ABSTRACT

Title of Thesis:	OPTIMIZING RAIL NETWORK TOPOLOGY ATTRIBUTES FOR TRACK RECLASSING, ACQUISITION AND REPURPOSING
	Yujie Mao, Master of Science, 2021
Thesis Directed By:	Professor Bilal M. Ayyub, Department of Civil and Environment Engineering

In the United States, railroads have been well established over several decades to meet the needs and demands of transporting freight and passengers. Given the mature rail network, the network's enhancement is often done by reclassing or repurposing existing rails or acquiring new rails. This thesis focuses on optimizing the topology of a network and proposes a methodology to optimize its efficiency using track reclassing, acquisition, and repurposing as means for topological changes. The network efficiency is selected as the primary network attribute. Due to the computational burden associated with computing network efficiency, this study proposes the use of the standard deviation of the node degree as an approximation of network efficiency in identifying optimal solutions. The approximate solution produces results reliably with computational efficiency and accuracy. A case study of a single Class I rail network is introduced to compare the solutions of these two optimization criteria. The results show that the standard deviation of node degree can be used to obtain an optimal solution and offers an adequate and more computationally efficient approximation than the direct use of network efficiency with differences less than 0.2% based on adding 40 links.

OPTIMIZING RAIL NETWORK TOPOLOGY ATTRIBUTES FOR TRACK RECLASSING, ACQUISITION AND REPURPOSING

by

Yujie Mao

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

2021

Advisory Committee:

Professor Bilal M. Ayyub, Chair

Professor Paul Schonfeld

Dr. Jennifer F. Helgeson (Fed), National Institute of Standards and Technology

Acknowledgements

First, I would like to pay particular appreciation to my advisor, Professor Bilal M. Ayyub, for his helpful advice and support of the thesis at the University of Maryland, College Park. He always advises and guides me patiently. It is wholeheartedly appreciated that without his persistent help, this thesis's goal cannot be realized.

I wish to show my gratitude to my thesis committee, Professor Paul Schonfeld and Dr. Jennifer F. Helgeson (Fed), for their attention to my work. I want to thank all the people whose assistance was a milestone in the completion of this project. I am also indebted to Dr. Yalda Saadat for her valuable suggestions, encouragement, and comments on my thesis's technical writing.

I wish to acknowledge the support and great love of my parents. They kept me going on, and without their support, this work would not be finished. Many thanks to all my friends particularly, Mr. Siqi Cao, Ms. Sally Saleem, Mr. Qianglong Wang, Mr. Chang Liu.

Table of Contents

Chapter 1. Introduction

The concept of networks is used in many fields. The Internet, transportation including railroads, supply chain, power systems, and so on are all networks. Planners and designers can increase the efficiency of a particular system based on the study of its underlying network. Researchers have come up with methods and concepts related to networks to address varied needs and study pursuits. For example, Seman et al. (2012) proposed ways to optimize the Internet with an improved network centric model. Dorigo and Gambardella (1997) used the ant colony method to solve the traveling salesman problem within the context of a network. Zhao et al. (2019) optimized the structure of China Rail Express to lower costs using network concepts.

Of particular interest is studying network topology in order to characterize and enhance the understanding of network characteristics. Network models provide basis for enhancing the efficiency of a network. For example, researchers can optimize the railroad network in the United States and make the network more efficient. An appropriate topology can enhance the network's performance and reduce losses when the network is disrupted by enhancing recovery profiles, robustness, and resilience. This chapter introduces some background of network topology, the Class I railroad in the United States, and some practical ways to expand the railroad network.

1.1. Network Topology

Networks, such as transportation networks including railroads, can be abstracted for defining topology. Network topology describes the relationship between nodes and links.

Researchers analyzed such topology types of networks to gain insights based on studying its attributes for network enhancements. For example, the adjacency matrix describes the connectivity of the network (Newman 2018). The number of nodes representing things such as rail waypoints or cities and links representing railroads could reflect the size of the network. Such a representation offers a basis to estimate network efficiency as a measure of the performance of the network (Latora and Marchiori, 2001).

Many researchers focused on and examined network vulnerability and resilience, given a disruption to a network. Zhang et al. (2013) assessed the resilience of the Shanghai metro network based on hypothesized disruptions. They estimated the impact of disabled nodes and links for the network and found the optimal recovery strategy for the disability of the most critical node. Chakraborty and Ikeda (2020) studied the topological properties of the global supply chain network by testing the nodal, structural, and flow characteristics of the network.

Other researchers examined network topology to estimate the characteristics of a network quantitatively. Network topology describes the relationship between nodes and links as a model that focuses on its transmissive ability for achieving a particular functionality. Newman (2018) provides background information on typical standard networks, such as technological networks, information networks, social networks, and biological networks, and several fundamental concepts, such as node degree, average node degree, characteristic path length, and so on. Based on network theories, the networks are categorized into unweighted, weighted, undirected, and directed networks that can be used to describe the network more accurately (Newman 2018). For example, using the distance of links as weights, the abstracted models are closer to reality and more convenient for the calculation. Also, the directed network can depict the structure of some specific networks, such as the neural network. Latora and Marchiori (2001) introduced network efficiency as an estimate of its ability to transmit goods, information, etc.

Johnson et al. (2020) estimated the influence of spatial correlation associated with failures based on the topological characteristic of networks and the dimensional properties of hazards. They used more than a thousand random networks to assess the relationship among topologies and failures (Johnson et al. 2020). Saadat et al. (2019) analyzed Metrorail networks in Washington, DC, with network topology methods. Node degree, average node degree, characteristic path length, network efficiency, efficiency impacts of nodes and links, and resilience are used to estimate the Metrorail networks. She tested the efficiency impacts with the disabled nodes or links. For the resilience analysis, she advised recovery strategies to reduce the loss when the most impactful nodes on the network's performance are disabled (Saadat et al. 2019).

1.2. Railroads in United States

In the United States, the significant railroads include Amtrak, Burlington Northern and Santa Fe (BNSF) Railway, Canadian Pacific Railway, Chessie Seaboard Consolidated (CSX) Transportation, Kansas City Southern Railway, Norfolk Southern Railway, and Union Pacific Railway. By the end of 2020, the railroads' passenger miles are up to

583,637,283 miles (Bureau of Transportation Statistics 2021). Railroads are also used for freight movement. Several modes of freight shipments are used in America, such as truck, rail, water, air, and pipeline. Figure 1.1 presents the percentage of freight of these modes. Freight transportation in ton-miles by rail accounts for 39.5% (Bureau of Transportation Statistics 2010). The rail has the highest proportion of any other transportation mode. The freight railroads can have considerable economic impacts. With more jobs and revenue provided by freight railroads, people have more income to stimulate the economy, and the government has more tax revenue to fund projects. Based on the Association of American Railroads (2019), the freight railroads could provide many jobs. Around 140,000 employees worked for the freight railroads. The annual average salary for them is more than $94,400, and they have $38,500 in fringe benefits. The tax revenues from railroads are around $26 billion per year. Almost one-third of U.S. exports depend on railroads.

