defined in compliance to and building on prior work on network efficiency that inspired this thesis, including Hamed et al. (Hamed et al., 2022). A network's topology, G, abbreviated as $G(n, m)$ is defined by *n* nodes and *m* links. Based on their size and connections, networks have computable characteristics that may be used for evaluation purposes. They are described formally under separate headings in subsequent sections.

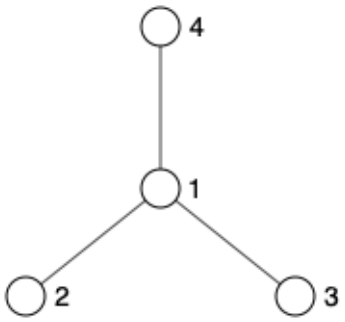
2.2. Connectivity and Adjacency

One critical network characteristic is the connectivity arrangement of links among nodes. Connectivity can be represented by the adjacency matrix *A* of $n \times n$ size, which has elements a_{ij} as its elements for column *i* and row *j* (M. E. J. Newman, 2004). Element a_{ij} describes a direct link between nodes *i* and *j* as follows:

$$a_{ij} = \begin{cases} 1 & \text{if node } i \text{ is linked to node } j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The term $link_{ij}$ describes the link between nodes *i* and *j*. In the case where $i = j$, $link_{ij}$ is called a self-link, i.e., the link starts from a node and ends with the same node. In the absence of self-links, matrix *A*'s diagonal elements have values of zero. Self-links represent an uncommon connection type in real physical networks and are therefore not considered in this thesis. For networks with links that do not reflect directionality, i.e., undirected links, the adjacency matrix *A* is symmetric, and in the case of networks with links that reflect directionality, i.e., directed links, matrix *A* is non-symmetric. For example, in the case where there is only a link from node 1 to

node 2, element a_{12} in matrix A equals 1 while element a_{21} equals 0 since no link is directed from node 2 to 1. In this schema multiple direct links from i and j can lead the value of a_{ij} to be greater than one. Matrix A for a topology of n nodes and m links $G_1(n = 4, m = 3)$ in Figure 1(a) is shown in Figure 1(b). Subsequent paragraphs build on topology $G_1(4,3)$ as an example for demonstration purposes.



$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

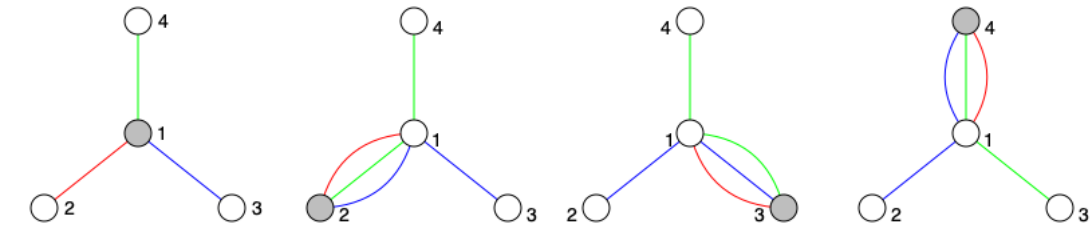
(a) Network topology $G_1(4,3)$ (b) Adjacency matrix A for G_1

Figure 1. Network topology example

2.3. Shortest Link Paths

The number of links along the shortest possible path connecting nodes i and j is considered the shortest path (p_{ij}) (M. Newman, 2010). Typically, a variation of one of many All-Pairs-Shortest-Path (APSP) algorithms is used to search for the shortest paths on more complex networks (Seidel, 1995). The code library used in this thesis, "igraph", selects an algorithm for best performance based on network characteristics; Dijkstra's algorithm is primarily used for railroad network analysis given the sparseness of most railroad networks (Csardi, 2023). These shortest path lengths can be represented in a matrix P with elements p_{ij}. Figure 2 shows the piecewise

construction of matrix P for each node for the topology G_1. Like adjacency matrix A, diagonal elements on matrix P are typically zero because the shortest path from an element to itself is zero, unless other assumptions are made.



$$P = \begin{bmatrix} 0 & - & - & - \\ 1 & - & - & - \\ 1 & - & - & - \\ 1 & - & - & - \end{bmatrix} \quad P = \begin{bmatrix} 0 & 1 & - & - \\ 1 & 0 & - & - \\ 1 & 2 & - & - \\ 1 & 2 & - & - \end{bmatrix} \quad P = \begin{bmatrix} 0 & 1 & 1 & - \\ 1 & 0 & 2 & - \\ 1 & 2 & 0 & - \\ 1 & 2 & 2 & - \end{bmatrix} \quad P = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 \\ 1 & 2 & 0 & 2 \\ 1 & 2 & 2 & 0 \end{bmatrix}$$

(a) (b) (c) (d)

Figure 2. Shortest path matrix (P) analysis and construction for each grey highlighted node for topology $G_1(4,3)$ with shortest paths shown in link counts, color matched in red, blue and green

2.4. Node Degree

The number of nodes adjacent or linked to node *i* is defined as the node degree k_i. k_i can computed from the adjacency matrix A, as the count of identity (1) values on column i. Node degree k_i values for all nodes are included in vector K with size $1 \times n$. The k_i value can be an indicator of the importance of a node in terms of its influence on the whole network and the number of links attached to it. For example, a node with a low k_i value has less influence and is attached to fewer links compared to another node with high k_i value that is attached to more links. The average number of

links connected to the nodes in a network is the average node degree ($\bar{K}$) calculated as follows:

$$\bar{K} = \sum_{i=1}^{n} \frac{k_i}{n} \quad (2)$$

Node degree is not used extensively as a network characteristic in this thesis but may be used in describing certain network language or as a comparison point with other computed characteristics.

2.5. Network Density

Network density D, sometimes also called graph density, is another network characteristic related to node degree. The density of a network is computed by taking the number of links on a network m and dividing by the number of total possible links as the combination of 2 out of *n*, i.e., $\binom{n}{2}$.

$$D = \frac{m}{\binom{n}{2}} \quad (3)$$

Density is not discussed extensively in this thesis's network analysis but is used to describe the network topology's degree of grid-like structure. Those networks which are more grid like are considered denser, and networks with fewer grid like connections are called sparse.

2.6. Component Networks

For railroad networks and other networks, component network structures are disjointed or otherwise not connected in any way to each other in the BTS data set. An example of a network with two components is shown in Figure 3. Component structures are commonly found in the BTS railroad network for independent railroad

lines, some short lines, or possibly other situations. In these cases, component networks must be filtered and analyzed separately because they will have no path between nodes to other components in the networks. Typically, there exists a dominant component on each railroad network, and subnetwork, which comprises most of the nodes and links. The dominant component is selected by the greatest size by node count and smaller components are excluded although can be examined in future studies or by other researchers. In Figure 3, Component 2 would be selected for analysis as the dominant component of the disjoint network.

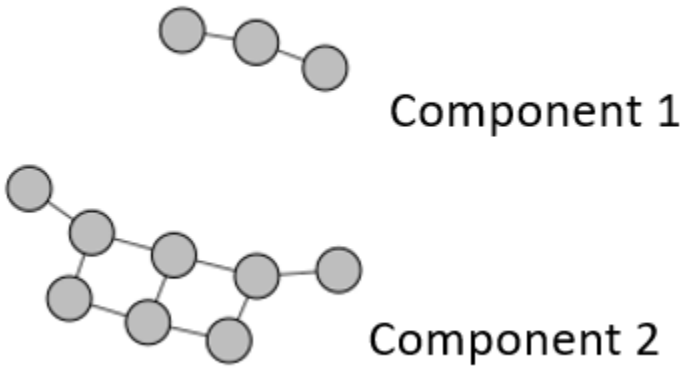


Figure 3. A network with two disjoint components

Chapter 3: Network Attributes and Associated Weights

3.1. Overview of Network Attributes

On a physical network, such as a railroad network, some links (i.e., lines) are longer in distance than others, and it is desirable to factor this into the network analysis. Several other railroad specific line attributes, such as track class, allowable travel speed, proximity to critical infrastructure, signaling, and presence of train automation could also differentiate or add information to parts of a railway network (Hamed et al., 2022). Analysis of these physical networks under the schema described thus far would not be able to account for differing link attributes along the network. Therefore, a system of incorporated weights for network link elements and shortest path computation is proposed which can account for these differences.

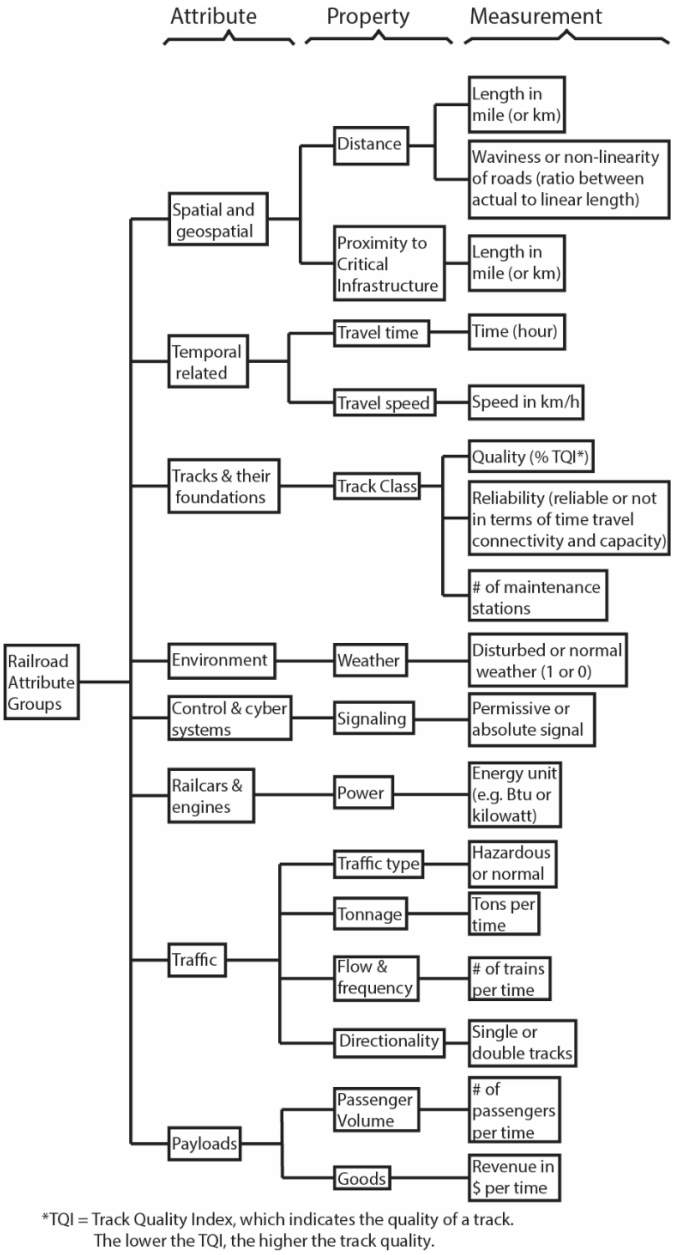


Figure 4. Railroad attribute hierarchy adapted after (Hamed et al., 2022)

Early work by Newman (2004) on weighted networks suggested that weighted network analysis had received less attention than unweighted networks. Unweighted network characteristic computation is notably simpler and computationally faster than that of a weighted network. The most widely used APSP algorithms needed

modification; otherwise they were ineffective for analyzing weighted networks similar to the work of Chan (2010) for unweighted networks.

The desire to include weights on network attributes can be traced back to some of the first papers in network theory, on scientific paper citation networks (M. E. J. Newman, 2001). Newman, attempting to answer the question of who the best-connected researcher was, argued that co-authorship relationship strength was stronger if there were fewer authors on a paper. Using this extra information to weigh the citation networks resulted in a more nuanced and detailed network analysis. If attributes for potential weight factors are not included in an analysis of a network, potentially relevant information is being ignored, and the resulting analysis may not reflect the true nature of the network.

3.2. Link Weights

An example of an arbitrarily link weighted topology G_1 from Figure 1 is shown in Figure 5. The link weights are used as link labels and are colored purple to distinguish them from node labels. The link weights are considered as link lengths or link costs for the purposes of network traversal and shortest path computation.

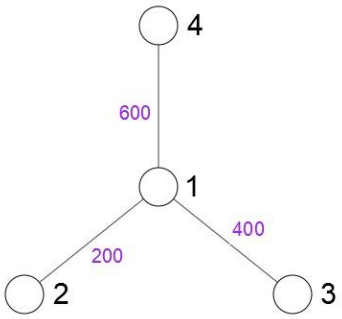


Figure 5. $G_1(4,3)$ Network topology example, with arbitary link weights

A link weight applied to $link_{ij}$, called w_{ij}, is considered the length of that link in the set of links that make up a path in the computations of a shortest path on a network. For example, on Figure 3, $w_{12} = 200$, and therefore traversing $link_{12}$ is the equivalent of traversing 200 unweighted links in the computation of a path length. The total link weights of all links along a path is the length of that path. For our analysis of railroad networks, only the geodesic line length in kilometers or miles are used as link weights. As more data becomes available for research, some of the other railroad attributes may be incorporated into a weighting system.

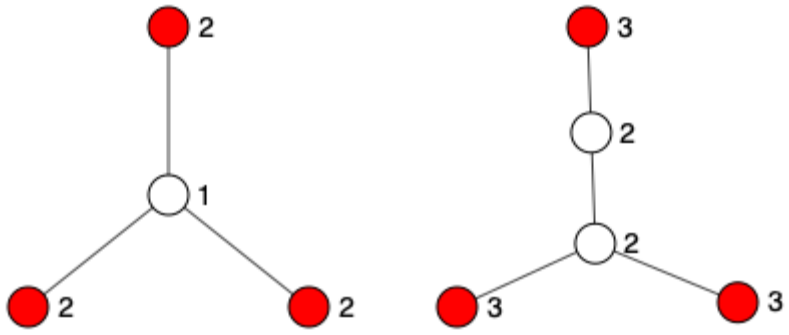
Chapter 4: Node and Network Eccentricity Characteristics

4.1. Node Eccentricity

Node eccentricity for a particular node in a network is the maximum of the set of shortest path distances from that node to all other nodes in the network. Representing all the shortest path distances for all the nodes by matrix P such as in Figure 2, the node eccentricity for node 1 is the maximum value of the first column or generally for node i, node eccentricity e_i is defined as follows:

$$e_i = \max\left(p_{i1}, p_{i2}, \dots, p_{ij}\right) \tag{4}$$

Node eccentricity is a worst-case scenario travel distance, measured in a link unit count. Some focused work on eccentricity refers to a matrix constructed with node eccentricities as an "anti-adjacency" matrix, in contrast to the adjacency matrix shown on Figure 1 (Wang et al., 2018). In terms of connectivity, nodes with higher eccentricity are usually more peripheral, while those with lower eccentricity are usually more central. Node eccentricity can be added as a label for each node on a network for a graphical representation. Figure 6(a) shows a small, unweighted network with the node eccentricity for all nodes as labeled by numeric values. The central node by eccentricity is colored white and has eccentricity of 1, and peripheral nodes are colored red. The central node can be verified by inspection. From it, only one link is required to reach the furthest point possible from it in the network as opposed to other nodes where two links are required. Figure 6(b) shows the case where there are two central nodes, when both are tied for minimum node eccentricity across all nodes in the network.



(a) *G*(4,3) (b) *G*(5,4)

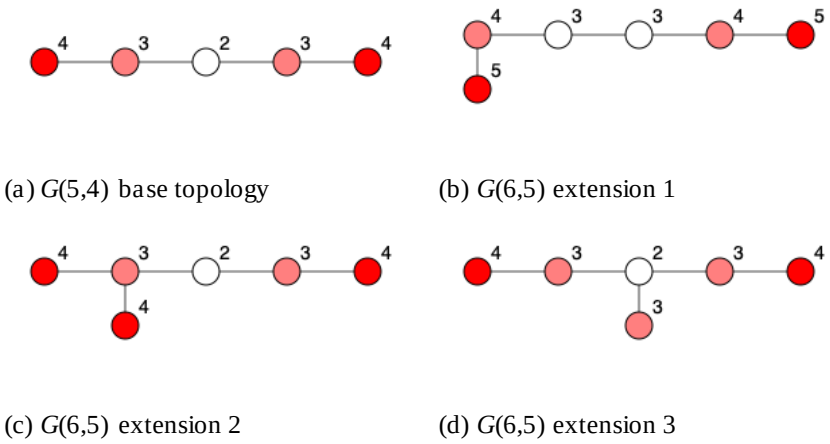
Figure 6. Example topologies $G(n, m)$ with node eccentricities as noted

4.2. Network Eccentricity

Eccentricity is a node characteristic, but the eccentricity distribution of a network, defined earlier as network eccentricity, can be characterized by the statistical measures on the set of all node eccentricities, such as the minimum, maximum, average, variance and skewness. These distribution parameters and the effect of different network shapes on them are explored in the next sections.

The examples in Figure 7 show the effect of adding one node with one link to a series of different positional nodes on the node eccentricities of an initial *G*(5,4) linear topology network. Each node is colored with different shades of red based on its node eccentricity. Higher node eccentricity is darker red, and lower node eccentricity is lighter red. Different expansions have different effects on node eccentricity throughout the network, and therefore its distribution. Adding peripheral nodes has the greatest effect on node eccentricity, causing a shift of the central nodes and

altering the node eccentricity of almost all other nodes. Adding central nodes has minimal or no effect on other node's eccentricity.



(a) $G(5,4)$ base topology (b) $G(6,5)$ extension 1

(c) $G(6,5)$ extension 2 (d) $G(6,5)$ extension 3

Figure 7. $G(5,4)$ Example linear network with node eccentricities noted, and three cases of $G(6,5)$ extensions

4.2.1. Network Radius and Diameter

The radius of a network is defined as the minimum eccentricity out of all node eccentricities (M. Newman, 2010). This is closely analogous to the shape description of the radius of a circle, which similarly is the maximum distance from the center or central point to the outer edge. On Figure 6(a) the radius of the network is equal to 1, and on Figure 6(b) the radius of the network is equal to 2. A network with a smaller radius will have some node or set of nodes which can more easily reach all other nodes.

The diameter of a network is its maximum eccentricity (M. Newman, 2010). $G(4,3)$ of Figure 6(a) has a diameter of 2, whereas $G(5,4)$ of Figure 6(b) has a diameter of 3. This again is closely analogous to the diameter of a circle, which is the distance from

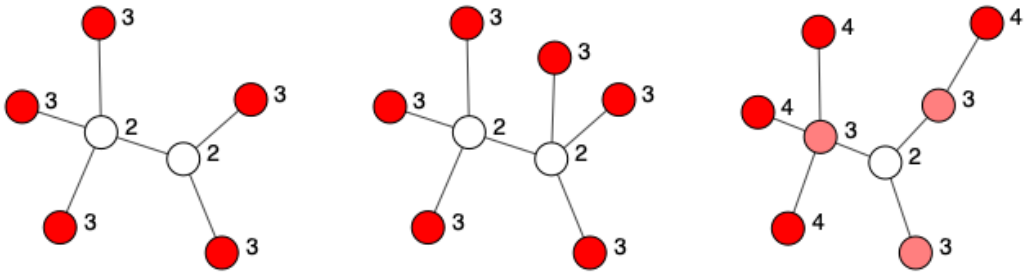
one point on the outermost edge to the furthest point possible also on the outermost edge. A network with a larger diameter will have at least 2 nodes that are farther apart than the most separated pair of nodes on the network with smaller diameter. The worst-case travel time or travel cost along the network with larger diameter is therefore higher.