Freight railroads are categorized into four different classes: (1) Class I railroad, (2) regional railroad, (3) local line-haul railroad, and (4) switching and terminal carriers. Among all the types, the Class I railroad is the most important with its freight revenue and employees and is defined as those with an operating income of $433.2 million or more (Association of American Railroads 2013). The Class I rail network accounts for 90% of employees and 93% of freight revenue among all types of freight railroads. Figure 1.2 presents the operating revenues of Class I railroads from 1990 to 2019 (Statista 2020). The figure shows an increasing trend of Class I railroads' payments from

28.4 billion dollars in 1990 to 80 billion dollars in 2019. The revenues almost tripled. Freight railroads in the United States have become a considerable economic resource.

The seven significant railroads in the United States are introduced in Table1.1 (Brooks 2018). These railroads pass through 44 states in North America and some areas in Canada and Mexico, and account for 68 percent of the railway mileage (Association of American Railroads 2020). The Union Pacific Railroad is the largest. However, the BNSF railroad has the highest annual revenues among all seven Class I railroads. Figure 1.3 shows the annual revenues of Class I railroads (Statista 2020a). Based on the figure, the Union Pacific Railroad provided the second-highest revenues in 2019, which is 21.708 billion dollars. The highest revenue received by BNSF Railroad is 23.133 billion dollars in 2019.

Table 1.1. Seven significant railroads in the United States

Name	Date of incorporation	Description
Union Pacific railroad	1862	Union Pacific railroad operates around 52,00 miles in 23 states
BNSF Railway	1849	BNSF Railway owns 32,500 miles of railroads across 28 states
CSX Transportation	1827	CSX Transportation owns 21,000-mile railroads in 23 states
Canadian National Railway	1918	Canadian National Railway has become the largest rail network in Canada. It has 19,600 miles of railroad in 16 U.S. states and 8 Canadian provinces
Norfolk Southern Railway	1838	Norfolk Southern Railway owns 19,500 miles of tracks in more than 20 states
Canadian Pacific Railway	1881	Canadian Pacific Railway has 15,000 miles of rail in 13 states
Kansas City Southern	1887	Kansas City Southern operates 6,000 miles of rail in the United States and Mexico

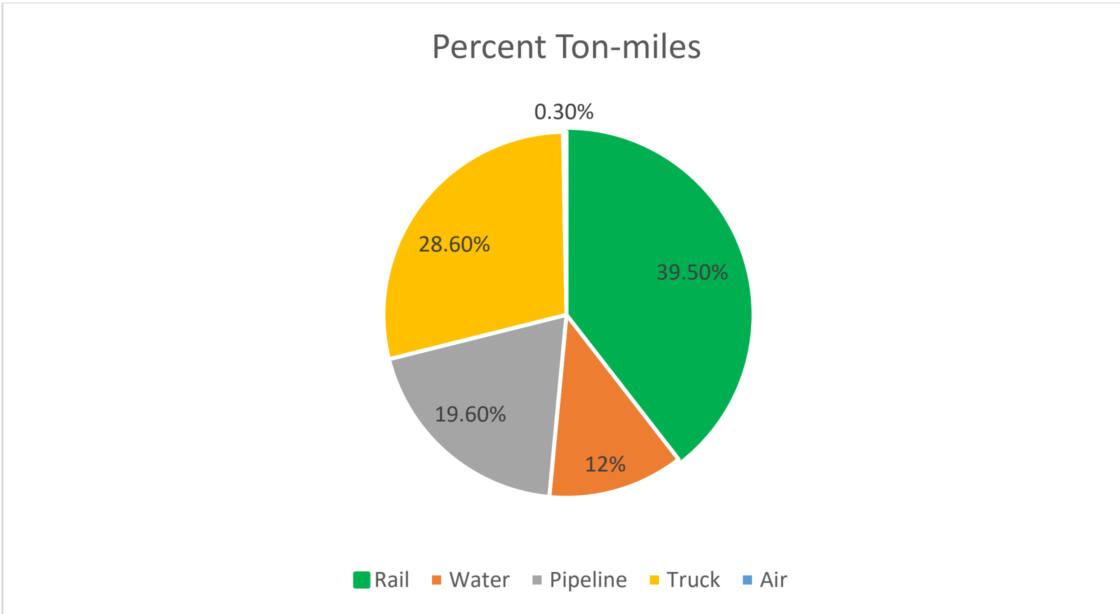


Figure 1.1. Percent ton-miles of each freight transportation mode (Bureau of Transportation Statistics 2010)

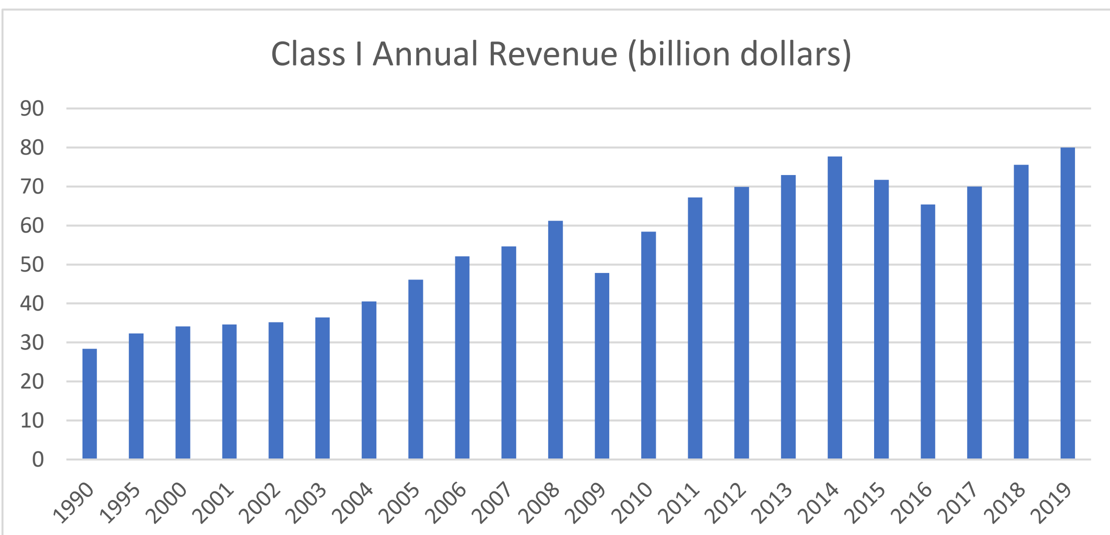


Figure 1.2. Operating revenues of Class I railroads from 1990 to 2019 (Statista 2020)

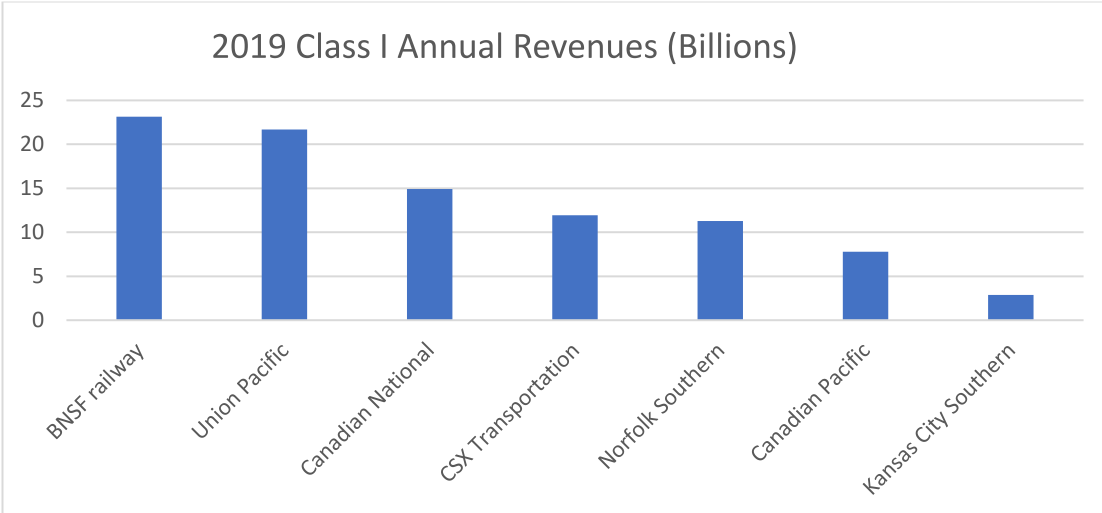


Figure 1.3. The annual revenues of railroads with different owners (Statista 2020a)

1.3. Network Enhancement Needs and Challenges

With the expansion of the economy, freight railroads developed dramatically in the past few decades with additional usage and good transportation. Railroads transported 168,000 carloads of trailer and containers in 1955 and 14.5 million containers and trailers in 2018 (Association of American Railroads 2020a). The Class I rail network plays a significant role in freight and is considered mature nowadays in terms of rails and associated networks. Further expansion is possible by changes in the use of available railroads. This study focuses on the expansion of the network efficiency by changes in the use of railroad through repurposing, reclassing, abandonment, etc. The reuse and repurposing of the isolated railroads can increase efficiencies and reduce cost with operational enhancement.

1.3.1. Network Expansion

Researchers examined network performance for enhancement by a wide range of methods. Dorigo and Gambardella (1997) used the ant colony optimization method to solve the traveling salesman problem. They accumulated and updated the result of each iteration to find the shortest route passing through all network nodes. In 2001, Zong et al. introduced harmony search to reduce the number of iterations solving the traveling salesman problem. They upgraded the initial network sequentially to get the shortest path connecting all network nodes (Zong et al. 2001). Feng et al. (2018) selected the Simulated Annealing Algorithm (SAA) to optimize the operation cost.

For freight networks, enhancement is typically based on increasing network efficiency. However, solving the traveling salesman problem cannot reduce transporting time given random origin and destination. In this thesis, network efficiency, as introduced by Latora and Marchiori (2001), is used as the standard to assess the network. Based on by Latora and Marchiori (2001), the network efficiency is used to estimate how well the information is transported in the network. Higher network efficiency means the network is more efficient with the same number of nodes. Thus, network efficiency can help find the optimal network for the extension of the network. Saadat et al. (2020) added three loop lines to Washington, DC Metro network to increase network efficiency and reduce the impact when the network suffers from a disruption. However, if the network is extensive with thousands of nodes and links, the calculation would become inefficient. A substitutable method is necessary to reduce computation. Bai et al. (2018) proved a positive relationship between the standard deviation of node degree (S) and network efficiency. The standard deviation of node degree could help reduce the calculation to find the optimal solution for network extension. The use of the standard deviation of node degree is investigated in this thesis.

1.3.2. Reclassing, Reuse, and Repurposing of Railroads

The railroad in the United States is mature as a transport network. The railway optimization cannot blindly build new railways without considering land use limitations and many surrounding urban and well-developed areas. New railways are not economical and take up much public space. Thus, taking advantage of the existing rails is necessary. A planner can upgrade regional railroad and local line-haul railroad to Class I railroad to enhance an underlying network. Reclassing, reuse, and repurposing can reduce the cost of optimizing the network.

According to the Merriam-Webster dictionary, the word "repurpose" means to give a new purpose or use to stuff and the word "reuse" means to use again in another way or after reprocessing. These concepts can be used to optimize the freight railroads. In the United States, thousands of miles of railroads were abandoned (Abandonedrails 2021). Several reasons can explain the abandonment. When the rails are no more profitable or built mostly for mines or industry, these rails are more likely to be abandoned. However, if decision-makers accept some of these rails into the freight rail network, the rail network's efficiency may be increased with less cost, and the solution is more practical to operate. Di Ruocco et al. (2017) pointed out the importance of using the existing railroad. They introduced some case studies proving the potential ways for sustainable regeneration.

Zhang et al. (2020) advised reusing the abandoned railroads as urban transportation to increase the passenger flow between old and new city centers. They found three strategies to reuse the railroads, including converting the abandoned rail into a new rail system, redevelopment of urban land, and reconstruction of the new public space system for the city. Using the existing resources to optimize the Class I rail network can increase network efficiency and is more practical to operate.

1.3.3. Challenges

The network enhancement needs, as provided in the previous sections, require addressing the following challenges associated with network enhancement:

- Estimating topological properties of Class I railroads in the United States;
- Assessing the criticality of nodes and links of the network;
- Optimizing the network with the potential reused or repurposed railroads and identification rails with great potential for development; and
- Enhancing the calculation efficiency of optimization with an appropriate approximate method suitable for network enhancement.

1.4. Organization of Thesis

Chapter 1 introduces this thesis containing the background of network topology, some railroad information in the United States, the network's enhancement needs and challenges, and the organization of view.

Chapter 2 is the literature review of network topology, efficiency impact, network resilience, and reuse of disused railroads. Also, this chapter will introduce some existing optimization methods with advantages and disadvantages. This chapter will state the gap and objectives of the thesis.

Chapter 3 presents the methods and processes with simple examples to analyze network topology and optimize the network with less calculation. This chapter compares two types of analysis to maximize network efficiency when adding links to the network.

Chapter 4 provides a case study containing how the decision-makers could use the method to optimize the Class I railroad network.

Chapter 5 presents the conclusion, contribution, and limitation of the thesis. Also, this chapter outlines future work for railroad networks.

Chapter 2. Optimization of Network Analysis: A Literature Review

2.1. Network Topology

Network topology depicts the relationship among nodes and links of the network. Based on the topological properties of the network, researchers can get insights of the network precisely and effectively. Garrison and Marble (1962) first mapped transportation networks topologically. They introduced links and nodes as the components of the network and described the characteristics of the network by defining the connection matrix, structural patterns, and cyclomatic numbers. Musso and Vuchic (1988) provided elements defining the network characteristics of metro networks systematically. They defined some indicators of network topology, such as length of the network, network complexity indicator, the density of the network, and so on.