4.2.2. Network Eccentricity Average, Variance, Skewness

The mean, variance, and skewness can be computed on network eccentricity to characterize the distribution of node eccentricities. These properties, their meanings, and their computation methods are well defined and can communicate important characteristics of network eccentricity and can be used as a basis to compare different networks.

Figures 8(a) to 8(c) plots the node eccentricities and network eccentricity statistical characteristics for different spoked or web networks which are approaching more typical network shapes. Figure 8(a) shows an initial $G(7,6)$ node eccentricity and network eccentricity statistics. Figures 9(b) and 9(c) show two different one node and one link expansions of Figure 8(a). The first one node and one link expansion on Figure 8(b) does not expand the periphery, and therefore has minimal effect on network eccentricity. The expansion at Figure 8(c) expands the periphery, which impacts the node eccentricity of most of the nodes on the network. The network eccentricity is therefore affected to a greater extent. Figures 9(a) and 10(b) show that this trend is maintained for more central and peripheral network expansions of larger topologies, respectively. Figure 9(c) represents a mix of the two expansion types and

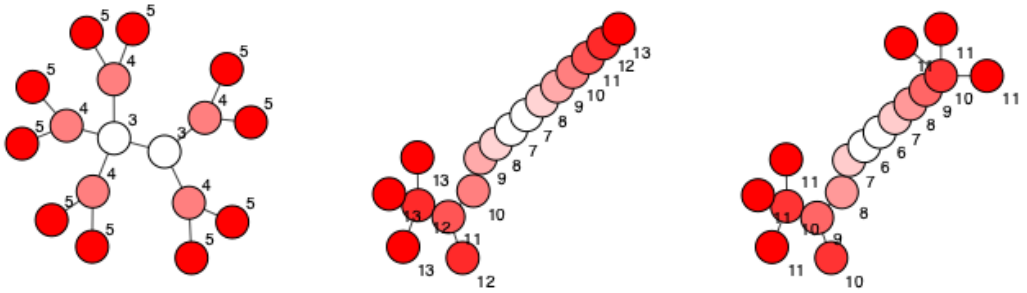
a more realistic connection pattern found in railroads and human travel routes, for example connecting two cities at distance with grid structures expanding access around high population or industrial areas. The central node by eccentricity is almost directly in between the two "cities." Node eccentricity increases monotonically traveling away from the central node in all cases.



(a) $G(7,6)$ (b) $G(8,7)$ (c) $G(8,7)$

(2, 3, 2.71, 0.20, -1.23) (2, 3, 2.75, 0.19, -1.44) (2, 4, 3.38, 0.48, -0.82)

Figure 8. Example node eccentricities and statistics on $G(7,6)$ compared to $G(8,7)$ expansions (radius, diameter, average, variance, skewness)



(a) $G(17,16)$ (b) $G(17,16)$ (c) $G(17,16)$

(3, 5, 4.47, 0.48, -1.03) (7, 13, 10.47, 4.37, -0.31) (6, 11, 9.18, 3.20, -0.56)

Figure 9. Example node eccentricities and statistics of $G(17,16)$ expansions of $G(7,6)$ (radius, diameter, average, variance, skewness)

Chapter 5: Node Efficiency Characteristics

5.1. Node Pairwise Efficiency

Node pairwise efficiency is a scaled measure of how closely two nodes, i and j where $i \neq j$, are linked. The best-case linkage distance possible, b, is divided by the shortest path distance, p_{ij} between the two nodes. For an unweighted network where link distances are assumed to be unitary, b takes a value of one, as the minimum link connection between nodes is for them to be directly linked by one link unit (Latora and Marchiori, 2001). Node pairwise efficiency pef_{ij} is therefore defined:

$$pef_{ij} = \frac{b}{p_{ij}} \text{ where } i \neq j \quad (5)$$

For the best-case network connection, the node pairwise efficiency is equal to one. For an unweighted network, the second-best network connection case is one where the shortest path distance between a pair of nodes is two links. The node pairwise efficiency in this case is equal to $\frac{1}{2}$. This trend would continue so that as the length of the shortest path distance between two nodes increases, their pairwise efficiency approaches zero.

5.2. Issues with Weighted Node Efficiency

For Railroad data using line distances as link weights, it is the case that some link weights are arbitrarily small, well below a unitary one mile or one kilometer. In different subnetworks of railroads, therefore the value of b can also be quite different, resulting in a different scaling factor being assigned to the efficiency of different subnetworks, unless other assumptions are made regarding the value of b.

Our proposal to avoid these problems is to drop the normalization and reciprocal in the original definition of node efficiency. To clarify, this new measure is defined as "node arithmetic inefficiency". The term inefficiency is used as in this schema lower node arithmetic inefficiency values are more connected on average to all other nodes. With this new definition, the pairwise inefficiency between two nodes *i* and *j* is simply p_{ij}. This also relates the two main network characteristics, eccentricity and efficiency, better and more closely to our original intention. That intention was that eccentricity represents the all-pairs worst case travel path (which it does without modification), and that network efficiency represents the all-pairs average travel path. The units of both network measures would then be the same and are equivalent to the units of the link weights.

Chapter 6: Node and Network Arithmetic Inefficiency Characteristics

6.1. Node Pairwise and Node Arithmetic Inefficiency

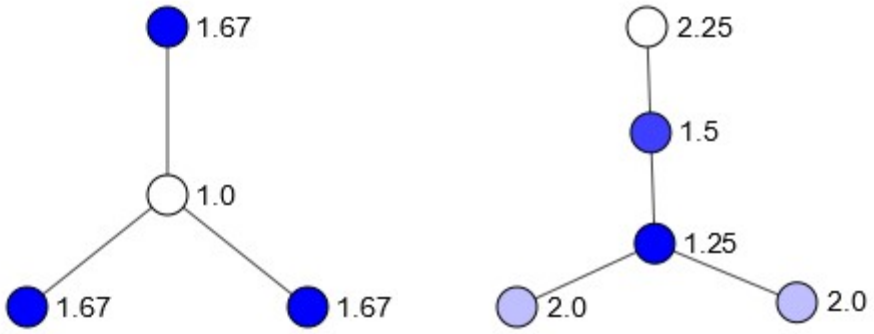
As previously stated, Node Pairwise Arithmetic Inefficiency (i.e., *paief*) between nodes *i* and *j*, is simply the shortest path between them, p_{ij}. Formally:

$$paief_{ij} = p_{ij} \text{ where } i \neq j \quad (6)$$

Node arithmetic inefficiency is a cumulative property of node pairwise arithmetic inefficiencies. The node arithmetic inefficiency for node i is the average of all node pairwise arithmetic inefficiency values between node i and other nodes. Formally:

$$ef_i = avg(paief_{i1},\ paief_{i2}, \dots, paief_{i(i-1)}, paief_{i(i+1)}, \dots, paief_{ij}) \quad (7)$$

Nodes with a lower arithmetic inefficiency on average will be able to reach all other nodes on the network with a smaller travel distance, but there may be outliers. Generally, more central nodes connected in a grid-like pattern will have a greater node efficiency than those in long chains or at the periphery of a network, but there are nuances. Figure 10(a) shows a small, unweighted network with the node efficiencies for all nodes as labeled by numeric values. The most efficient node on each network is colored white, and less efficient nodes are shaded darker blue. Node efficiency coloring on Figure 10(b) highlights some key differences emerging in efficiency centrality computations and coloring from the same network with eccentricity computations and coloring at Figure 6(b). There is no longer a tie in centrality measurement between the two more central nodes, and a better efficiency (lower inefficiency) is computed for the node with a better average connection to each node.



(a) *G*(4,3) Topology (b) *G*(5,4) Topology

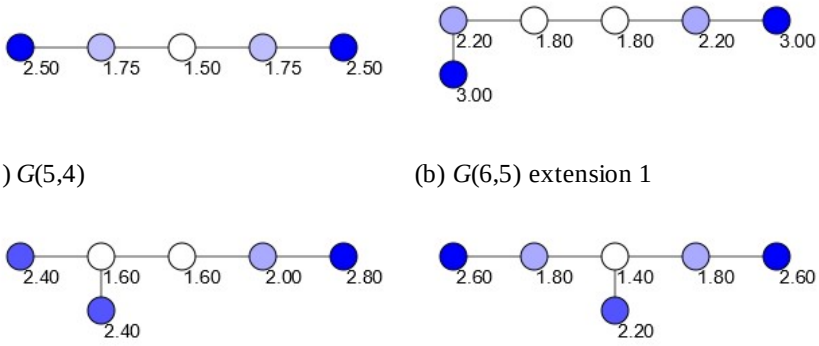
Figure 10. Example topologies $G(n,m)$ with node arithmetic inefficiencies as noted

6.2. Network Arithmetic Inefficiency

Once exhaustively computed, the distribution of node arithmetic inefficiencies on the whole network can be used to represent the network arithmetic inefficiency. Network arithmetic inefficiency can be characterized by the statistical measures on the set of all node arithmetic inefficiencies, such as the minimum, maximum, average, variance, and skewness which again are detailed in later examples.

To illustrate the effect of different types of network extensions on the set of node arithmetic inefficiencies, the same line network and line network expansions shown with node eccentricity distributions in Figure 7, are shown in Figure 11 with node arithmetic inefficiency distributions and blue coloring. The coloring of Figure 11(c) highlights differences between it and Figure 7(c). The most efficient nodes are pulled closer to the semi-peripheral extension, as opposed to the least eccentric node which is unchanged by the extension. Network topologies with more grid-like topology

structures, or greater density, tend to have greater influence on node arithmetic inefficiency than on eccentricity.



(a) $G(5,4)$ (b) $G(6,5)$ extension 1

(c) $G(6,5)$ extension 2 (d) $G(6,5)$ extension 3

Figure 11. $G(5,4)$ Example line network with node efficiencies with 3 cases of $G(6,5)$

6.2.1. Minimum and Maximum Node Arithmetic Inefficiency

The range of values of node inefficiencies for a network describe its general topology. In comparing two networks, the existence of a less efficient node on one of the networks indicates that network contains structures with longer linear or loop portions and more sparsely connected grid structures, whereas the existence of a more efficient node implies the opposite. Figure 10(a) illustrates the range of node arithmetic inefficiencies is $[1, 1.67]$. When one link and one node is added at Figure 10(b), the minimum and maximum are both increased, and the range becomes $[1.25, 2.25]$. The line extension networks show some of the same patterns. Addition of one terminal node on the line network from Figure 11(a), shown at Figure 11(b) increases both the maximum and minimum of node arithmetic inefficiencies from $[1.50, 2.50]$ to $[1.80, 3.00]$. A line network is a poor configuration of nodes and links if efficiency is to be prioritized. The less linear expansions of Figure 11(c) results in a smaller

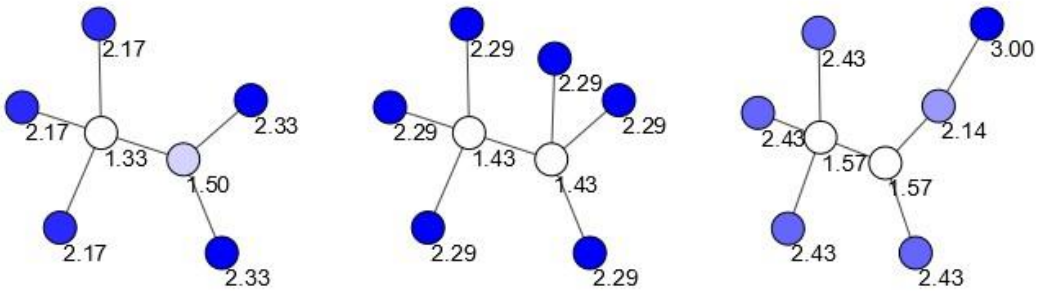
increase in the maximum and minimum arithmetic inefficiency, from $[1.50, 2.50]$ to $[1.60, 2.80]$. Finally, the most central expansion possible illustrated in Figure 11(d) results in a range in inefficiency of $[1.40, 2.60]$. The most central expansion still increases maximum inefficiency slightly, but also decreases minimum inefficiency.

6.2.2. Network Arithmetic Inefficiency Mean, Variance, and Skewness

The moments, such as the mean, variance, and skewness, can be computed for all node arithmetic inefficiencies on a network to characterize the distribution of network arithmetic inefficiency. Figure 12 shows the network inefficiency moments and node inefficiency labels and colorings for the cases of network topologies of $G(7,6)$ and $G(8,7)$ to demonstrate similar centrality trends as the eccentricity case illustrated in Figure 8. The differences can be explained by the greater influence of denser, grid-like network topologies on arithmetic inefficiency versus eccentricity. Eccentricity's focus on worst-case shortest path lengths as opposed to arithmetic inefficiency's focus on average shortest path lengths can result in denser structures not influencing computation as much for eccentricity unless a dense structure extends the worst-case shortest path for a node. The impact of central expansions of Figure 12(a) shown on Figure 12(b) has a more significant impact on network arithmetic inefficiency than it did for network eccentricity at Figure 8(a) versus Figure 8(b). The average arithmetic inefficiency was changed from 2.00 to 2.07, a 3.5% increase, whereas average node eccentricity was only increased from 2.71 to 2.75, an approximate 1.5% increase on the same expansion. Peripheral expansion at Figure 11(c) still causes greater effects on network arithmetic inefficiency than central expansions. The average node arithmetic inefficiency is increased by a larger margin, 12.5%. The minimum

inefficiency is increased, from 1.33 to 1.57, and maximum increased from 2.33 to 3.00.

Larger expansions of $G(17,16)$, starting at Figure 13(a) continue to show similar trends, as well as continuing to show the added impact of larger and denser grid-like network topologies. The most efficient node on Figure 13(a), or the one with smallest arithmetic inefficiency, has a higher node degree and is linked to 4 separate travel chains while the second most efficient node is only linked to 3 chains. Figure 13(b) demonstrates the strength of the impact of the denser grid structure on dictating which node is most efficient. Along the purely peripheral linear expansion, the most efficient node is still closer to the original grid structure at Figure 12(a). This contrasts with eccentricity at Figure 9(b), where the peripheral expansion greatly impacted the least eccentric node, pulling it more in the direction of the expansion. Finally, the mixed case network structure expansion at Figure 13(c) demonstrates competing trends. The most efficient nodes on this network are in between the two opposite grid topologies but are closer to the larger original denser grid topology. This contrasts with eccentricity at Figure 9(c) where the least eccentric nodes are further away from the original denser grid topology.



(a) $G(7,6)$

(1.33, 2.33, 2.00, 0.143, -1.14)

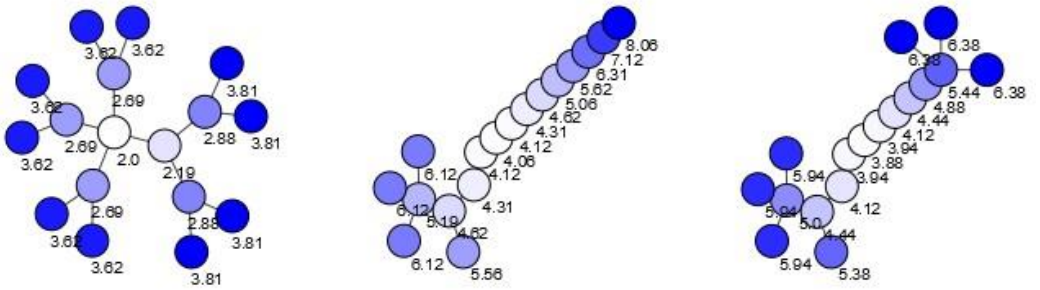
(b) $G(8,7)$

(1.43, 2.29, 2.07, 0.138, -1.44)

(c) $G(8,7)$

(1.57, 3.00, 2.25, 0.203, -0.31)

Figure 12. Example node arithmetic inefficiencies and statistics on G(7,6) vs G(8,7) expansions (min, max, mean, variance, skewness)



(a) $G(17,16)$

(0.32, 0.58, 0.39, 0.0072, 0.86)

(b) $G(17,16)$

(0.21, 0.41, 0.32, 0.0025, -0.11)

(c) $G(17,16)$

(0.27, 0.41, 0.33, 0.0025, 0.023)

Figure 13. Example node efficiencies and statistics on $G(17,16)$ expansions of $G(7,6)$ (min, max, mean, variance, skewness)

Chapter 7: North American Railroad Network

A network representation of the North American Railroad system is made available by the FRA and BTS together with geographical information. The full network consists of 302,943 links, 250,388 nodes, covering almost the entire United States, and parts of Mexico and Canada. After filtering out abandoned railroad lines, abandoned railroad lines than have been physically removed, lines that are out of service, yard tracks, tourist lines, museum lines, science passenger service lines, and trails on former railroad right-of-way and selecting only those lines used by Class I carriers (whether by ownership or track rights), the size of the network is reduced to 142,462 links and 130,971 nodes, with 792 component networks. Class I carriers are the largest classification of railroad carriers, and are defined by statute and the Surface Transportation Board (STB) as railroad carriers which have an annual operating revenue of $900 million or more in 2019 dollars, defined by a deflator formula in the statute (Railroad Companies, 1887). The dominant component was chosen which contains 139,170 links and 126,993 nodes, and the other 791 components were discarded. The total length of all track in this network is 242,500 km. As detailed earlier, actual physical line lengths, also contained in the BTS data, are used as weights for each link for all analysis.

7.1. North American Railroad Weighted Network Eccentricity

The visualized node eccentricities of the full network described above are shown on Figure 14. Greater values of node eccentricity are colored darker red, while lower values are lighter red.

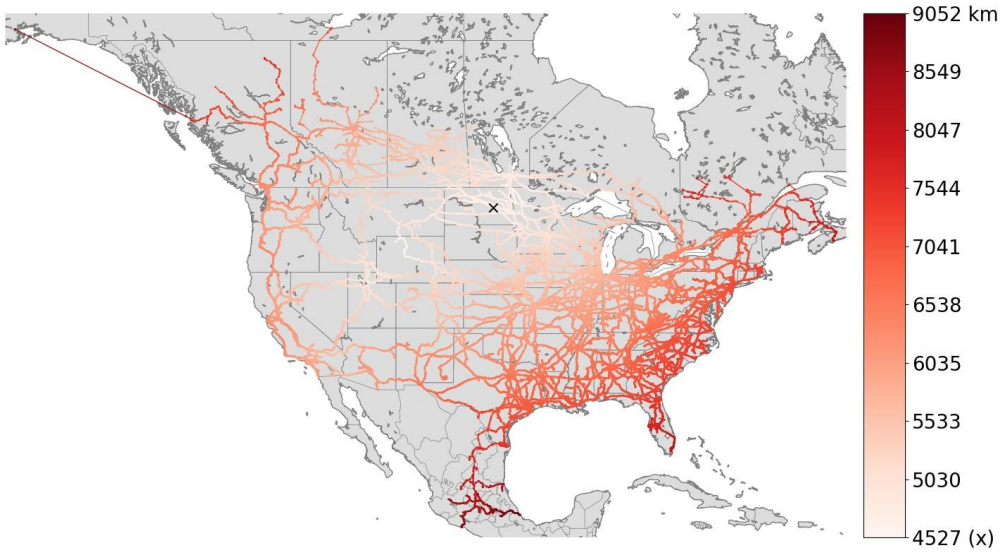


Figure 14. Weighted node eccentricity of all Class I Carriers combined

The central node by eccentricity is FRA node ID 351220, marked by an "x" on Figure 14, and is in Foster County, North Dakota. FRA node IDs are unique identifiers assigned to each node in the BTS data set. The minimum eccentricity across the network, which is at that central node is 4,527 km. This quantity also defines the radius of the network. The maximum eccentricity is 9,052 km, and that node eccentricity is shared by the two most peripheral nodes in Alaska and Mexico. This quantity also represents the diameter of the network. Finally, reportable parameters of the distribution of node eccentricities, shown in Figure 15, include the average node eccentricity as 6,240 km, median of 6,201 km, variance of 512,331 km^2, and skewness of -0.032.