Latora and Marchiori (2001) defined network efficiency to estimate the transmission ability of a network, which is used in the analysis of the Washington D.C. Metro network and Shanghai Metro network (Saadat et al. 2019; Saadat et al. 2020; Zhang et al. 2018). Also, Johnson et al. (2020) focused on the effect of spatially- correlated failures on the robustness of the network by introducing network efficiency. Small world networks and scale-free networks are also identified as significant properties. Watts and Strogatz (1998) introduced the Watts-Strogatz model to produce a small-world network based on the characteristic path length and network clustering coefficient. A scale-free network is a network with the node degree distributed according to a power law.

Zhang et al. (2018) estimated the Shanghai network based on network efficiency. They listed the ten most critical nodes and projected the recovery strategies when the most critical node is disabled due to the disruption on the network. Also, Saadat et al. (2019) assessed the Washington D.C. Metro network identifying the critical nodes and links. They selected the most significant node for the recovery strategy analysis that estimated the resilience loss for different recovery processes.

2.2. Problems of Optimization

Different objectives lead to various optimization problems. The traveling salesman problem is about the shortest route passing through all nodes of a network. This problem is an NP-hard problem in combinatorial optimization. This problem assumes a list of cities and the distance between each pair of cities and asks the shortest route that passes all cities. The route returns to the original city. Ant colony and bee colony optimization methods are used to solve this problem (Dorigo and Gambardella 1997; Nikolić and Teodorović 2013).

The transit network design problem is more about public transportation. Researchers try to design the shape of the public network to minimize or maximize intended objectives, such as maximizing the number of passengers and minimizing the cost of time (Fan and Machemehl 2006; Nikolić and Teodorović 2013; Nayeem et al. 2014).

The network flow problem is viewed as a transportation problem with the cost of transporting materials. The directed networks are usually used in the analysis. Decision-makers try to minimize the cost with various demands and parameters. This problem can be categorized into maximum flow problems, minimum-cost flow problems, and multi-commodity flow problems. Table 2.1 summarizes the problems mentioned previously.

In this study, the optimization's objective is to find the optimal solution of network expansion with cost-efficient computation. The optimal solution is the network with the highest network efficiency. When adding more links to the network, these links' potential locations lead to different network efficiency. However, if the network size is considerable, the comparison of potential solutions is complex and would cost much time. The computation is not efficient. Thus, an alternative method is necessary to find the optimal solution.

Table 2.1. Summary of optimization problem

Problems	Objective
Traveling salesman problem	Finding the shortest route passing through all nodes of the network
Transit network design problem	Finding the network topology that maximizes or minimizes the intended purpose
Maximum flow problem	Maximizing the flow into the sink terminals
Minimum-cost flow problem	Minimizing the cost for transporting a specific number of flows
Multi-commodity flow problem	Finding an assignment of all flow variables to satisfy the demands
Maximum network efficiency	Finding the network with the highest efficiency

2.3. Optimization Methods

Researchers optimize the network in order to increase efficiency and reliability. Murray-Tuite and Mahmassani (2004) introduced a game between an "evil entity" and the traffic agency. In this study, the agency decides the routes of vehicles while the evil entity fully exploits the network's disruption. They selected the vulnerability index to identify the importance of a particular link. Disabled links with higher indexes are more influential to the network. Decision-makers can use this information to reduce the influence of disruption with a pre-arranged planning. Zhang et al. (2018) estimated the change of network efficiency with a disrupted node finding that the influence of disruption is dependent on node degree and the impact of network efficiency. Also, they designed the recovery strategy to reduce the resilience loss. Saaddat et al. (2020) enhanced the Washington D.C. metro network based on a pre-failure strategy. This method reduced the characteristic path length. They reduced the efficiency impact of disabled nodes and links by adding loop lines in the network.

The ant colony optimization method imitates the ants to find the shortest paths traveling all network nodes (Dorigo and Gambardella 1997). The network is unweighted and undirected. They assigned ants in each simulation, letting them go through all nodes, and the routes are used as information accumulated for further simulation. Based on the information, the solution is updated after each simulation until the end of the computation. Zong et al. (2001) also solved this problem by using the harmony search method to do simulations. And they compared the results and updated the solution.

Nikolić and Teodorović (2013) solved the transit network design problem by bee colony optimization. They studied bees' behavior in nature and generated artificial bees to get a feasible solution to the problem. Furthermore, with the information shared by all bees, the optimal solution can be identified. Nayeem et al. (2014) introduced a genetic algorithm to solve the transit network design problem. They maximize the number of passengers and minimize the transfer number and traveling time. For the network flow problem, Majumder et al. (2017) used a genetic algorithm to solve the maximum flow problem. They introduced the expected value model and chance-constrained model. Jiang et al. (2020) parallelized network simplex algorithm to solve the minimum-cost flow problems.

However, these methods cannot solve the maximum network efficiency problem for the expansion of the network efficiently. The computation using network efficiency is expensive and complicated. Bai et al. (2018) found the relationship between the network efficiency and the standard deviation of node degree. The network efficiency increased with the rise of the standard deviation of node degree. Thus, brute force is selected, and the standard deviation of node degree is used instead of network efficiency to find the optimal solution. Table 2.2 presents the optimization measures and methods.

Table 2.2. Summary of optimization measures and methods

Optimization methods	Corresponding problem
Network efficiency (Murray-Tuite and Mahmassani 2004)	Vulnerability of links
Network topology (Zhang et al. 2018)	Topology analysis of Shanghai metro network
Network topology (Saadat et al. 2019)	Topology analysis of Washington D.C. metro network
Pre-failure strategy (Saadat et al. 2020)	Reduction of the efficiency impact of disabled nodes and links
Ant colony optimization method (Dorigo and Gambardella 1997)	Traveling salesman problem
Harmony search (Zong et al. 2001)	Traveling salesman problem
Bee colony optimization (Nikolić and Teodorović 2013)	Transit network design problem
Genetic algorithm (Nayeem et al. 2014)	Transit network design problem
Genetic algorithm (Majumder et al. 2017)	Maximum flow problem
Parallel network simplex algorithm (Jiang et al 2020)	Minimum-cost flow problems

2.4. Reclassing, Reuse, and Repurposing of Railroads

Given the identifying problems and optimization methods, a practical solution to expand the Class I railroad network can significantly impact. Building new rails are expensive and sometimes not practical due to unavailable land public space. One way to expand the railroad network is to upgrade tracks to another class level. Another way is to reuse or repurpose the abandoned or disused railroads.

Railroads are well developed in the United States, and some of them are disused or abandoned. These disused rails can have other functions. Di Ruocco et al. (2017) introduced Reduce/Reuse/Recycle concepts for railroads' infrastructure. They made fair use of the existing resources for urban built environments. The reuse of the abandoned railroad can be used to recover degraded areas. The Delaware Valley Regional Planning

Commission (1991) identified around 40 inactive rail lines reused in the Delaware Valley region. The commission categorized three types of potential rails that can be used in the future for freight rail services, such as high potential, medium potential, and low potential rails. Zhang et al. (2020) introduced three ways to reuse the abandoned railways containing the converting the disused railways into new rail systems, redesigning the urban land around these railways to stimulate business and tourism, and planning a new public space system for the city.

2.5. Knowledge Gaps and Objectives

The literature review points out a need to have a practical solution for optimizing network efficiency. Based on the literature review, the following knowledge gap helps identify the objectives of the thesis:

- The concept of network topology offers a basis to characterize the network. The topology focuses on the connectivity among nodes and links. However, the topological analysis does not offer a suitable basis to optimize the expansion of a network efficiently;
- Researchers have come up with several methods to solve the optimization problems for networks, such as the ant colony method, bee colony method, and genetic algorithm. However, these methods are not suitable for examining network topology optimization by expansion or reduction;
- Network efficiency offers an appropriate way to find the optimal solution to network expansion or reduction. However, the calculation can be computationally

taxing and complicated when the network is considerable in size. Thus, a new method is needed to reduce the computational effort; and

- Since building new rails in a mature network is not practical and economical, reclassing, acquiring, and repurposing the railroads are required to implement optimization results by practical means.

The primary objectives of the study are to:

- Estimate the topological properties of the Class I rail network;
- Find the optimal solution of the network expansion;
- Reduce the computation of the network optimization and prove the feasibility of the method for network expansion or reduction; and
- Find a practical way to optimize a network.

Chapter 3. Methodology and Illustrative Examples

3.1. Topology Optimization Statement

The topology of a network can be optimized in several ways, such as changing the location of links or adding or removing links with other options related to changes to nodes. In this work, the focus is on link addition or removal under the constraints of not changing the number of nodes, and starting with an initial topology of minimally connected nodes. This study projects method to enhance the efficiency of the whole network. Thus, the optimal solution for the network is based on maximizing network efficiency. Consider a network G_0 with *n* nodes and *m* links, where *m* is larger than or equal to *n-1* and smaller than or equal to *n*(*n*-1)/2. Then, consider *k* links to be added to the network G_t, where 1 < *k* < *n*(*n*-1)/2-*m*. The objective is to find the placement of the additional *k* links to maximize the efficiency *E* of the network. The same case can be formulated in terms of link removal to get to the case of adding *k* links by starting with a fully connected network and removing *n*(*n*-1)/2-*k*.

The objective function for link addition can be expressed as follows:

$$E(G_t) \;=\; max(E(G_0, n, m + k)) \tag{3.1}$$

where *E* is the efficiency of the network; G_0 is the initial network topology; G_t is the new network topology at stage or time *t*; *n* is the number of nodes in network G_0 and G_t, *m* is the number of links in network G_0; *k* is the number of additional links to obtain G_t.

The objective function for link removal can be expressed as follows:

$$E(G_t) = max(E(G_f, n, \frac{n(n-1)}{2} - k)) \quad (3.2)$$

where *E* is the efficiency of the network; G_f is the network with fully connected nodes; G_t is the new network topology at stage or time *t*; *n* is the number of nodes in network G_f and G_t; *k* is the number of removed links to obtain G_t.

The optimization variables are the placement of the additional *k* links necessary to define the G_t topology corresponding to meeting the objective function.

The following constraints are set:

- Number of nodes for G_0, G_t, and G_f: *n*
- Number of initial links of G_0, *m*: n-1 ≤ m ≤ $n(n$-1)/2
- Number of additional links (*k*) set or assumed: n-1-m ≤ k ≤ $n(n$-1)/2-m

The optimal solution results in the network efficiency: *E* for the weights set for the links as 1 in the case of two nodes connected; otherwise, 0.

This optimization problem has a discrete set of potential solutions corresponding to different link placement to define G_t with one or more of the solutions meeting the objective function.

3.2. Background

Meeting the objective of Eq.3.1 requires solving for the shortest path between pairs of nodes (j,i) as follows:

The adjacency matrix A describes a network by the connectivity by k links of the nodes (column i, row j) with weights consisting of elements labeled as A_{ji} such that:

$$A_{ji} = \begin{cases} \text{Weight } if\ there\ is\ a\ link \text{ with a weight } between\ nodes\ i\ and\ j \\ 0\ otherwise \end{cases}$$

(3.3)

Note that the sum is computed over the columns of the matrix. The matrix A has a diagonal of zeros, and is symmetric for cases with links that do not convey directionality, i.e., undirected links.

The node degree K_i of node i is the number of links connected directly to node i. Thus, the average node degree is calculated as follows:

$$\bar{K} = \sum_{i=1}^{n} \frac{K_i}{n} \tag{3.4}$$

where n is the number of nodes; K_i is the node degree of node i.

The length of the shortest path, i.e., distance, in the network between nodes i and j is defined as d_{ij}.

The characteristic path length L is the average length of all possible pairs of nodes in a network shown as follows:

$$L = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ji} \tag{3.5}$$

where d_{ij} is the shortest path length between node i and node j, which is the number of links in the path.

The network efficiency (E) reflects how well the information is transported and is defined as follows:

$$E = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ji}} \tag{3.6}$$

where d_{ij} is the shortest path length between nodes *i* and *j*.

An approximate solution can be based on the relationship between standard deviation of node degree (*S*) and network efficiency (*E*). The standard deviation of node degree (*S*) reflects the distribution of links to the nodes and can be expressed as follow:

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (K_i - \bar{K})^2} \tag{3.7}$$

where *S* is the standard deviation of node degree; *n* is the number of nodes; K_i is the node degree of node I; $\bar{K}$ is the average node degree.

Bai et al (2018) have shown a positive relationship between the standard deviation of node degree (*S*) and network efficiency. Based on the standard deviation of node degree, the calculation to find the optimal solution for network extension could be reduced.

3.3. Fundamental Cases and Variants

Tables 3.1 and 3.2 provide examples of optimization by adding links (*k*) to the initial network $G\_0(n = 6, m = 5)$ with the assumption of a sequential topology, as shown in Figure 3.1. Table 3.1 shows the change of network efficiency with *k* = 1 with 10 ways to add a single link to the network G_0. Table 3.2 shows the change of network efficiency with k = 2 with 22 ways to add two links to the network G_0.

Starting with an initial network $G_0(n, m)$, different connectivity of node by different positions of the links lead to different network efficiency. The initial network G_0 with the same number of nodes and links can have different standard deviation of node degree owing to the different location of links. Networks are estimated based on max E values in case of multiple solutions with the same S. Figure 3.1 to 3.3 shows the relationship between the standard deviation of node degree and network efficiency for networks with 6, 7, and 8 nodes. Generally, networks with a higher standard deviation of node degree are more efficient.