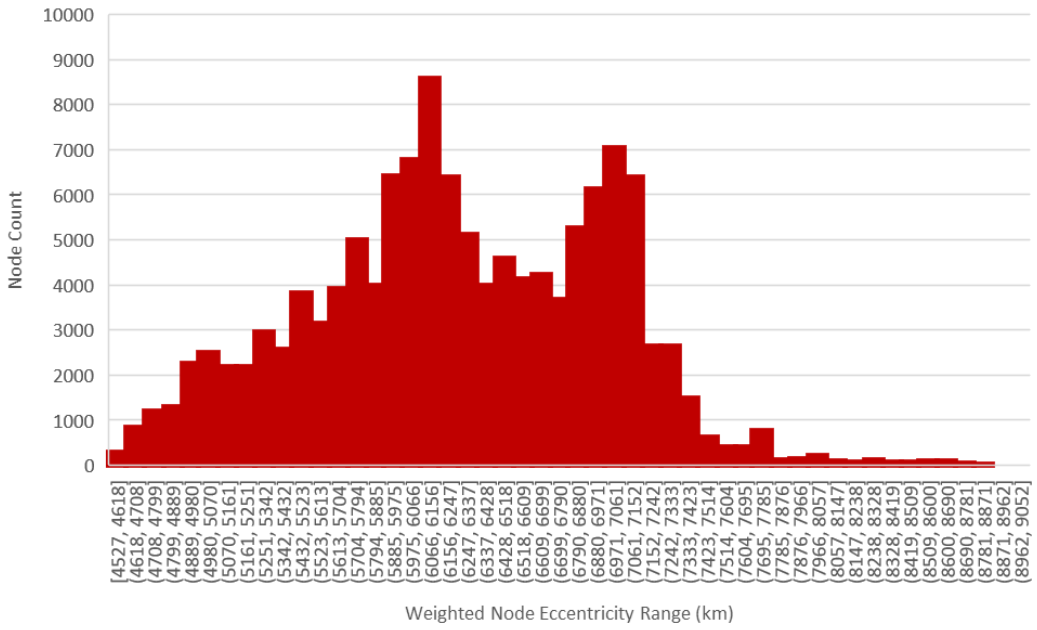


Figure 15. Histogram of weighted node eccentricity of all Class I Carriers combined.

7.2. North American Railroad Weighted Network Arithmetic Inefficiency

A colormap of node arithmetic inefficiencies of this full network are shown on Figure 16. Greater values of node arithmetic inefficiency are colored darker blue, while lower values are colored lighter blue.

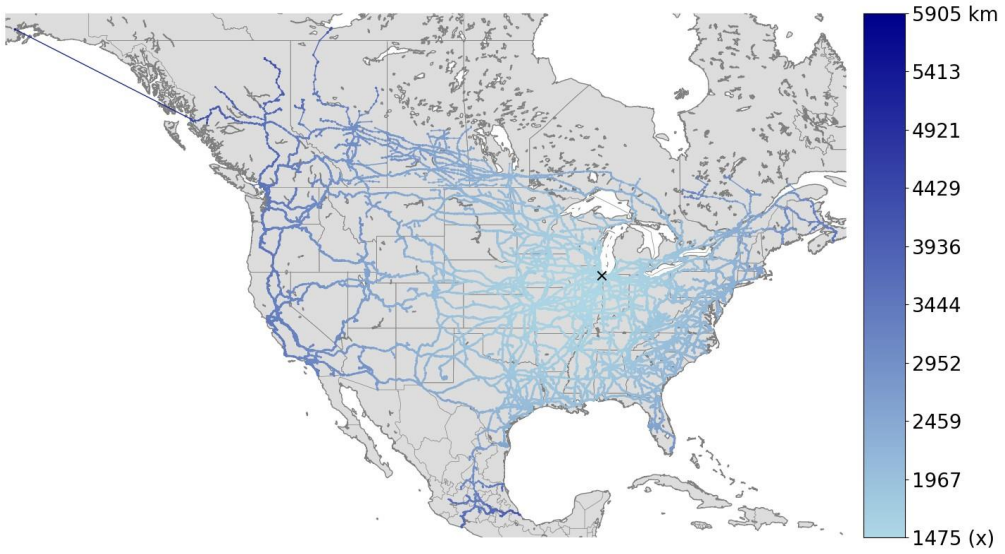


Figure 16. Weighted node arithmetic inefficiency of all Class I Carriers combined

The node with best arithmetic efficiency (lowest inefficiency) is FRA node ID 414270, marked by an "x" on Figure 16, and is in Cook County, Illinois. The arithmetic inefficiency of that node is 1,474 km. The maximum arithmetic inefficiency is 5,905 km, at the furthermost peripheral node in Alaska. Finally, reportable parameters of the distribution of node arithmetic inefficiency, shown in Figure 17, include the average node arithmetic inefficiency as 2,143 km, median of 1,979 km, variance of 335,196 km^2, and skewness of 1.093.

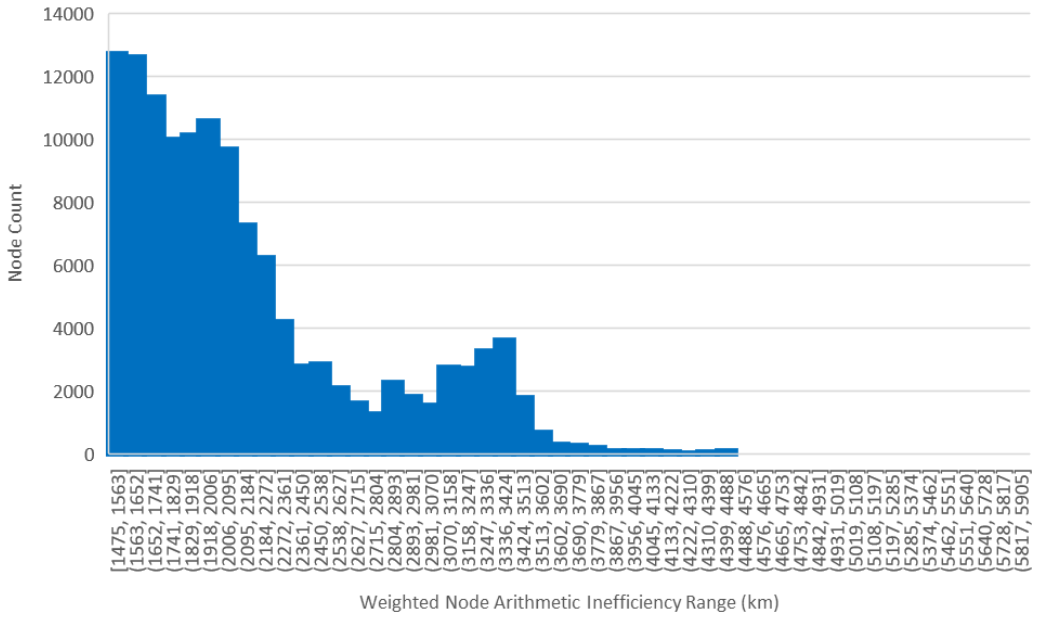


Figure 17. Histogram of weighted node arithmetic inefficiency of All Class I Carriers Combined.

Chapter 8: Class 1 Railroad Operators

As stated earlier, North American railroad carriers are classified by the STB by their annual operating revenue, and Class I carriers are the largest. The Association of American Railroads (AAR) indicates that Class I carriers account for 94% of the revenue of goods transported in North America by rail, and according to the BTS data all together own or have rights to the majority of track (Association of American Railroads, 2023). There are seven entities that qualify as a Class I freight carrier. They are, in no specific order, BNSF Railway (formerly Burlington Northern and Santa Fe Railway), CSX Transportation (formerly Chessie System and Seaboard System Railroad), Canadian Pacific Railway, Kansas City Southern Railway Company, Norfolk Southern Railway, Canadian National Railway Company, and Union Pacific Railroad. All seven have track in the United States, two have substantial rail lines extending into Canada, and one has substantial rail lines that extend into Mexico. There is a long history of changes in ownership structures as well for these railroad carriers, so this information is subject to change. There is also a passenger railroad carrier, Amtrak, that would qualify as a Class I carrier if they carried freight, but they will be included in this analysis for comparison purposes. The Class I carrier subnetworks are analyzed similarly to the full railroad network, and similar moments are generated and displayed in the following sections. The Class I subnetworks were constructed using BTS data and included any track that was not filtered according to the prior criteria for inclusion in the full network. For each Class I carrier a line of rail was included in the network if it was owned or co-owned by the carrier, or if the carrier had track rights recorded for that line of track.

8.1. BNSF Railway Network

The BNSF Railway network chosen according to the previously discussed criteria consists of 32,297 links and 29,783 nodes, with 214 component networks. The dominant component was chosen which contains 31,297 links and 28,604 nodes, and the other 213 components are discarded. The total length of all track in this network is 58,477 km. As detailed earlier, actual physical line lengths, also contained in the BTS data, are used as weights for each link for all analysis.

8.1.1. BNSF Railway Weighted Network Eccentricity

The visualized node eccentricities of the BNSF Railway subnetwork described previously are shown on Figure 18. Greater values of node eccentricity are colored darker red, while lower values are lighter red.

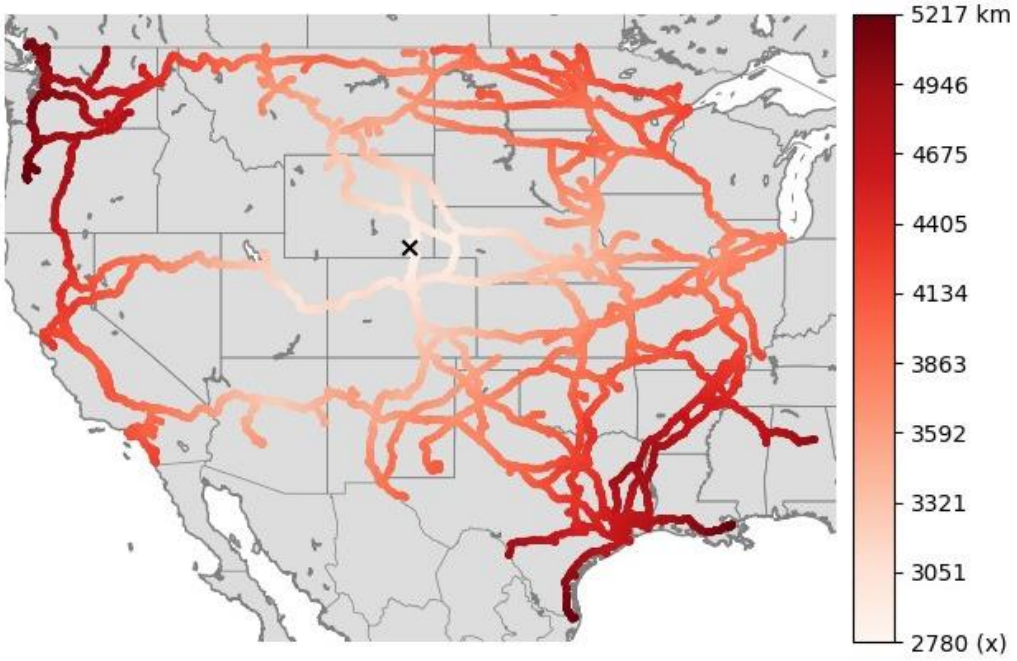


Figure 18. Weighted node eccentricity of the BNSF Railway Subnetwork

The central node by eccentricity is FRA node ID 340326, marked by an "x" on Figure 18, and is in Laramie County, Wyoming. The minimum eccentricity across the network, which is at that central node is 2,779 km. This quantity also defines the radius of the network. The maximum eccentricity is 5,216 km, and that node eccentricity is shared by the two most peripheral nodes in Oregon and Louisiana. This quantity also represents the diameter of the network. Finally, reportable parameters of the distribution of node eccentricities, shown in Figure 19, include the average node eccentricity as 4,060 km, median of 4,038 km, variance of 281,866 km^2, and skewness of 0.14.

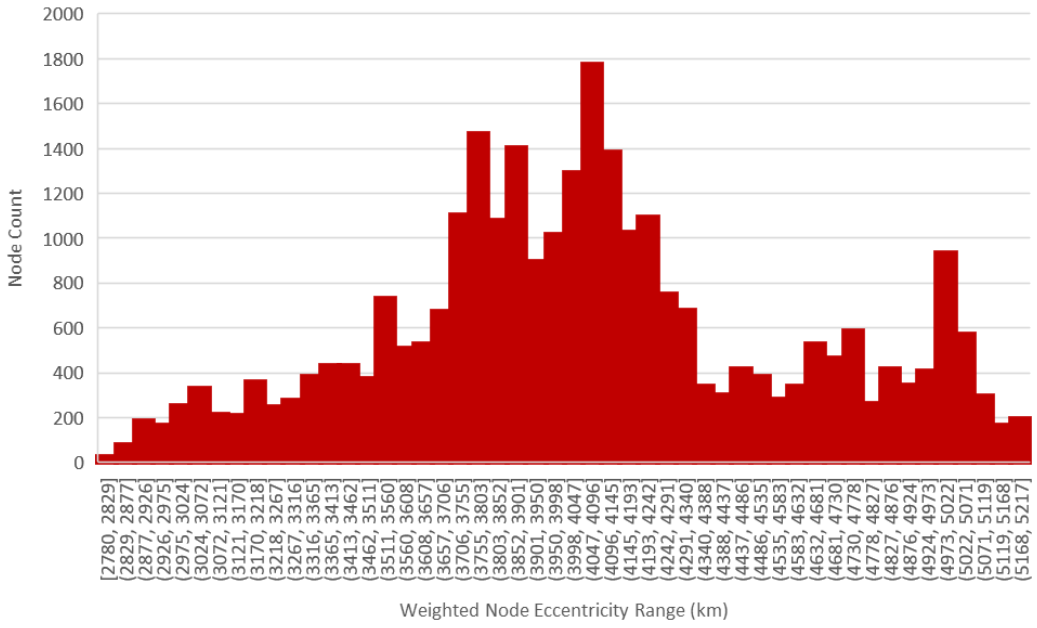


Figure 19. Histogram of Weighted node eccentricity of the BNSF Railway Subnetwork

8.1.2. BNSF Railway Weighted Network Arithmetic Inefficiency

A colormap of node arithmetic inefficiencies of the BNSF subnetwork are shown on Figure 20 Greater values of node arithmetic inefficiency are colored darker blue, while lower values are colored lighter blue.

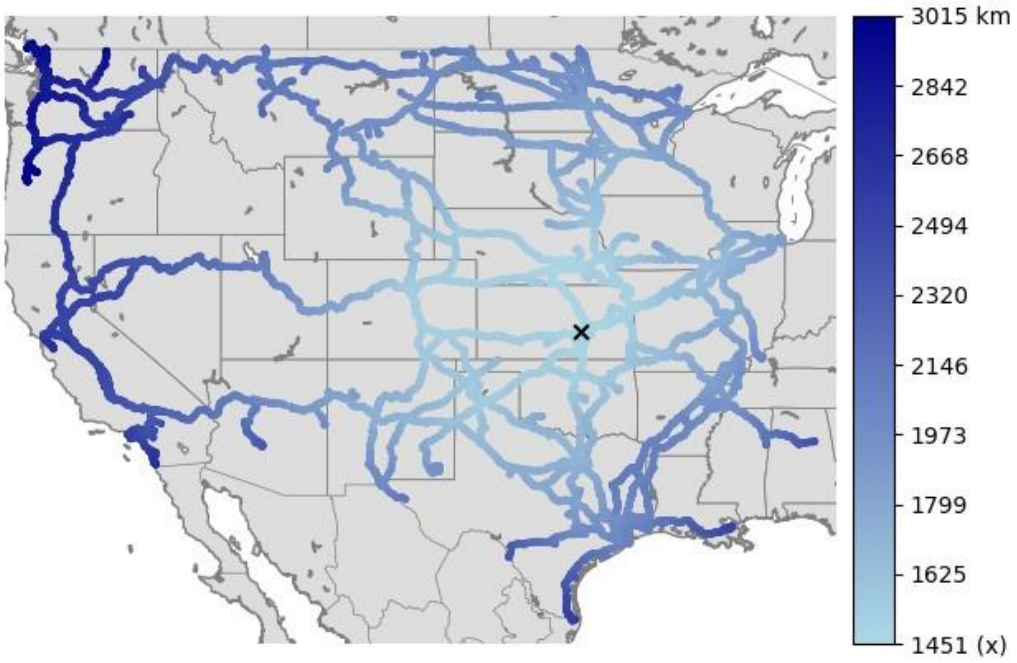


Figure 20. Weighted node arithmetic inefficiency of the BNSF Railway Subnetwork

The node with best arithmetic efficiency (lowest inefficiency) is FRA node ID 359509, marked by an "x" on Figure 20, and is in Marion County, Kansas. The arithmetic inefficiency of that node is 1,451 km. The maximum arithmetic inefficiency is 3,015 km, at the furthermost peripheral node in Oregon. Finally, reportable parameters of the distribution of node arithmetic inefficiency, shown in Figure 21, include the average node arithmetic inefficiency as 1,983 km, median of 1,873 km, variance of 175,090 km^2, and skewness of 0.717.

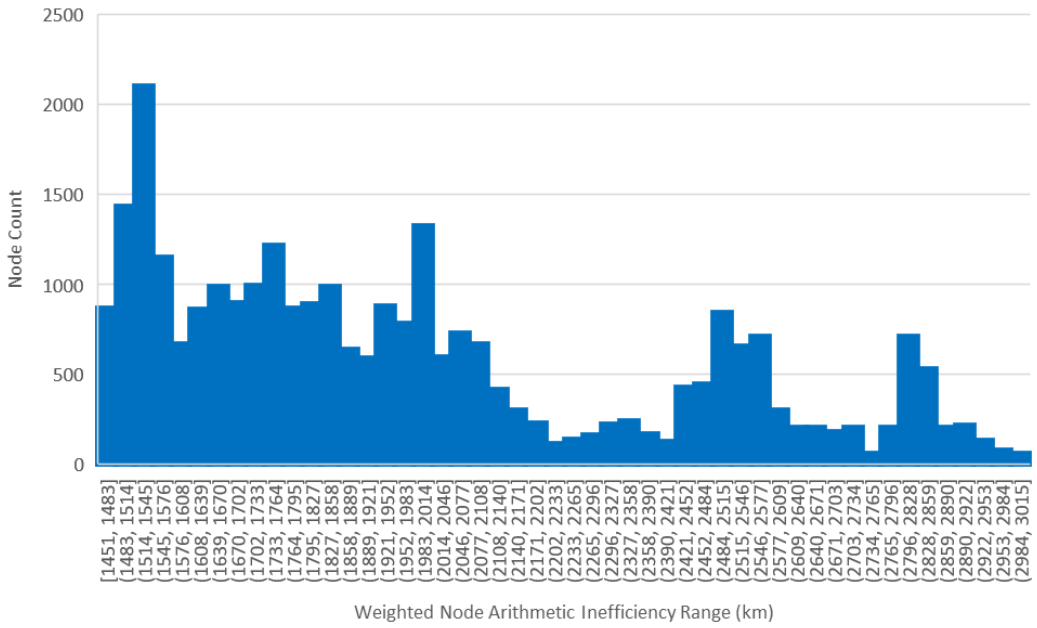


Figure 21. Histogram of Weighted node arithmetic inefficiency of the BNSF Railway Subnetwork

8.2. CSX Transportation Network

The CSX Transportation network chosen according to the previously discussed criteria consists of 25,827 links and 24,524 nodes, with 144 component networks. The dominant component was chosen which contains 25,248 links and 23,821 nodes, and the other 143 components are discarded. The total length of all track in this network is 37,447 km. As detailed earlier, actual physical line lengths, also contained in the BTS data, are used as weights for each link for all analysis.