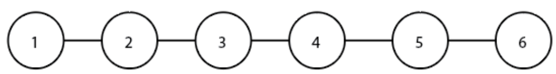


Figure 3.1. Sequential network with n=6 and m=5

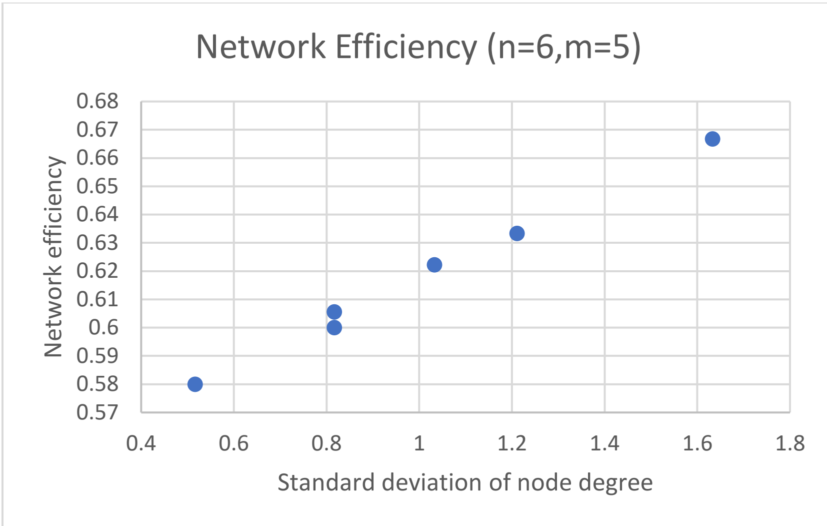


Figure 3.2. Relationship between network efficiency and standard deviation of node degree when *n*=6 and *m*=5

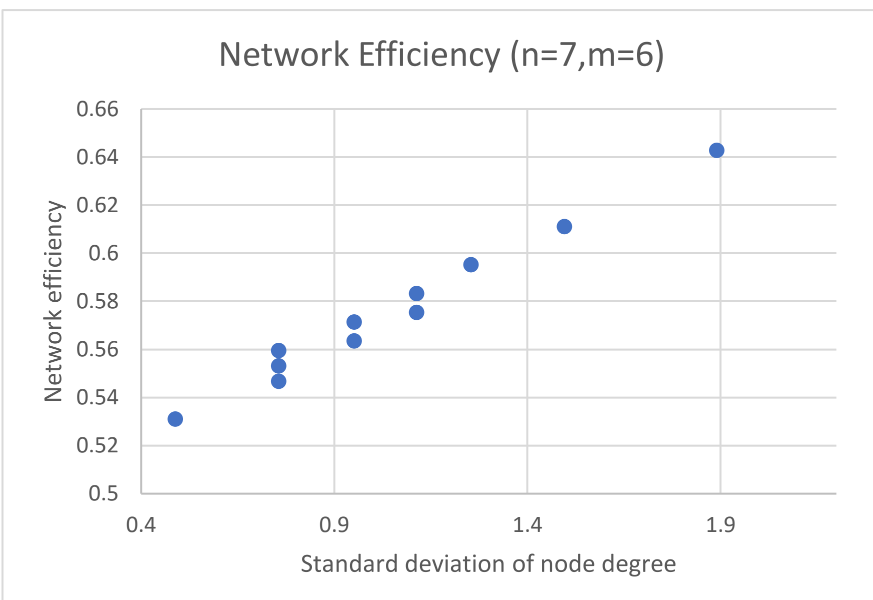


Figure 3.3. Relationship between network efficiency and standard deviation of node degree when *n*=7 and *m*=6

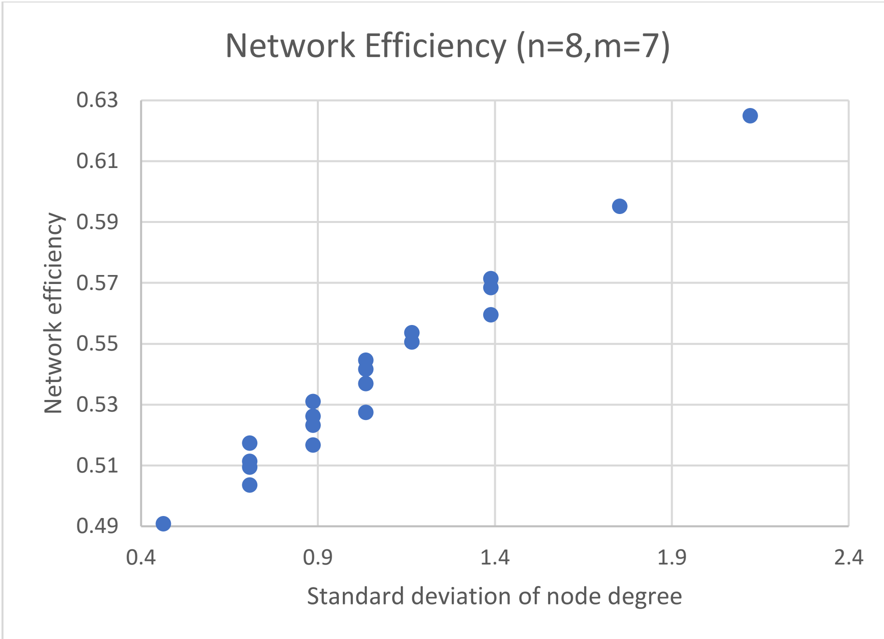


Figure 3.4. Relationship between network efficiency and standard deviation of node degree when n=8 and m=7

Table 3.1 Comparison between the initial network and new network with k=1 rank-ordered by new efficiency