8.2.1. CSX Transportation Weighted Network Eccentricity

The visualized node eccentricities of the CSX Transportation subnetwork described previously are shown on Figure 22. Greater values of node eccentricity are colored darker red, while lower values are lighter red.

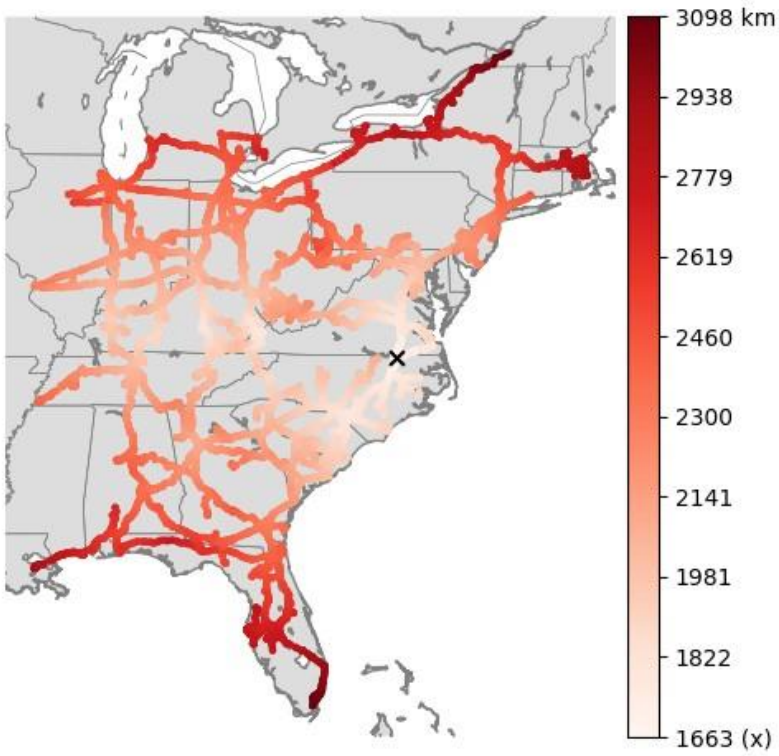


Figure 22. Weighted node eccentricity of the CSX Transportation Subnetwork

The central node by eccentricity is FRA node ID 473751, marked by an "x" on Figure 22, and is in Northampton County, North Carolina. The minimum eccentricity across the network, which is at that central node is 1,662 km. This quantity also defines the radius of the network. The maximum eccentricity is 3,097 km, and that node eccentricity is shared by the two most peripheral nodes in Florida and Quebec, Canada. This quantity also represents the diameter of the network. Finally, reportable parameters of the distribution of node eccentricities, shown in Figure 23, include the average node eccentricity as 2,246 km, median of 2,216 km, variance of 81,432 km^2, and skewness of 0.408.

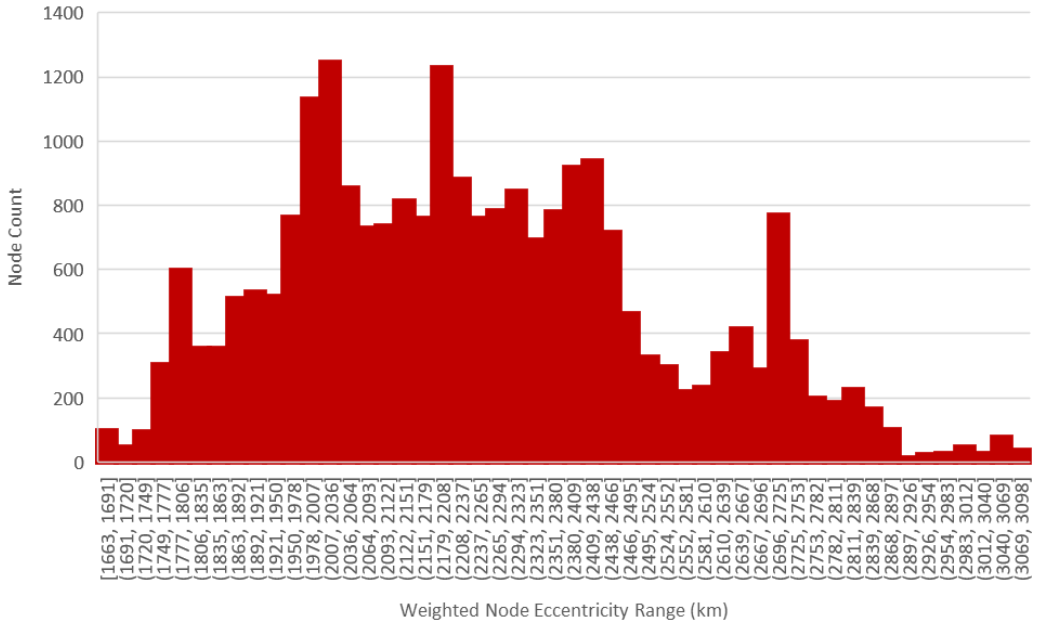


Figure 23. Histogram of Weighted node eccentricity of the CSX Transportation Subnetwork

8.2.2. CSX Transportation Weighted Network Arithmetic Inefficiency

A colormap of node arithmetic inefficiencies of the CSX Transportation subnetwork are shown on Figure 24 Greater values of node arithmetic inefficiency are colored darker blue, while lower values are colored lighter blue.

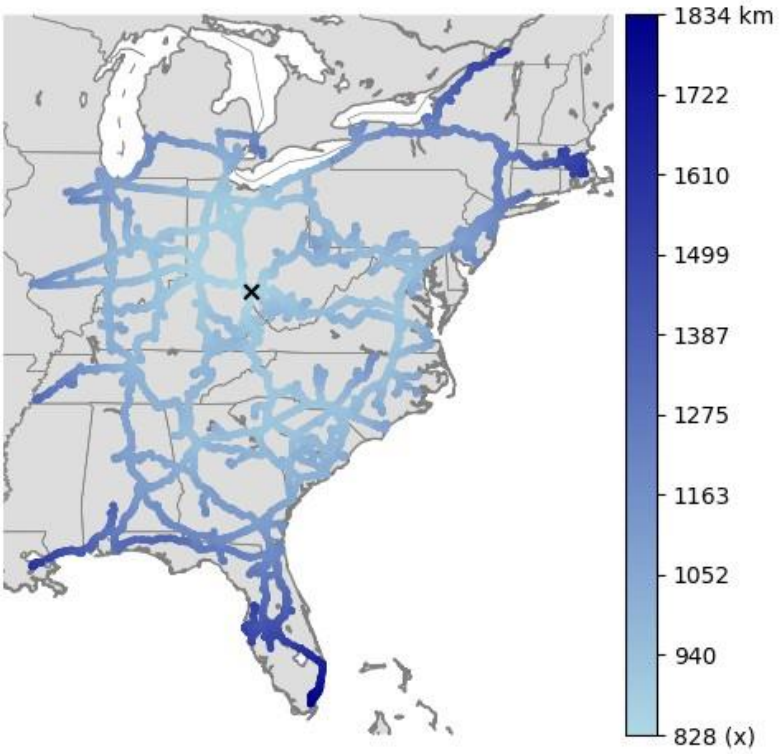


Figure 24. Weighted node arithmetic inefficiency of the CSX Transportation Subnetwork

The node with best arithmetic efficiency (lowest inefficiency) is FRA node ID 444054, marked by an "x" on Figure 24, and is in Boyd County, Kentucky. The arithmetic inefficiency of that node is 828 km. The maximum arithmetic inefficiency is 1,833 km, at the furthermost peripheral node in Florida. Finally, reportable parameters of the distribution of node arithmetic inefficiency, shown in Figure 25, include the average node arithmetic inefficiency as 1,066 km, median of 1,008 km, variance of 29,426 km^2, and skewness of 1.43.

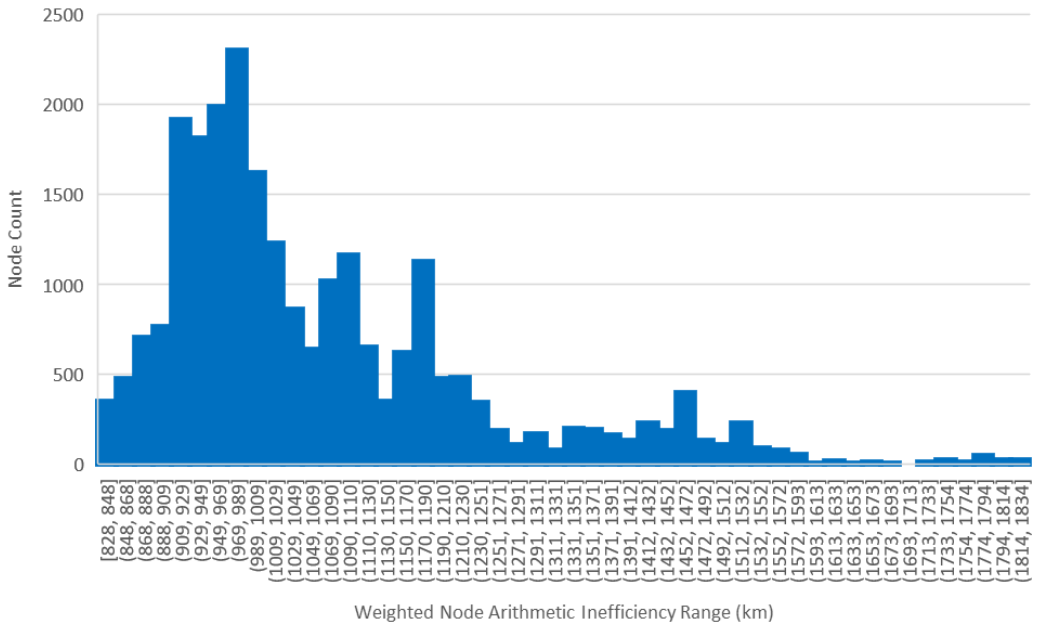


Figure 25. Histogram of Weighted node arithmetic inefficiency of the CSX Transportation Subnetwork

8.3. Canadian Pacific Railway Network

The Canadian Pacific Railway network chosen according to the previously discussed criteria consists of 13,718 links and 12,519 nodes, with 53 component networks. The dominant component was chosen which contains 13,483 links and 12,245 nodes, and the other 52 components are discarded. The total length of all track in this network is 29,600 km. As detailed earlier, actual physical line lengths, also contained in the BTS data, are used as weights for each link for all analysis.

8.3.1. Canadian Pacific Railway Weighted Network Eccentricity

The visualized node eccentricities of the Canadian Pacific Railway subnetwork described previously are shown on Figure 26. Greater values of node eccentricity are colored darker red, while lower values are lighter red.

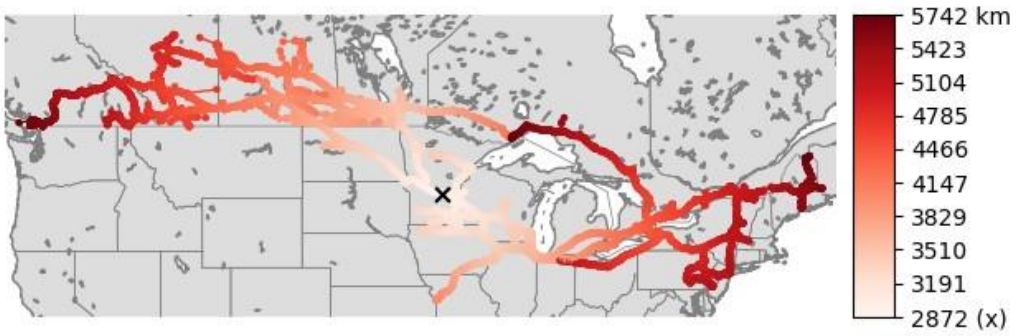


Figure 26. Weighted node eccentricity of the Canadian Pacific Railway Subnetwork

The central node by eccentricity is FRA node ID 377712, marked by an "x" on Figure 26, and is in Wright County, Minnesota. The minimum eccentricity across the network, which is at that central node is 2,872 km. This quantity also defines the radius of the network. The maximum eccentricity is 5,741 km, and that node eccentricity is shared by the two most peripheral nodes in Maine and British Columbia, Canada. This quantity also represents the diameter of the network. Finally, reportable parameters of the distribution of node eccentricities, shown in Figure 27, include the average node eccentricity as 4,296 km, median of 4,418 km, variance of 570,903 km^2, and skewness of -0.0960.

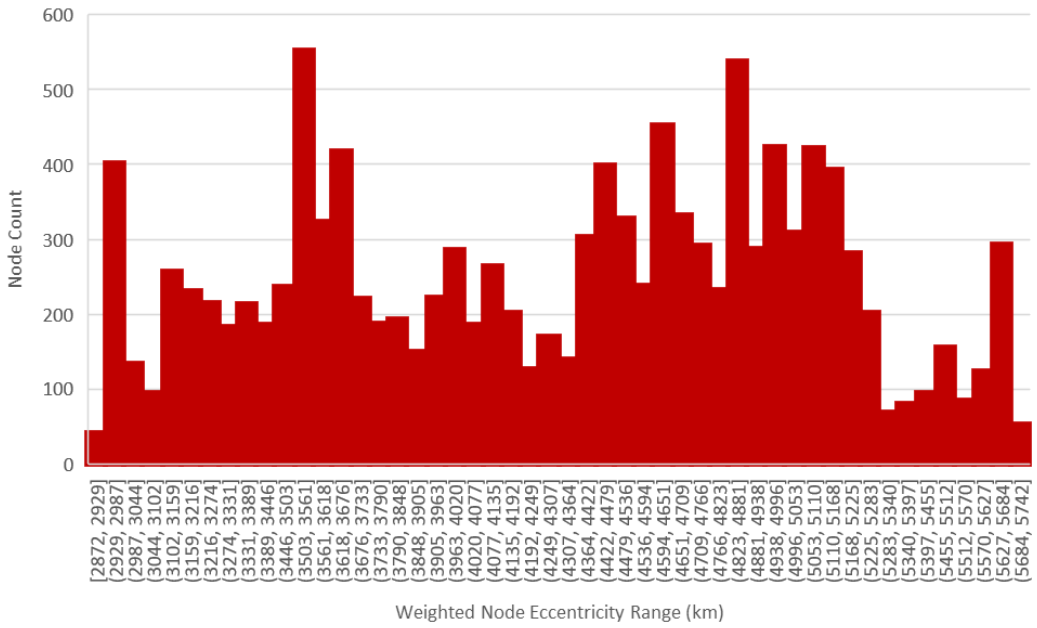


Figure 27. Histogram of Weighted node eccentricity of the Canadian Pacific Railway Subnetwork

8.3.2. Canadian Pacific Railway Weighted Network Arithmetic Inefficiency

A colormap of node arithmetic inefficiencies of the Canadian Pacific Railway subnetwork are shown on Figure 28 Greater values of node arithmetic inefficiency are colored darker blue, while lower values are colored lighter blue.

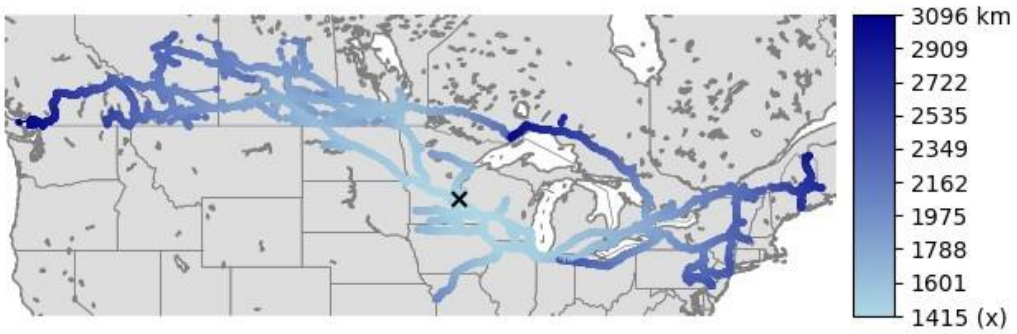


Figure 28. Weighted node arithmetic inefficiency of the Canadian Pacific Railway Subnetwork

The node with best arithmetic efficiency (lowest inefficiency) is FRA node ID 383127, marked by an "x" on Figure 28, and is in Ramsey County, Minnesota. The arithmetic inefficiency of that node is 1,414 km. The maximum arithmetic inefficiency is 3,095 km, at the furthermost peripheral node in Florida. Finally, reportable parameters of the distribution of node arithmetic inefficiency, shown in Figure 29, include the average node arithmetic inefficiency as 1,959 km, median of 1,880 km, variance of 168,168 km^2, and skewness of 0.615.

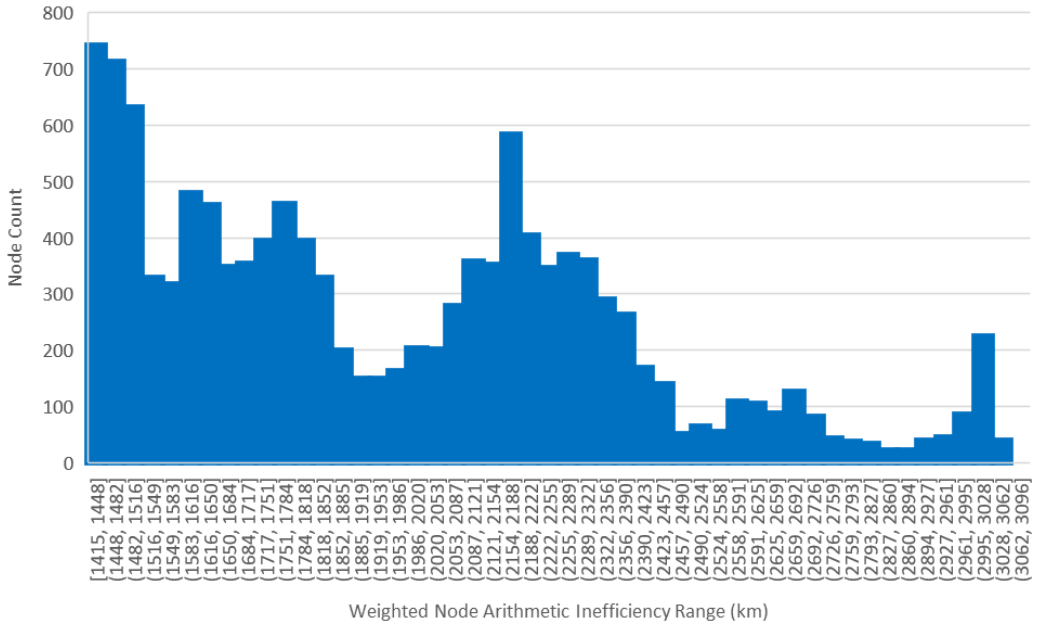


Figure 29. Histogram of Weighted node arithmetic inefficiency of the Canadian Pacific Railway Subnetwork

8.4. Kansas City Southern Railway Company Network

The Kansas City Southern Railway Company network chosen according to the previously discussed criteria consists of 5,304 links and 5,012 nodes, with 28

component networks. The dominant component was chosen which contains 4,951 links and 4,647 nodes, and the other 27 components are discarded. The total length of all track in this network is 11,620 km. As detailed earlier, actual physical line lengths, also contained in the BTS data, are used as weights for each link for all analysis.

8.4.1. Kansas City Southern Railway Company Weighted Network Eccentricity

The visualized node eccentricities of the Kansas City Southern Railway Company subnetwork described previously are shown on Figure 30. Greater values of node eccentricity are colored darker red, while lower values are lighter red.

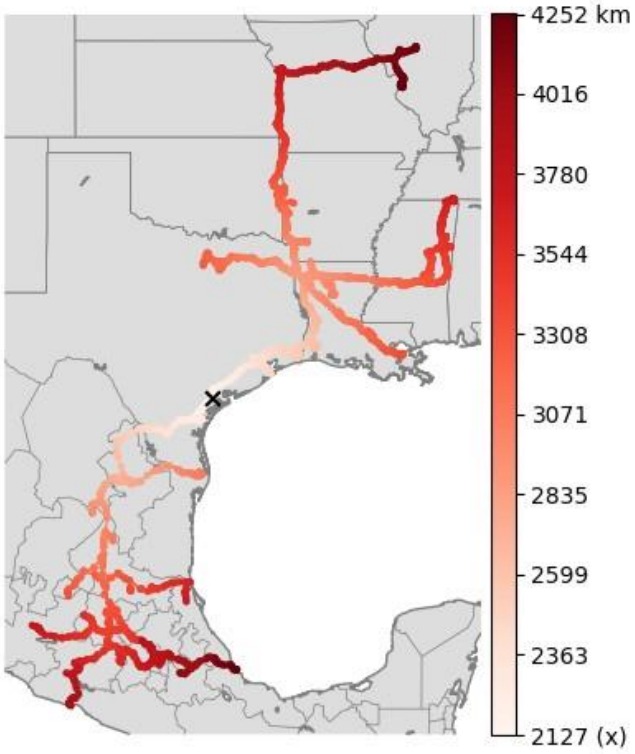


Figure 30. Weighted node eccentricity of the Kansas City Southern Railway Company Subnetwork

The central node by eccentricity is FRA node ID 360021, marked by an "x" on Figure 30, and is in Refugio County, Texas. The minimum eccentricity across the network, which is at that central node is 2,126 km. This quantity also defines the radius of the network. The maximum eccentricity is 4,252 km, and that node eccentricity is shared by the two most peripheral nodes in Illinois and Veracruz, Mexico. This quantity also represents the diameter of the network. Finally, reportable parameters of the distribution of node eccentricities, shown in Figure 31, include the average node eccentricity as 3,235 km, median of 3,246 km, variance of 245,311 km^2, and skewness of -0.121.

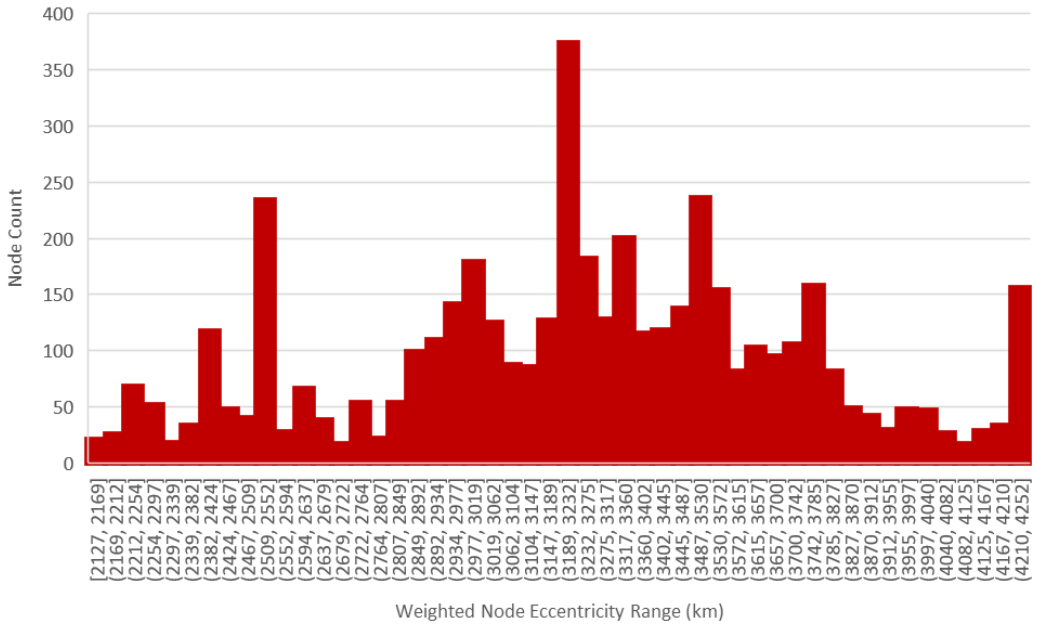


Figure 31. Histogram of Weighted node eccentricity of the Kansas City Southern Railway Company Subnetwork

8.4.2. Kansas City Southern Railway Company Weighted Network Arithmetic Inefficiency

A colormap of node arithmetic inefficiencies of the Kansas City Southern Railway Company subnetwork are shown on Figure 32 Greater values of node arithmetic inefficiency are colored darker blue, while lower values are colored lighter blue.

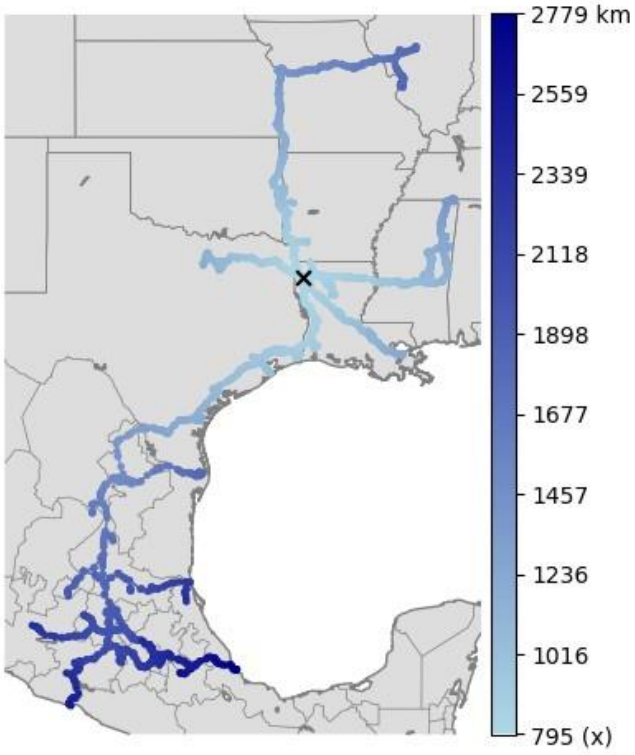


Figure 32. Weighted node arithmetic inefficiency of the Kansas City Southern Railway Company Subnetwork

The node with best arithmetic efficiency (lowest inefficiency) is FRA node ID 379303, marked by an "x" on Figure 32, and is in Caddo Parish, Louisiana. The arithmetic inefficiency of that node is 795 km. The maximum arithmetic inefficiency is 2,779 km, at the furthermost peripheral node in Veracruz Llave, Mexico. Finally, reportable parameters of the distribution of node arithmetic inefficiency, shown in

Figure 33, include the average node arithmetic inefficiency as 1,294 km, median of 1,140 km, variance of 200,522 km^2, and skewness of 1.225.

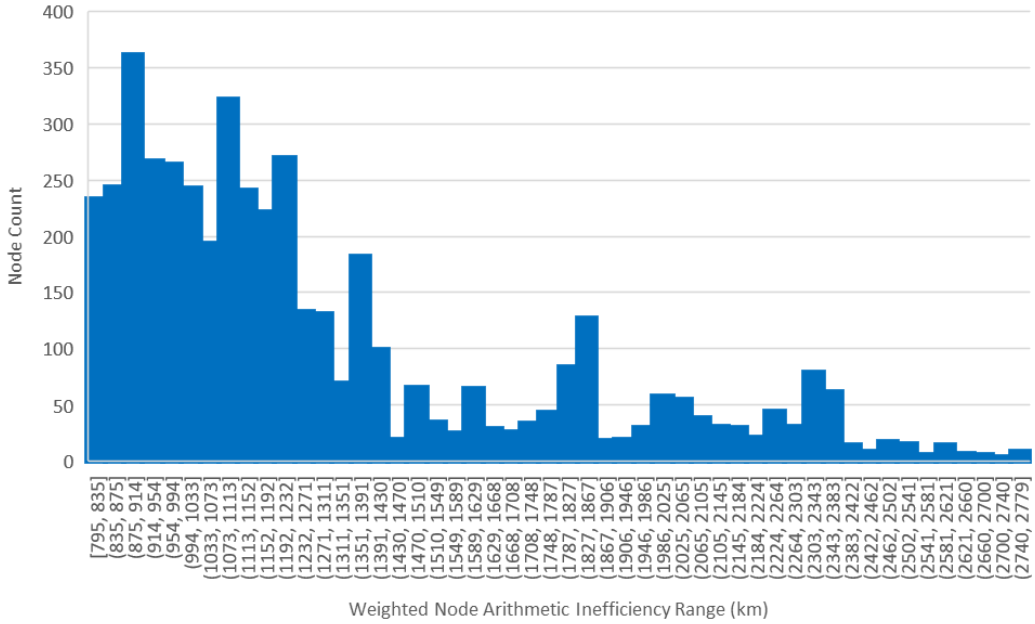


Figure 33. Histogram of Weighted node arithmetic inefficiency of the Kansas City Southern Railway Company Subnetwork

8.5. Norfolk Southern Railway Network

The Norfolk Southern Railway network chosen according to the previously discussed criteria consists of 26,339 links and 25,012 nodes, with 176 component networks. The dominant component was chosen which contains 25,634 links and 24,149 nodes, and the other 175 components are discarded. The total length of all track in this network is 37,656 km. As detailed earlier, actual physical line lengths, also contained in the BTS data, are used as weights for each link for all analysis.

8.5.1. Norfolk Southern Railway Weighted Network Eccentricity

The visualized node eccentricities of the Norfolk Southern Railway subnetwork described previously are shown on Figure 34. Greater values of node eccentricity are colored darker red, while lower values are lighter red.

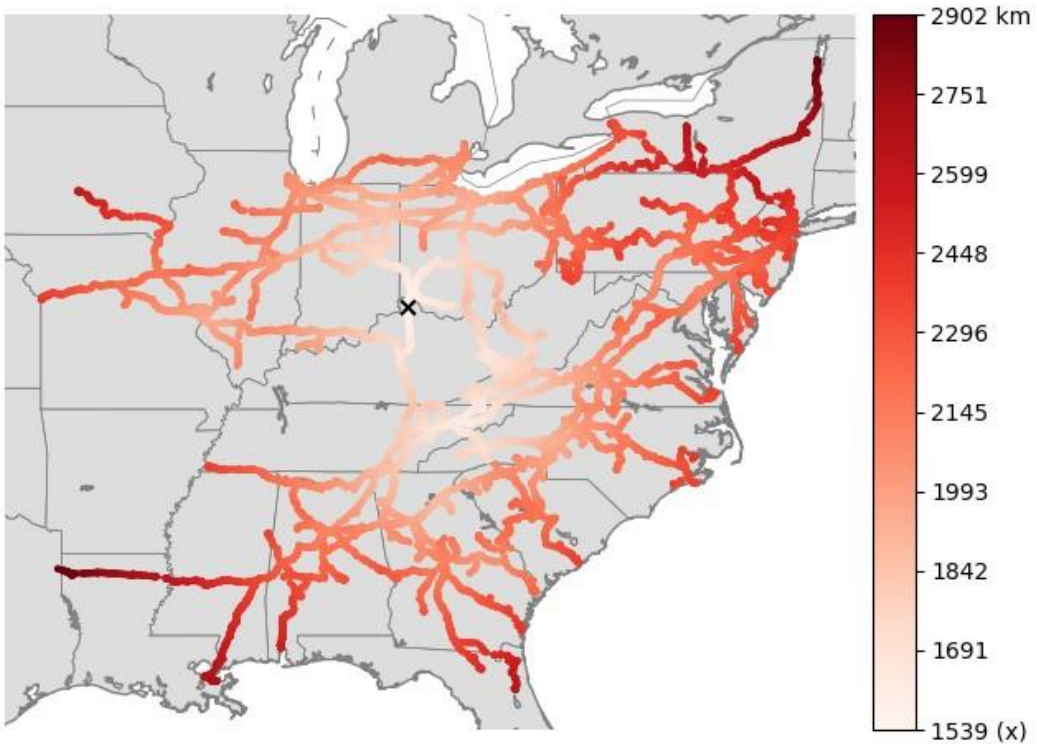


Figure 34. Weighted node eccentricity of the Norfolk Southern Railway Subnetwork

The central node by eccentricity is FRA node ID 431086, marked by an "x" on Figure 34, and is in Boone County, Kentucky. The minimum eccentricity across the network, which is at that central node is 1,539 km. This quantity also defines the radius of the network. The maximum eccentricity is 2,901 km, and that node eccentricity is shared by the two most peripheral nodes in Texas and New York. This quantity also represents the diameter of the network. Finally, reportable parameters of the

distribution of node eccentricities, shown in Figure 35, include the average node eccentricity as 2,104 km, median of 2,109 km, variance of 52,495 km^2, and skewness of 0.0318.

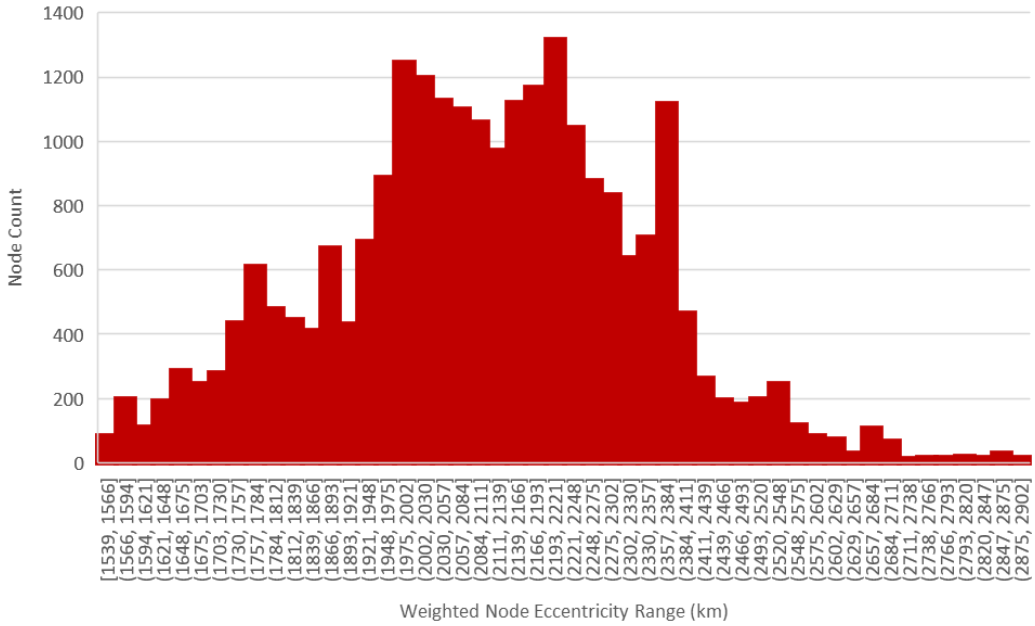


Figure 35. Histogram of Weighted node eccentricity of the Norfolk Southern Railway Subnetwork

8.5.2. Norfolk Southern Railway Weighted Network Arithmetic Inefficiency

A colormap of node arithmetic inefficiencies of the Norfolk Southern Railway subnetwork are shown on Figure 36 Greater values of node arithmetic inefficiency are colored darker blue, while lower values are colored lighter blue.

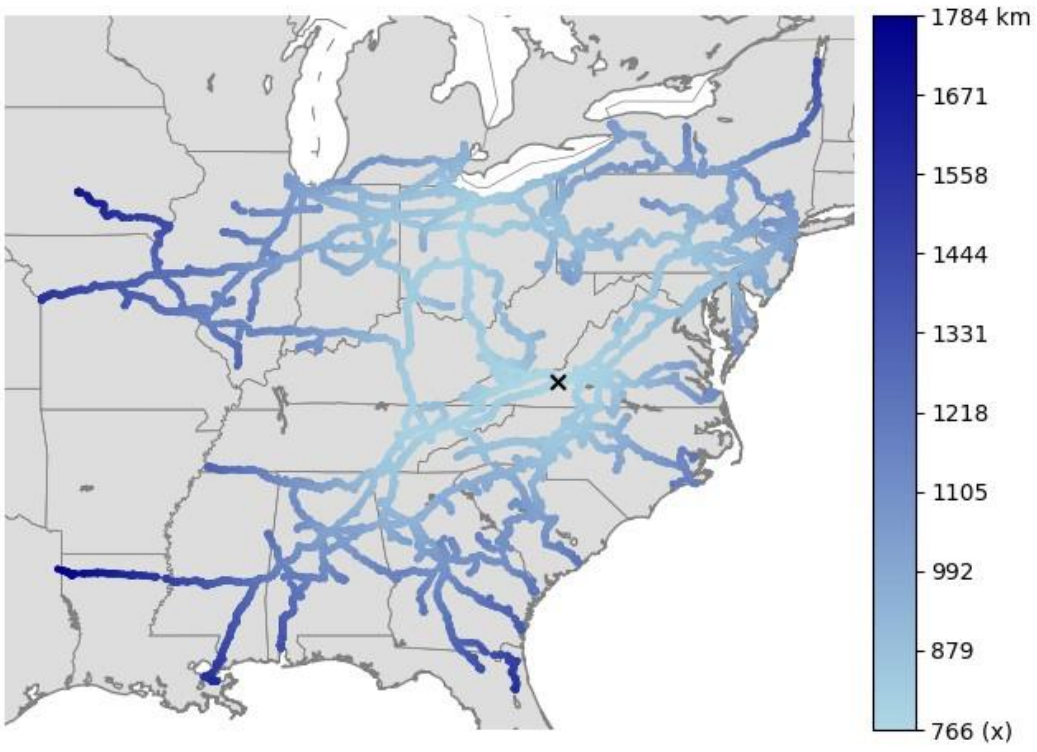


Figure 36. Weighted node arithmetic inefficiency of the Norfolk Southern Railway Subnetwork

The node with best arithmetic efficiency (lowest inefficiency) is FRA node ID 456911, marked by an "x" on Figure 36, and is in Montgomery County, Virginia. The arithmetic inefficiency of that node is 765 km. The maximum arithmetic inefficiency is 1,783 km, at the furthermost peripheral node in Texas. Finally, reportable parameters of the distribution of node arithmetic inefficiency, shown in Figure 37, include the average node arithmetic inefficiency as 979 km, median of 934 km, variance of 27,176 km^2, and skewness of 1.486.