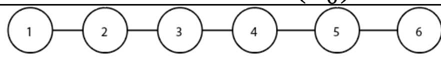
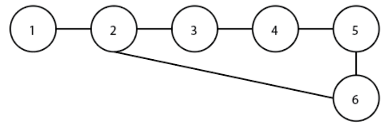
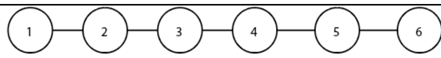
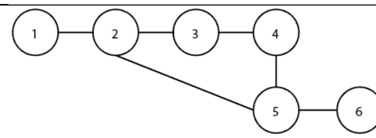
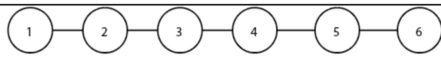
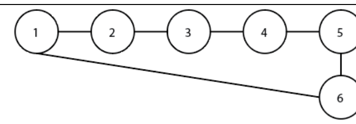
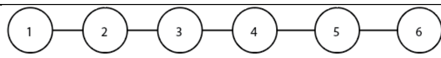
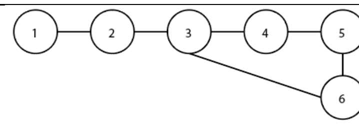
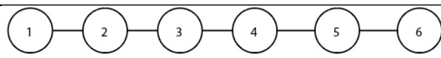
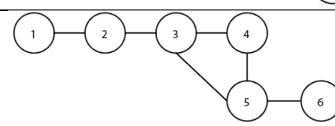
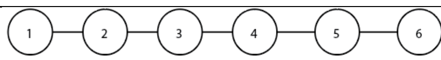
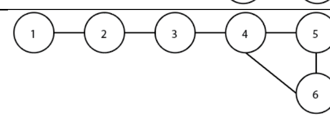
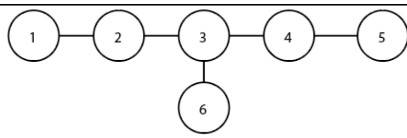
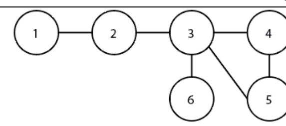
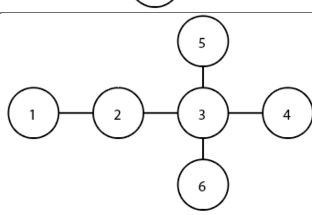
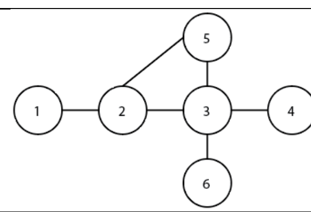
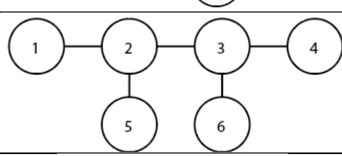
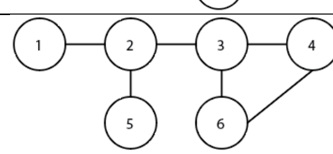
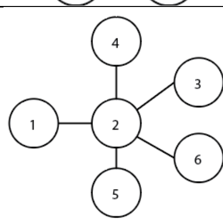
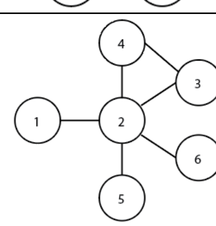
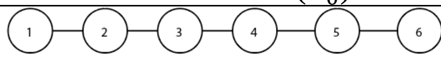
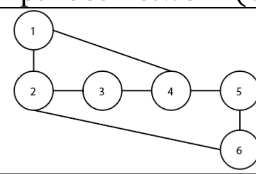
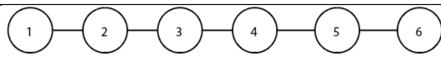
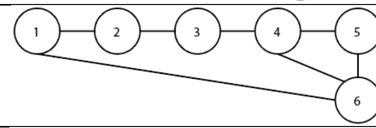
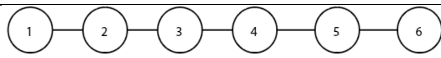
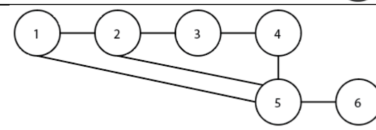
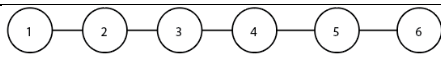
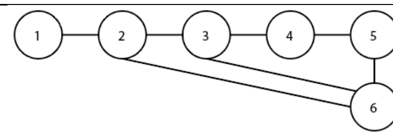
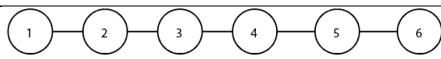
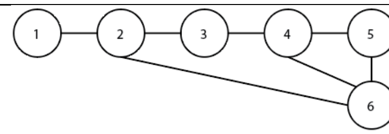
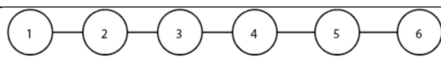
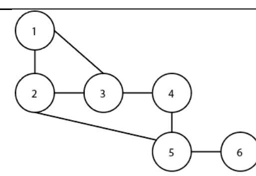
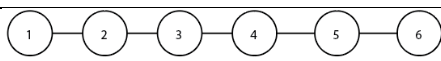
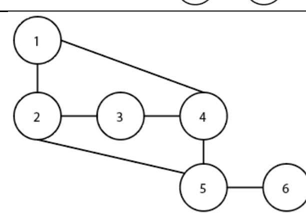
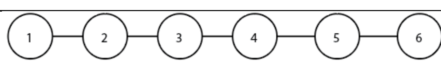
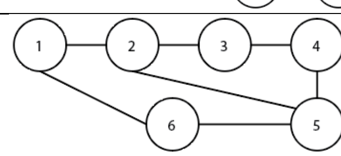
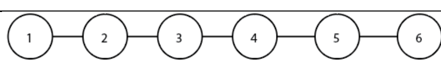
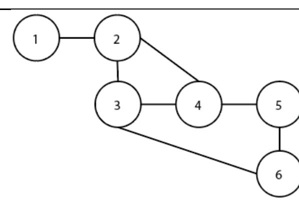
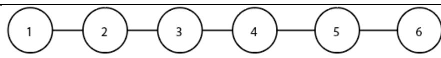
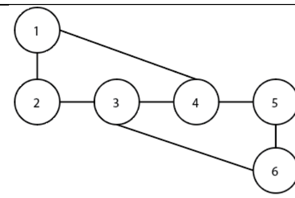
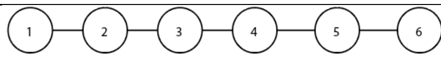
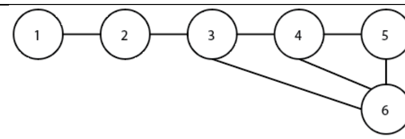
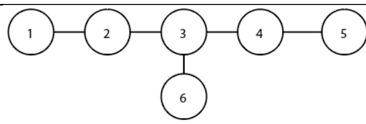
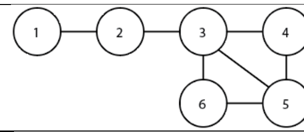
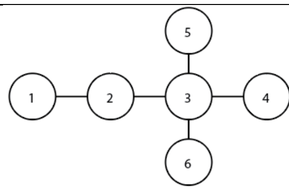
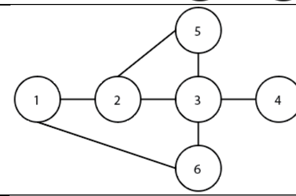
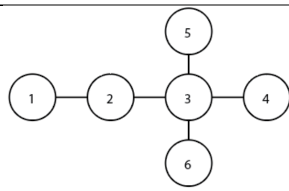
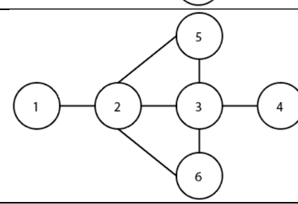
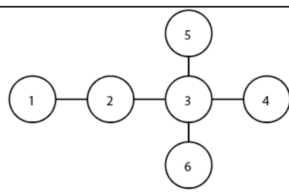
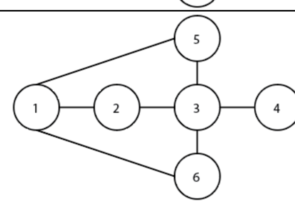
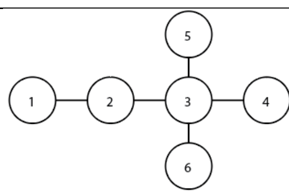
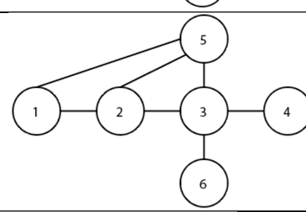
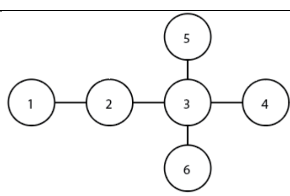
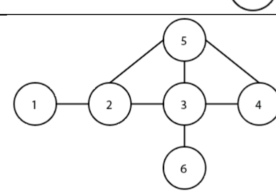
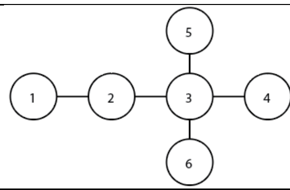
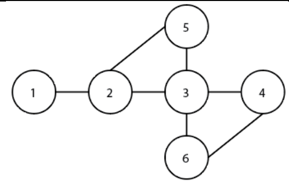
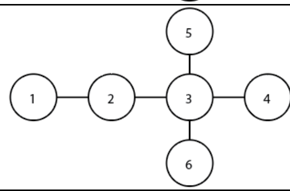
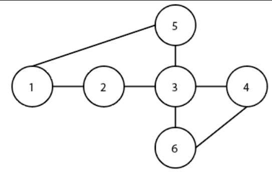
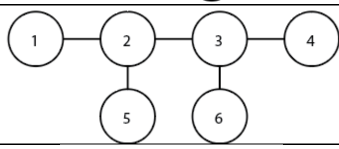
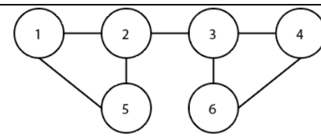
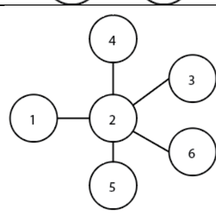
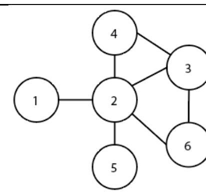
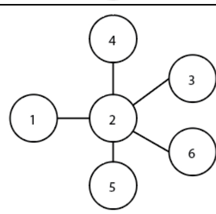
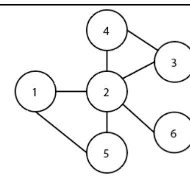
Initial network (G_0)	$E(G_0)$	Expanded network (G_t)	$E(G_t)$
	0.58		0.6778
	0.58		0.6667
	0.58		0.6667
	0.58		0.65
	0.58		0.65
	0.58		0.6333
	0.6056		0.6667
	0.6333		0.6778
	0.6222		0.6556
	0.6667		0.7