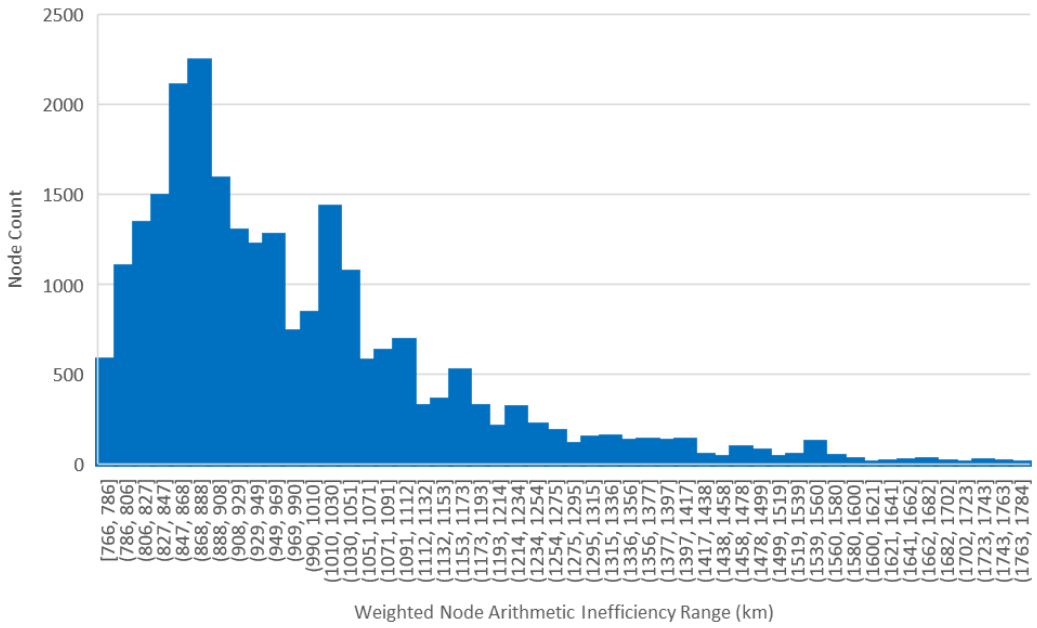


Figure 37. Histogram of Weighted node arithmetic inefficiency of the Norfolk Southern Railway Subnetwork

8.6. Canadian National Railway Company Network

The Canadian National Railway Company network chosen according to the previously discussed criteria consists of 20,380 links and 18,830 nodes, with 91 component networks. The dominant component was chosen which contains 19,580 links and 18,005 nodes, and the other 90 components are discarded. The total length of all track in this network is 42,080 km. As detailed earlier, actual physical line lengths, also contained in the BTS data, are used as weights for each link for all analysis.

8.6.1. Canadian National Railway Company Weighted Network Eccentricity

The visualized node eccentricities of the Canadian National Railway Company subnetwork described previously are shown on Figure 38. Greater values of node eccentricity are colored darker red, while lower values are lighter red.

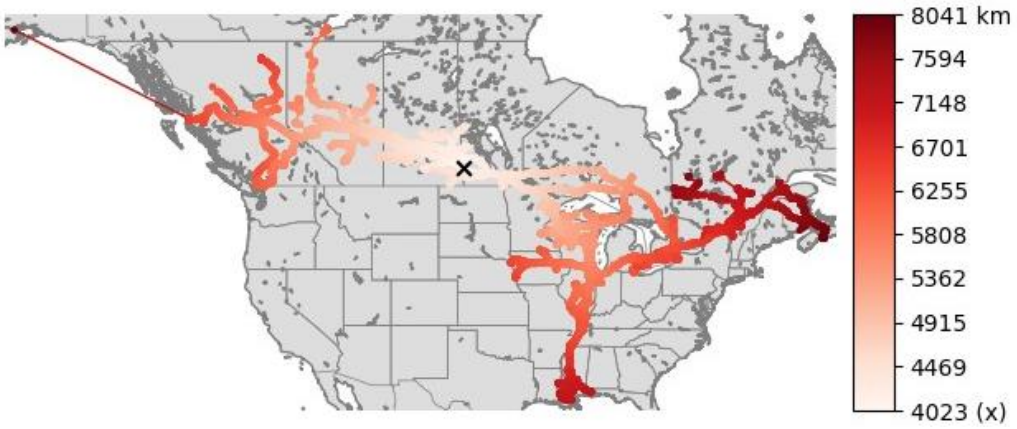


Figure 38. Weighted node eccentricity of the Canadian National Railway Company Subnetwork

The central node by eccentricity is FRA node ID 620176, marked by an "x" on Figure 38, and is in Saskatchewan, Canada. The minimum eccentricity across the network, which is at that central node is 4,022 km. This quantity also defines the radius of the network. The maximum eccentricity is 8,040 km, and that node eccentricity is shared by the two most peripheral nodes in Alaska and Nova Scotia, Canada. This quantity also represents the diameter of the network. Finally, reportable parameters of the distribution of node eccentricities, shown in Figure 39, include the average node eccentricity as 5,970 km, median of 6,053 km, variance of 759,337 km^2, and skewness of -0.149.

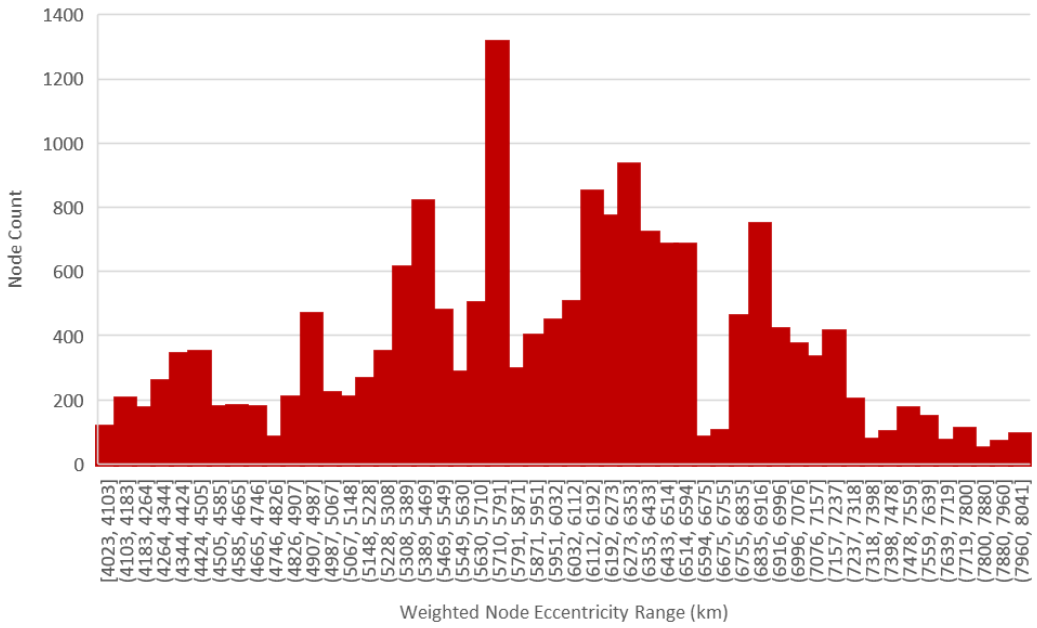


Figure 39. Histogram of Weighted node eccentricity of the Canadian National Railway Company Subnetwork

8.6.2. Canadian National Railway Company Weighted Network Arithmetic Inefficiency

A colormap of node arithmetic inefficiencies of the Canadian National Railway Company subnetwork are shown on Figure 40 Greater values of node arithmetic inefficiency are colored darker blue, while lower values are colored lighter blue.

Figure 40. Weighted node arithmetic inefficiency of the Canadian National Railway Company Subnetwork

The node with best arithmetic efficiency (lowest inefficiency) is FRA node ID 414606, marked by an "x" on Figure 40, and is in Cook County, Illinois. The arithmetic inefficiency of that node is 1,362 km. The maximum arithmetic inefficiency is 5,496 km, at the furthermost peripheral node in Alaska. Finally, reportable parameters of the distribution of node arithmetic inefficiency, shown in Figure 41, include the average node arithmetic inefficiency as 2,092 km, median of 1,890 km, variance of 394,543 km^2, and skewness of 1.034.

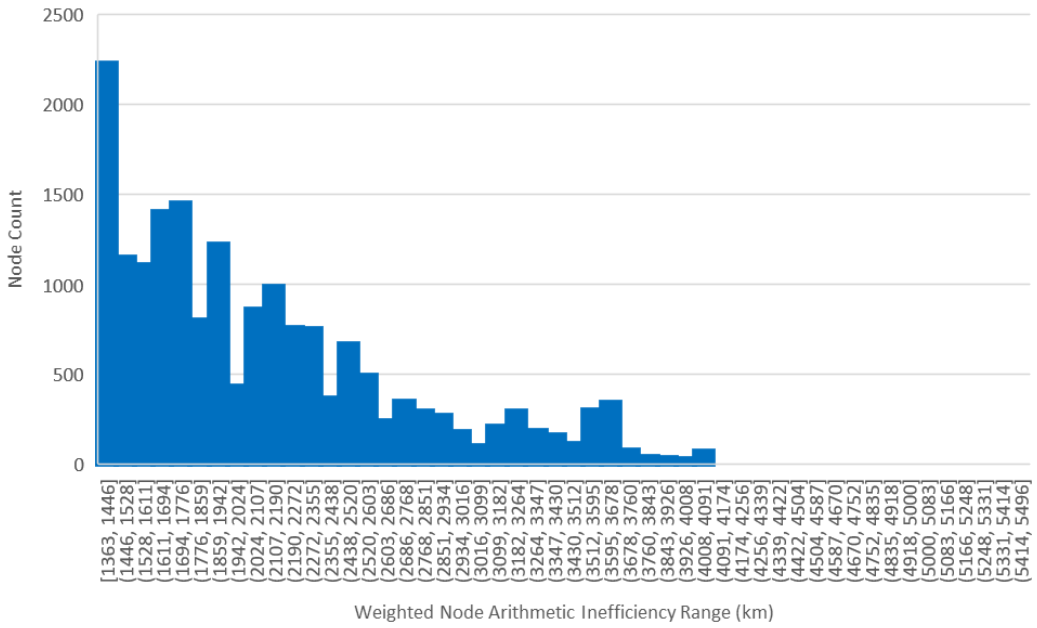


Figure 41. Histogram of Weighted node arithmetic inefficiency of the Canadian National Railway Company Subnetwork

8.7. Union Pacific Railroad Network

The Union Pacific Railroad network chosen according to the previously discussed criteria consists of 37,595 links and 34,706 nodes, with 217 component networks. The dominant component was chosen which contains 36,766 links and 33,693 nodes, and the other 216 components are discarded. The total length of all track in this network is 60,834 km. As detailed earlier, actual physical line lengths, also contained in the BTS data, are used as weights for each link for all analysis.

8.7.1. Union Pacific Railroad Weighted Network Eccentricity

The visualized node eccentricities of the Union Pacific Railroad subnetwork described previously are shown on Figure 42. Greater values of node eccentricity are colored darker red, while lower values are lighter red.

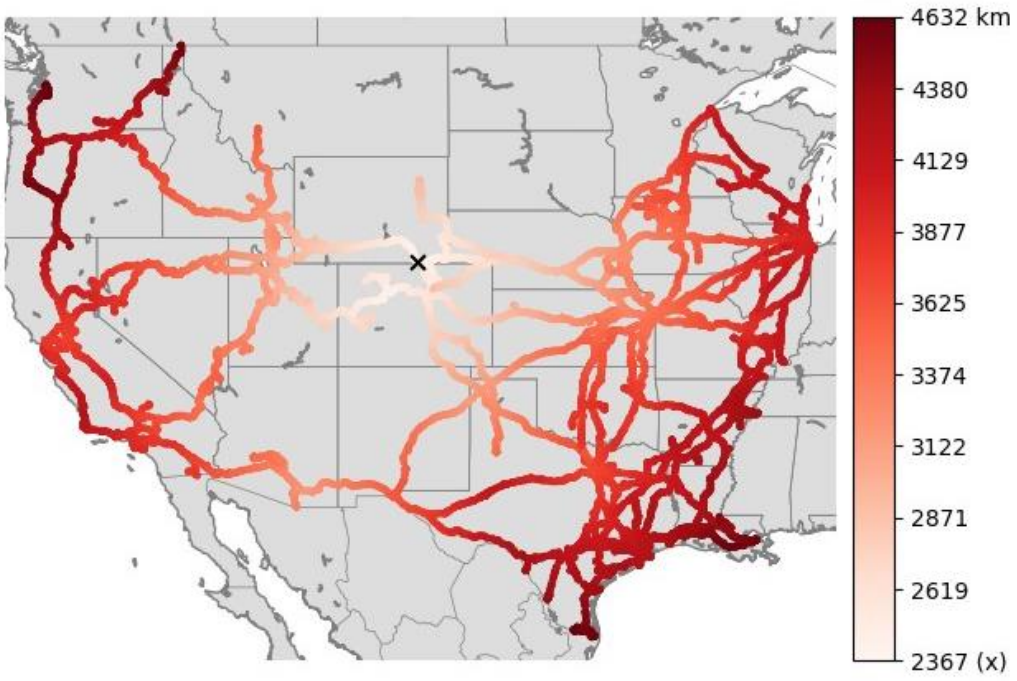


Figure 42. Weighted node eccentricity of the Union Pacific Railroad Subnetwork

The central node by eccentricity is FRA node ID 339954, marked by an "x" on Figure 42, and is in Albany County, Wyoming. The minimum eccentricity across the network, which is at that central node is 2,367 km. This quantity also defines the radius of the network. The maximum eccentricity is 4,631 km, and that node eccentricity is shared by the two most peripheral nodes in Washington State and Texas. This quantity also represents the diameter of the network. Finally, reportable parameters of the distribution of node eccentricities, shown in Figure 43, include the average node eccentricity as 3,760 km, median of 3,847 km, variance of 229,878 km^2, and skewness of -0.628.

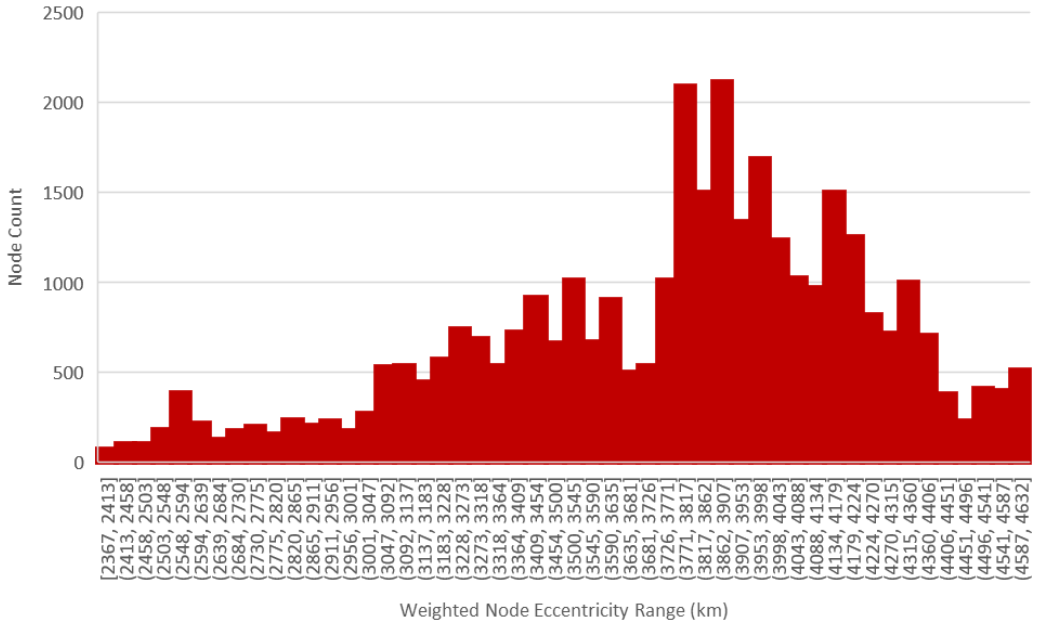


Figure 43. Histogram of Weighted node eccentricity of the Union Pacific Railroad Subnetwork

8.7.2. Union Pacific Railroad Weighted Network Arithmetic Inefficiency

A colormap of node arithmetic inefficiencies of the Union Pacific Railroad subnetwork are shown on Figure 44 Greater values of node arithmetic inefficiency are colored darker blue, while lower values are colored lighter blue.

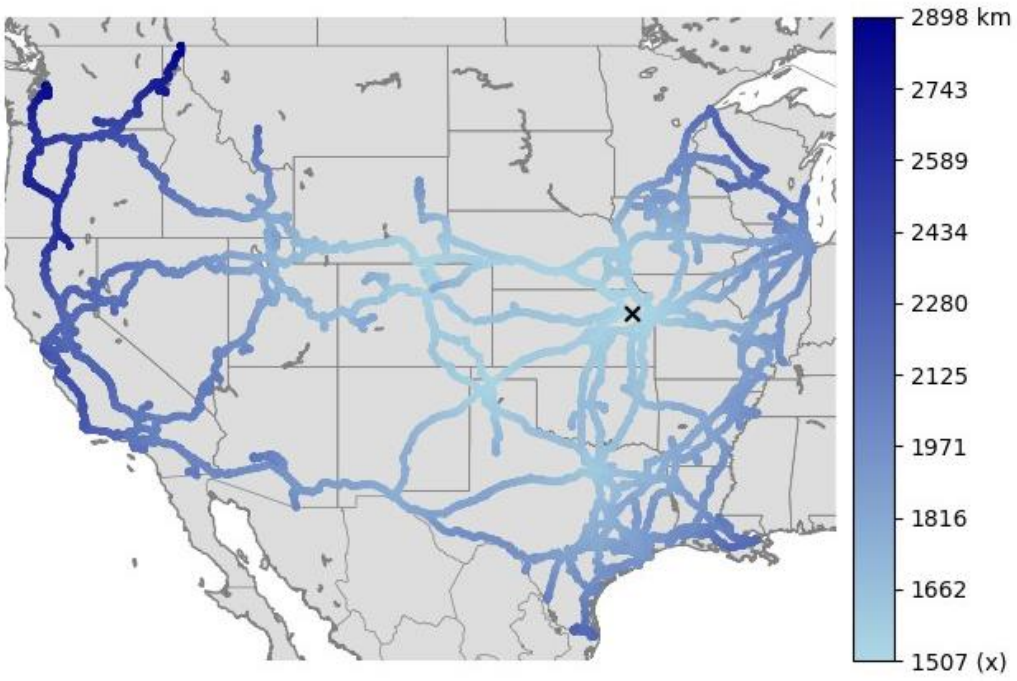


Figure 44. Weighted node arithmetic inefficiency of the Union Pacific Railroad Subnetwork

The node with best arithmetic efficiency (lowest inefficiency) is FRA node ID 367772, marked by an "x" on Figure 44, and is in Shawnee County, Kansas. The arithmetic inefficiency of that node is 1,507 km. The maximum arithmetic inefficiency is 2,897 km, at the furthermost peripheral node in Idaho. Finally, reportable parameters of the distribution of node arithmetic inefficiency, shown in Figure 45, include the average node arithmetic inefficiency as 1,945 km, median of 1,896 km, variance of 87,741 km^2, and skewness of 0.845.

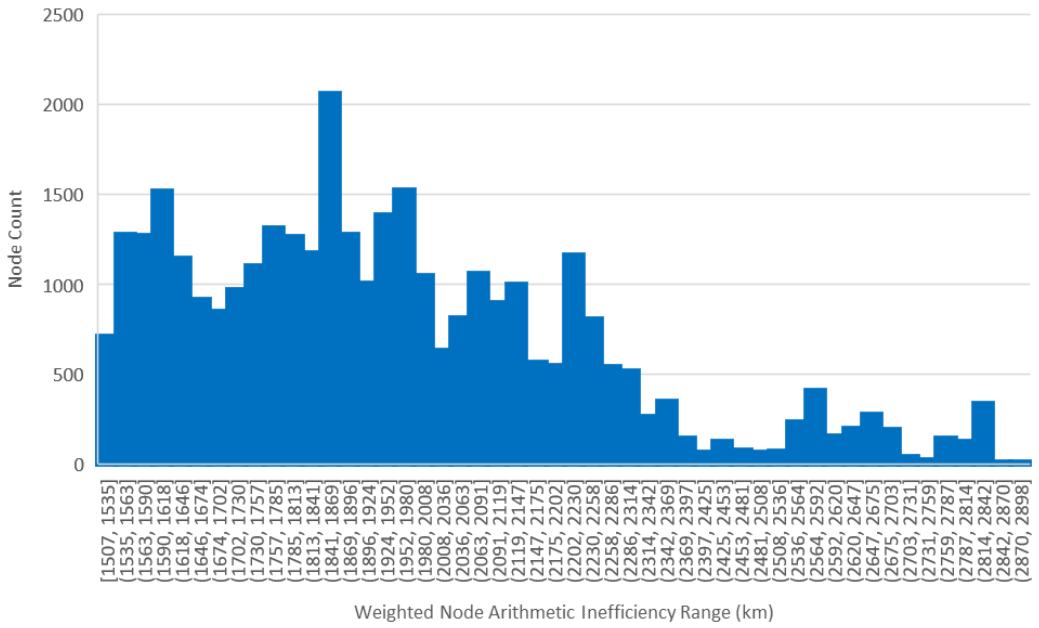


Figure 45. Histogram of Weighted node arithmetic inefficiency of the Union Pacific Railroad Subnetwork

8.8. Amtrak Network

The Amtrak network chosen according to the previously discussed criteria consists of 14,427 links and 14,071 nodes, with 27 component networks. The dominant component was chosen which contains 14,192 links and 13,821 nodes, and the other 26 components are discarded. Amtrak does not own or co-own a majority of their network, unlike Class I carriers. Their network is mostly made of lines they operate with track rights. The total length of all track in this network is 35,730 km. As detailed earlier, actual physical line lengths, also contained in the BTS data, are used as weights for each link for all analysis.

8.8.1. Amtrak Weighted Network Eccentricity

The visualized node eccentricities of the Amtrak subnetwork described previously are shown on Figure 46. Greater values of node eccentricity are colored darker red, while lower values are lighter red.

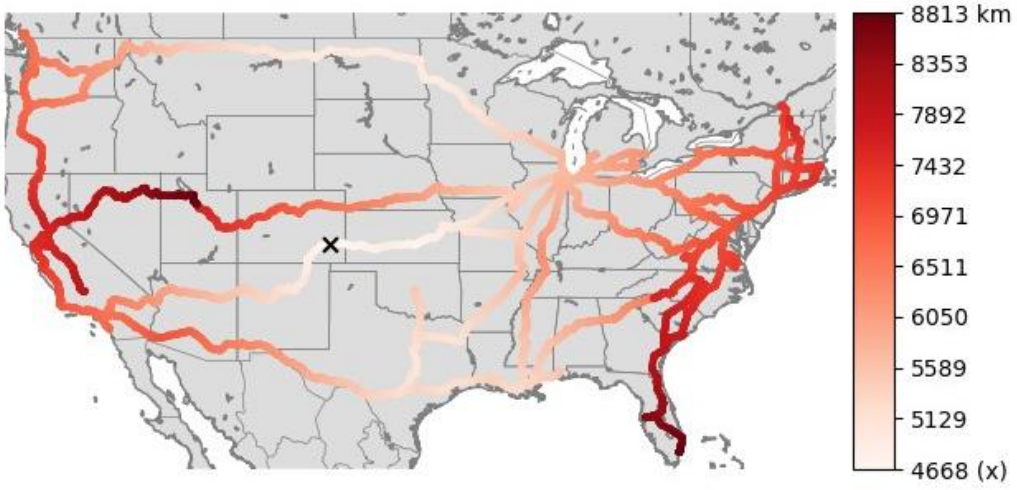


Figure 46. Weighted node eccentricity of the Amtrak Subnetwork

The central node by eccentricity is FRA node ID 344429, marked by an "x" on Figure 46, and is in Bent County, Colorado. The minimum eccentricity across the network, which is at that central node is 4,668 km. This quantity also defines the radius of the network. The maximum eccentricity is 8,813 km, and that node eccentricity is shared by the two most peripheral nodes in Florida and Utah. That Utah is the most peripheral node indicates a discontinuity in the lines from California across the central Midwest, potentially a data error or a true gap in their network. This quantity also represents the diameter of the network. Finally, reportable parameters of the distribution of node eccentricities, shown in Figure 47, include the average node

eccentricity as 6,420 km, median of 6,441 km, variance of 895,429 km^2, and skewness of 0.205.

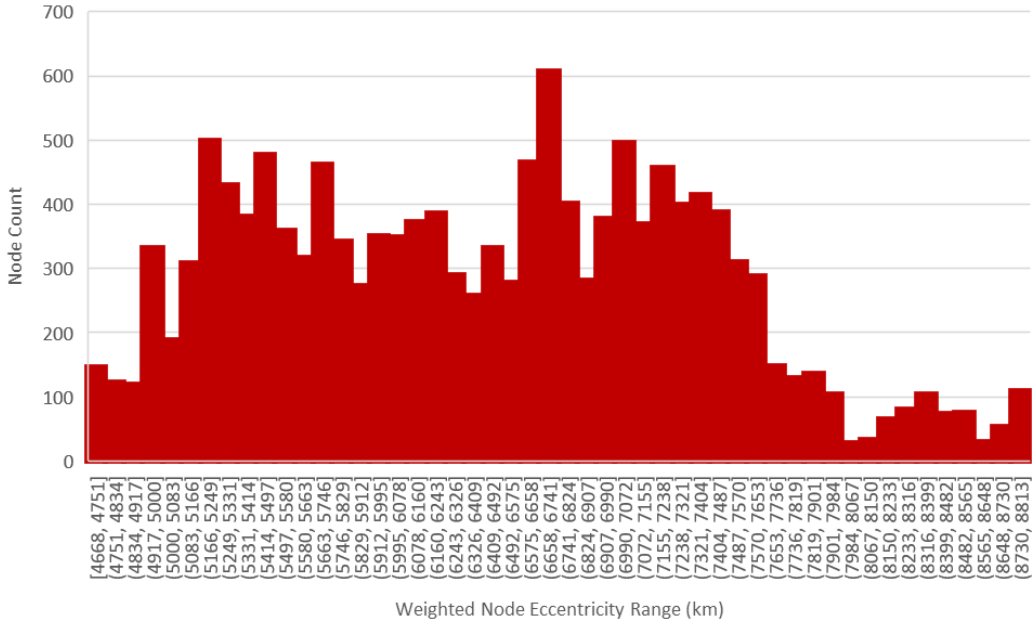


Figure 47. Histogram of Weighted node eccentricity of the Amtrak Subnetwork

8.8.2. Amtrak Weighted Network Arithmetic Inefficiency

A colormap of node arithmetic inefficiencies of the Amtrak subnetwork are shown on Figure 48 Greater values of node arithmetic inefficiency are colored darker blue, while lower values are colored lighter blue.

Figure 48. Weighted node arithmetic inefficiency of the Amtrak Subnetwork

The node with best arithmetic efficiency (lowest inefficiency) is FRA node ID 414657, marked by an "x" on Figure 48, and is in Cook County, Illinois. The arithmetic inefficiency of that node is 1,932 km. The maximum arithmetic inefficiency is 5,083 km, at the furthermost peripheral node in Utah. Finally, reportable parameters of the distribution of node arithmetic inefficiency, shown in Figure 49, include the average node arithmetic inefficiency as 2,953 km, median of 2,823 km, variance of 418,193 km^2, and skewness of 0.618.

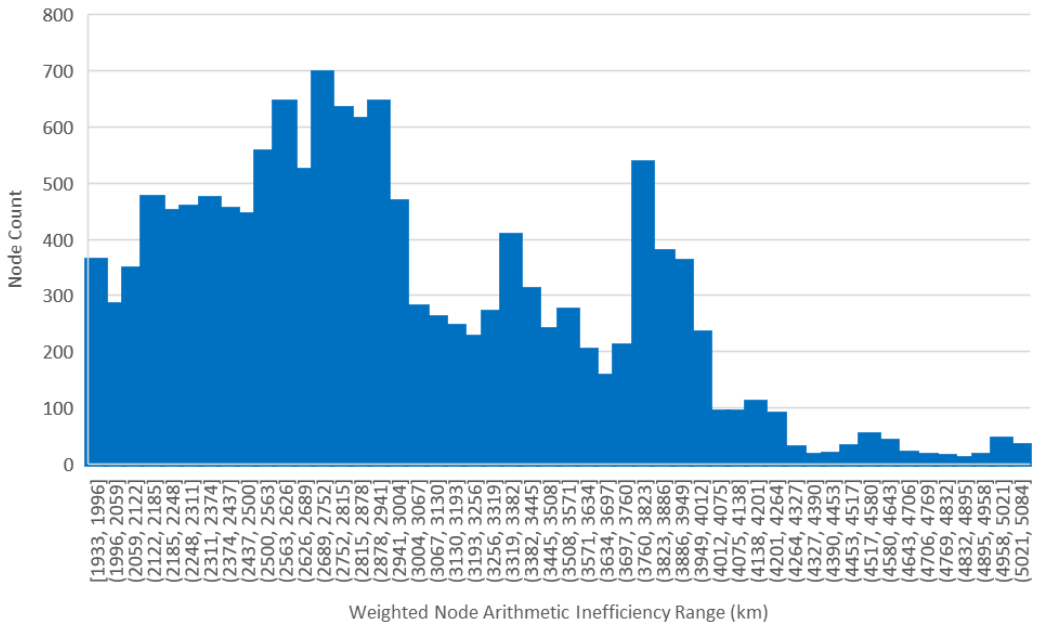


Figure 49. Histogram of Weighted node arithmetic inefficiency of the Amtrak Subnetwork

8.9. Canadian Pacific Kansas City Limited Network

While this thesis was being prepared two of the smallest Class I carriers by track length, Canadian Pacific Railway and The Kansas City Southern Railway, initiated and finalized a merger to become Canadian Pacific Kansas City Limited. Their new network will be analyzed and presented together with their prior separate networks and compared with other Class I carriers. This comparison will showcase a potential application of the network analysis presented herein, that is evaluating network changes.

The Canadian Pacific Kansas City Limited network chosen according to the previously discussed criteria consists of 19,001 links and 17,509 nodes, with 80

component networks. The dominant component was chosen which contains 18,413 links and 16,870 nodes, and the other 79 components are discarded. Amtrak does not own or co-own a majority of their network, unlike Class I carriers. Their network is mostly made of lines they operate with track rights. The total length of all track in this network is 41,207 km. As detailed earlier, actual physical line lengths, also contained in the BTS data, are used as weights for each link for all analysis.

8.9.1. Canadian Pacific Kansas City Limited Weighted Network Eccentricity

The visualized node eccentricities of the Canadian Pacific Kansas City Limited subnetwork described previously are shown on Figure 50. Greater values of node eccentricity are colored darker red, while lower values are lighter red.

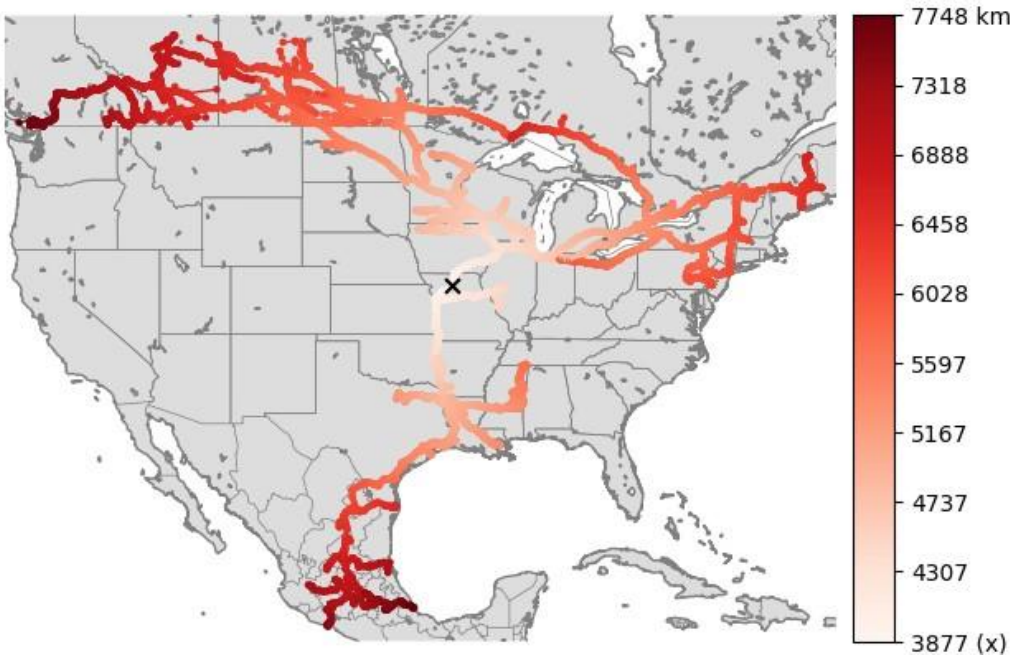


Figure 50.Weighted node eccentricity of the Canadian Pacific Kansas City Limited Subnetwork

The central node by eccentricity is FRA node ID 380436, marked by an "x" on Figure 46, and is in Livingston County, Missouri. The minimum eccentricity across the network, which is at that central node is 3,877 km. This quantity also defines the radius of the network. The maximum eccentricity is 7,748 km, and that node eccentricity is shared by the two most peripheral nodes in Veracruz, Mexico and British Columbia, Canada. This quantity also represents the diameter of the network. Finally, reportable parameters of the distribution of node eccentricities, shown in Figure 51, include the average node eccentricity as 5,660 km, median of 5,598 km, variance of 764,738 km^2, and skewness of 0.266.

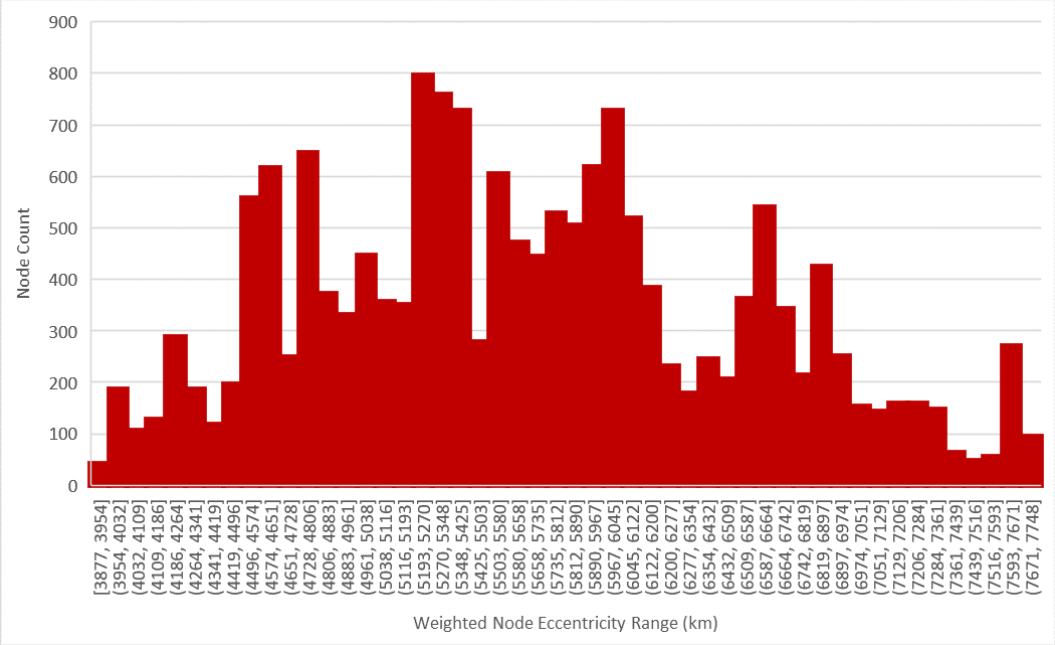


Figure 51. Histogram of Weighted node eccentricity of the Canadian Pacific Kansas City Limited Subnetwork

8.9.2. Canadian Pacific Kansas City Limited Weighted Network Arithmetic Inefficiency

A colormap of node arithmetic inefficiencies of the Canadian Pacific Kansas City Limited subnetwork are shown on Figure 52. Greater values of node arithmetic inefficiency are colored darker blue, while lower values are colored lighter blue.

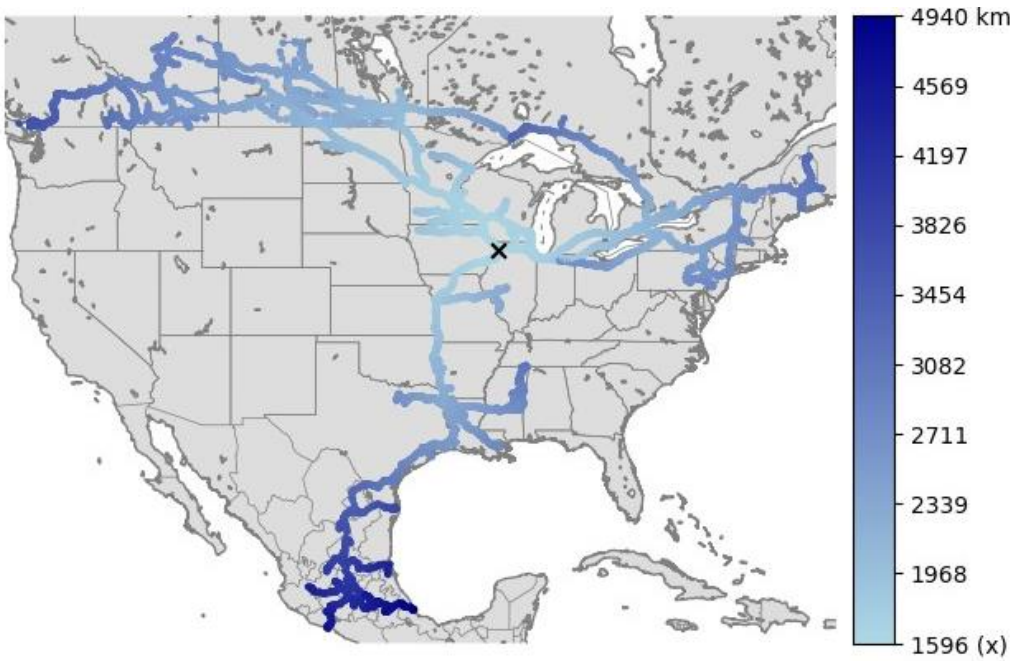


Figure 52. Weighted node arithmetic inefficiency of the Canadian Pacific Kansas City Limited Subnetwork

The node with best arithmetic efficiency (lowest inefficiency) is FRA node ID 395686, marked by an "x" on Figure 52, and is in Jackson County, Iowa. The arithmetic inefficiency of that node is 1,596 km. The maximum arithmetic inefficiency is 4,940 km, at the furthermost peripheral node in Veracruz, Mexico. Finally, reportable parameters of the distribution of node arithmetic inefficiency, shown in Figure 53, include the average node arithmetic inefficiency as 2,484 km, median of 2,492 km, variance of 375,586 km^2, and skewness of 1.023.

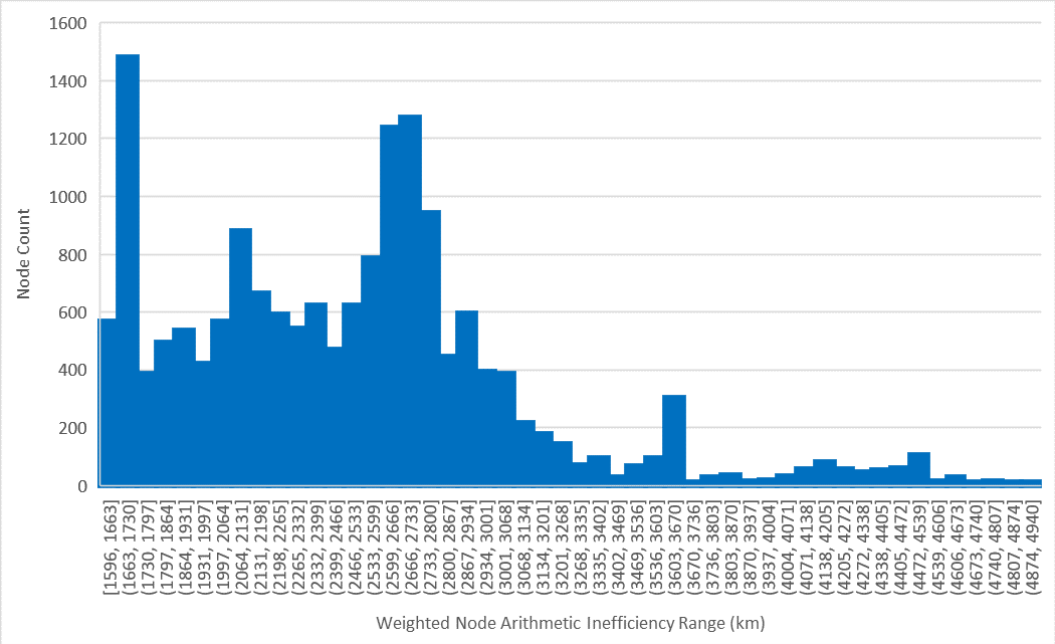


Figure 53. Histogram of Weighted node arithmetic inefficiency of the Canadian Pacific Kansas City Limited Subnetwork

Chapter 9: Rankings of Class I Carriers by Network Characteristics

One of the practical uses of network analysis is that entire networks and their structures can be ranked and compared. The Class 1 carrier networks' maximum eccentricity and average efficiency characteristics have been computed for each of their respective largest subnetworks for ranking and comparison purposes. The comparisons are limited and are not meant to render an absolute judgement on which network is best. For example, a larger network in terms of node/link count, area coverage, or diameter would potentially be much more costly and perhaps impractical to increase link density and efficiency than for a smaller network. Geographical concerns may dictate or limit the routes for some Class 1 carriers and the different operational needs of different subnetworks may further limit their achievable efficiency. Still, the rankings may be helpful to determine which network might be more vulnerable to disruption and to compare their geometric sizes.

Finally, some assumptions must be made about the underlying data for the subnetworks to be comparable. The nodes and links are assumed to be relatively equally distributed along the network and that they represent equivalent network elements on each subnetwork. This may not be a perfect assumption in all cases but should be for most. For example, most links are below 1 km in length, and often joined by several nodes in a long string, but there is one link in the Canadian National Railway Company subnetwork that is reported as 1554 km long. The average link size however is 1.4 km with a standard deviation of 4.9 km, and a median of 0.38 km.

The distribution of link lengths in km on the unfiltered list of 302,835 links is dominated by relatively short links and is shown on a histogram at Figure 54.

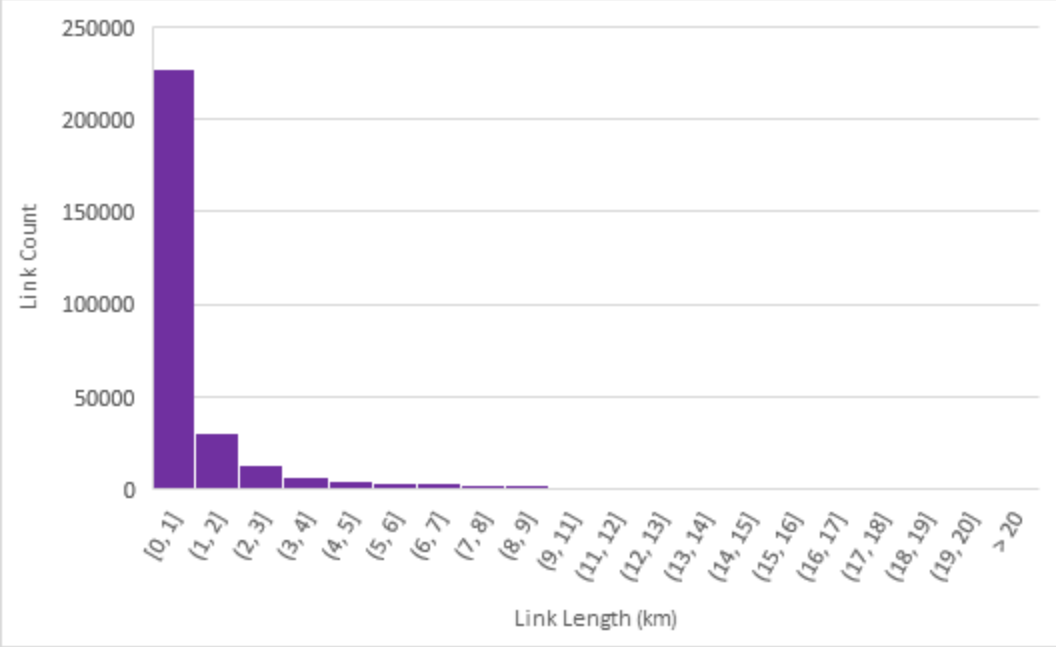


Figure 54. Histogram of link lengths for the complete unfiltered railroad link data

The Arithmetic Efficiency of each network is ranked in Figure 55. The subnetwork with the smallest arithmetic inefficiency is most efficient. As detailed earlier, arithmetic efficiency is a measure of the average shortest path for all nodes to all others. The most arithmetic efficient network can benefit from being slightly smaller, but also from being dense or more heavily grid structured. For example, Canadian Pacific Railway is slightly larger than BNSF Railway by diameter (5,741 km vs 5,216 km), but much smaller in terms of node code (12,245 vs 28,604) yet they have very similar arithmetic inefficiencies (1,959 km vs 1,983 km). BNSF Railway achieved this efficiency through the denser more grid-like structure of their network.

Interestingly, the new Canadian Pacific Kansas City Limited railroad is one of least efficient networks but is only the fourth largest Class I carrier by track length.

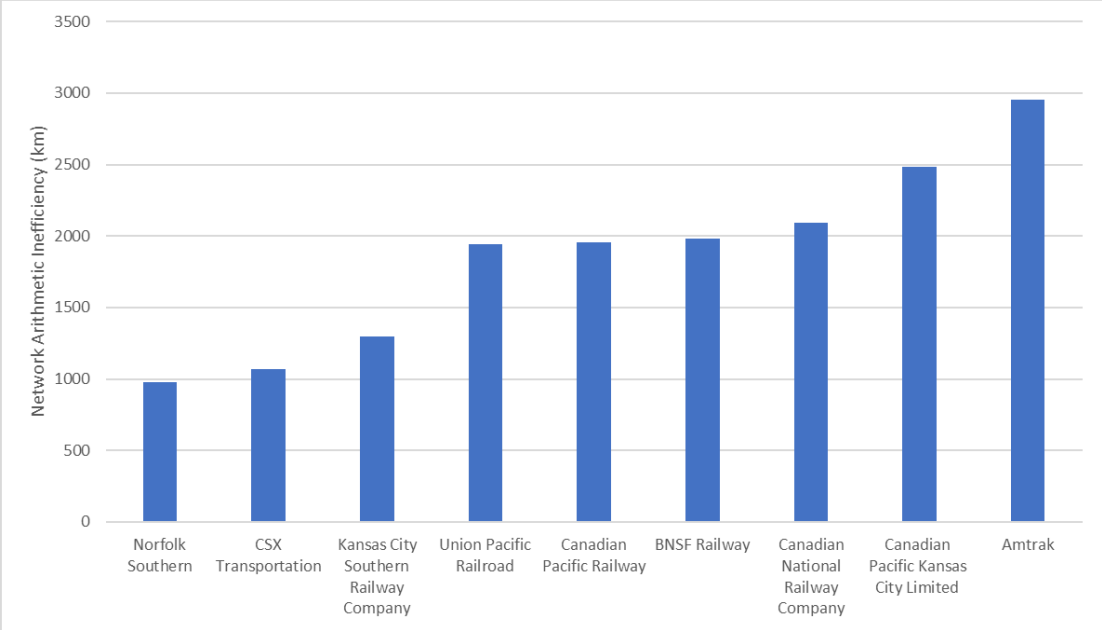


Figure 55. Class 1 Carrier Subnetworks ranked by arithmetic efficiency (largest to smallest average efficiency)

The Diameter of each network is ranked in Figure 56. Networks with a larger diameter cover greater distances at their maximum extents. By inspection, Amtrak does cover the largest area, reaching almost throughout the entire contiguous United States. Canadian National spans a large distance as well, reaching from Alaska to Eastern Canada and the Southern US. The new merged railroad, Canadian Pacific Kansas City Limited, is now close behind Canadian National.

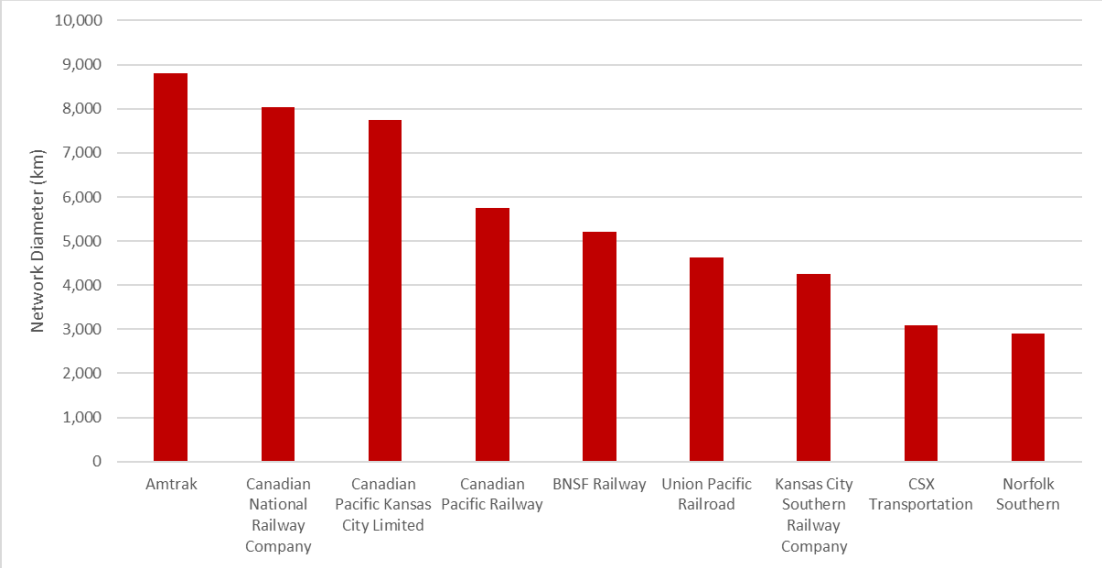


Figure 56. Class 1 Carrier Subnetworks ranked by diameter (largest to smallest length)

A common metric used to evaluate network size, particular for railroad networks, is the total length of all track. The BTS data set derived measure for this metric for all the subnetworks of interest was detailed in each network description and is ranked and shown on Figure 57. This size measure expectedly does not fully correlate to the diameter or inefficiency measure of each network. The newly merged railroad,

Canadian Pacific Kansas City Limited, formed from the two smallest Class I Carriers and is now the fourth largest, sitting roughly in the middle of the field.

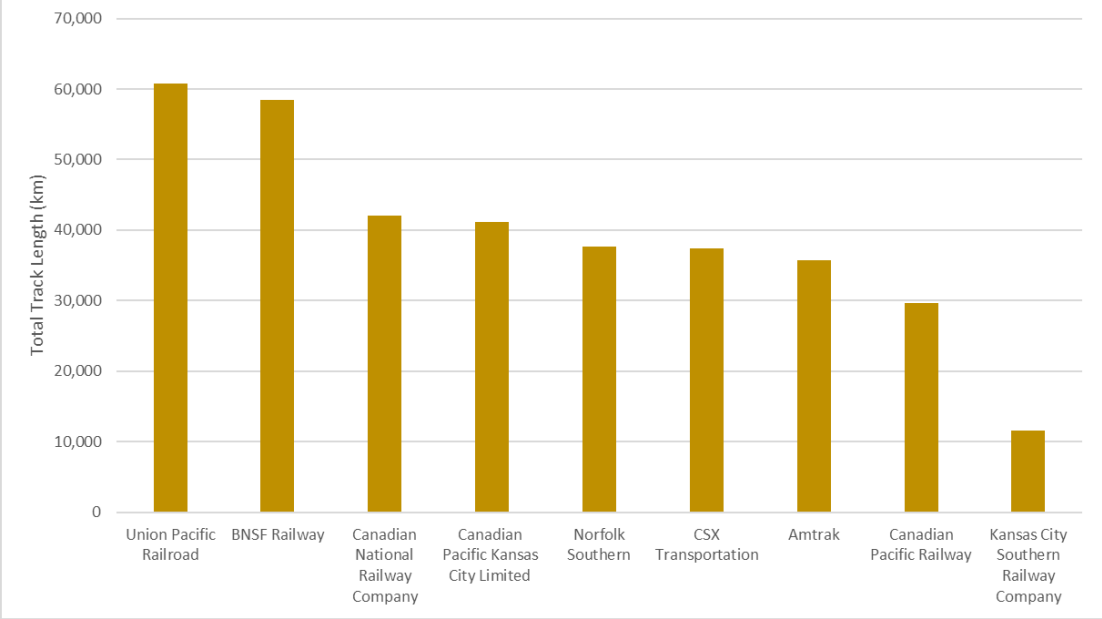


Figure 57. Class 1 Carrier Subnetworks ranked by track length (largest to smallest length)

Combining network characteristics of interest and traditional measures such as total track length can further enhance the evaluation of the subnetworks. For example, Union Pacific is the largest network by track length, yet maintains a middle of the pack efficiency and network diameter size. Taking these parameters together and without using the network visual map projection, Union Pacific's network can be concluded to be denser and more grid-like in shape than other subnetworks. The map/visual projection confirms this conclusion.

Chapter 10: Conclusion and Contribution

This thesis overviews network analysis and graph theory methods, outlines an established approach to represent network characteristics related to all pairs shorted path computations, eccentricity, and a novel approach to do the same, arithmetic efficiency, and applies the old and new methods to small example network and a detailed network data set representing the largest and most used portions of the North American railroad network, and relevant Class I subnetworks. The methods used to analysis network topology add to the potential language and characteristics usable to describe rail networks shape, size, and topology. In addition, some of the network characteristics can be used to rank subnetworks for comparison and analysis purposes.

The new arithmetic efficiency model allows for the computation of efficiency related network characteristics where the units of the link weights are preserved, so the ultimate network characteristics can be expressed with units that are equivalent to the link weights. The preservation of units allows for a clearer expression of the meaning of the efficiency characteristics and allows comparison and equivalency to established unit preserving methods such as eccentricity. The same methods detailed here can be expanded to other unit weighted networks where network size is not equivalent and there is no standard link unit weight with which to normalize network efficiency.

The detailed North American Railroad and Class I carrier subnetwork case studies document the application of eccentricity and arithmetic efficiency to ranking nodes,

identifying critical central and peripheral nodes, producing color map network visualizations, and describing each network with overall characteristics that can be used to rank entire networks. The ranking and terminology could assist railroad network stakeholders evaluate and improve their network for use or resiliency enhancements, or to plan for disruption and choose ideal locations for infrastructure where geometric centrality or average centrality is desirable. In addition, peripheral network sections can be identified as at-risk locations that cannot be reached as easily from the bulk central network.

This research can be expanded to factor other types of network link weight, and other types of networks. For the railroad network analysis in this study, the geodesic distances for each link are used, but an interesting comparative analysis could be performed combining the distance metric with a top railroad speed or average speed metric to arrive at a time to travel link weight. In this weighting system the actual central node might be different by travel time to the peripheral destinations, as railroad line speed can vary significantly depending on track quality, curvature, and acceleration time. Another area of weights that could be explored is link cost, combining further metrics such as land incline and any usage fees, or documented traversal costs, to arrive at a link weight in terms of monetary cost of traversal for the average loaded railroad locomotive. The railroad network could then be analyzed in terms of centrality of nodes on a cost and economic basis. Finally, other types of networks can be analyzed using these link weight unit preserving methods, such as

sea routes, combined sea and land transportation routes, or entirely separate network types such as electronic circuit networks, computer networks, or social networks.

Bibliography

Association of American Railroads. (2023). *Freight Railroad Fact Sheet.* The Association of American Railroads. https://www.aar.org/wp-content/uploads/2020/08/AAR-Railroad-101-Freight-Railroads-Fact-Sheet.pdf

Chan, T. M. (2010). More Algorithms for All-Pairs Shortest Paths in Weighted Graphs. *SIAM Journal on Computing*, *39*(5), 2075–2089. https://doi.org/10.1137/08071990X

Csardi, G. (2023). *Igraph Documentation*. https://igraph.org/r/doc/distances.html

Dankelmann, P., Goddard, W., & Swart, C. S. (2004). The average eccentricity of a graph and its subgraphs. *Utilitas Mathematica*, *65*, 41–52.

Euler, L. (1726). Solutio problematis ad geometriam situs pertinentis. *Comment. Acad. Sci. U., Petrop 8*, 128–140.

Geospatial at the Bureau of Transportation Statistics. (n.d.). Retrieved September 13, 2022, from https://data-usdot.opendata.arcgis.com/

Gleich, D. F. (2015). PageRank Beyond the Web. *SIAM Review*, *57*(3), 321–363. https://doi.org/10.1137/140976649

Hage, P., & Harary, F. (1995). Eccentricity and centrality in networks. *Social Networks*, *17*(1), 57–63. https://doi.org/10.1016/0378-8733(94)00248-9

Hamed, M., Mao, Y., Ayyub, B. M., Elsibaie, M., & Omar, T. (2022). Connectedness Efficiency Analysis of Weighted U.S. Freight Railroad Networks. *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, *8*(4). https://doi.org/10.1115/1.4054326

Harary, F. (1969). *Graph Theory*. Addison-Wesley.

Interstate Commerce Act, Code of Federal Regulations, Title 49 § Subtitle B, Chapter X, Subchapter C, Part 1201 (1887). https://www.ecfr.gov/current/title-49/subtitle-B/chapter-X/subchapter-C/part-1201

Jordan, C. (1869). Sur les assemblages de lignes. *Journal für die reine und angewandte Mathematik, 70*, 185–190.

Latora, V., & Marchiori, M. (2001). Efficient Behavior of Small-World Networks. *Physical Review Letters, 87*(19), Article 19. https://doi.org/10.1103/PhysRevLett.87.198701

Mao, Y. (2021). *Optimizing Rail Network Topology Attributes for Track Reclassing, Acquisition and Repurposing*. University of Maryland.

Newman, M. (2010). *Networks: An Introduction* (1st edition). Oxford University Press.

Newman, M. E. J. (2001). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, *64*(1), 016132. https://doi.org/10.1103/PhysRevE.64.016132

Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review E*, *70*(5), 056131. https://doi.org/10.1103/PhysRevE.70.056131

Peixoto, T. (2020, July 8). *Graph-tool performance comparison*. https://graph-tool.skewed.de/performance

Saadat, Y., Ayyub, B. M., Zhang, Y., Zhang, D., & Huang, H. (2019). Resilience of Metrorail Networks: Quantification With Washington, DC as a Case Study. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, *5*(4), 041011. https://doi.org/10.1115/1.4044038

Seidel, R. (1995). On the All-Pairs-Shortest-Path Problem in Unweighted Undirected Graphs. *Journal of Computer and System Sciences, 51*(3), 400–403. https://doi.org/10.1006/jcss.1995.1078

Takes, F. W., & Kosters, W. A. (2013). Computing the Eccentricity Distribution of Large Graphs. *Algorithms, 6*(1), Article 1. https://doi.org/10.3390/a6010100

Wang, J., Lu, M., Belardo, F., & Randić, M. (2018). The anti-adjacency matrix of a graph: Eccentricity matrix. *Discrete Applied Mathematics, 251*, 299–309. https://doi.org/10.1016/j.dam.2018.05.062

West, D. B. (2000). *Introduction to Graph Theory* (Second edition). Pearson College Div.