Table 3.2. Comparison between the initial network and new network with k=2 rank-ordered by new efficiency

Initial network (G_0)	$E(G_0)$	Expanded network (G_t)	$E(G_t)$
	0.58		0.7333
	0.58		0.7222
	0.58		0.7222
	0.58		0.7111
	0.58		0.7111
	0.58		0.7111
	0.58		0.7111
	0.58		0.7111
	0.58		0.7111

	0.58		0.7111
	0.58		0.6833
	0.6056		0.7
	0.6333		0.7222
	0.6333		0.7222
	0.6333		0.7222
	0.6333		0.7111
	0.6333		0.7111

	0.6333		0.7111
	0.6333		0.7111
	0.6222		0.6889
	0.6667		0.7333
	0.6667		0.7333

3.4. Optimization Solution Approach, Methods, and Results

A brute force by enumeration solution approach is used to enhance the understanding of the most efficient ways to expand a network from an initial topology G_0. Two methods are identified under this approach to find the most efficient expanded network of G_t with an initial or a set number of links: (1) link removal method consisting of removing links from a fully connected network by $n(n\text{-}1)/2$ links until the network becomes the initial network of interest; and (2) link addition method of expanding the initial network. In both methods, optimal solutions are sought, and the two methods are compared.

Since the standard deviation of node degree (*S*) is shown to have a relationship with network efficiency, it is used to find the optimal solution for adding *k* links to a network G_0. The network with six nodes (*n* = 6) is introduced as an example. The number of links (*m*) is between 5 and 15. In order to keep the integrality of the network, i.e., each node has at least one node degree, the sequential network G_0. is set to be the initial network which is shown in Figure 3.1. The links of the sequential network cannot be removed.

Two types of analysis, one driven by max *E* and the other by max *S* links, are introduced to find the optimal solution. The analysis driven by max *E* is to use network efficiency to find the optimal network after adding or removing links, and it can be formulated into four steps:

(1) Start with an initial network or a fully connected network, i.e., each node is connected directly to all other nodes, requiring *n*(*n*-1) links.

(2) Add links to the initial network or remove links from the fully connected network.

(3) Estimate the network efficiency for all networks after adding or removing links.

(4) Selected networks with the highest network efficiency and compute the standard deviation of the node degree of these networks.

The analysis driven by max *S* uses the standard deviation of node degree to find the most efficient network after adding or removing links. This analysis can also be formulated into four steps:

(1) Start with the initial network or a fully connected network, i.e., each node is connected directly to all other nodes, requiring $n(n$-1) links.

(2) Add links to the initial network or remove links from the fully connected network.

(3) Estimate the standard deviation of the node degree for networks after adding or removing links.

(4) Select networks with the highest standard deviation of node degree and calculate their network efficiency.

Thus, two methods and two types of analysis lead to four different results: the results based on link addition method and max *S* analysis, link addition method and max *E* analysis, link removal method and max *S* analysis, and link removal method and max *E* analysis. The superior method or analysis can be identified by comparing *S* and *E* of the network for each result.

Table 3.3 shows the results of two types of analysis for the link removal method. Table 3.4 shows the results for the link addition method. In the table, "*S*" means the standard deviation of node degree, and "*E*" means network efficiency. The addition and removal methods give the same results. For example, for max *S* analysis, the maximum *S* and the corresponding *E* of the network for each *m* based on the addition method are identical to those on the removal method. It is the same for max E analysis.

Thus, this study would focus on the link addition method. Also, in most cases, analysis driven by max *S* has the same results as analysis driven by max *E*. For example, when *m*

= 9 to 15, two types of analysis come up with networks with the same *S* and *E* for each *m* value. However, when *m* = 7 and 6, the results of two different analyses are slightly different.

Figures 3.6 to 3.15 refer to the network efficiency (*E*) and standard deviation of node degree (*S*) for networks with *m* = 5 to 14. The different connections among nodes can lead to the same network efficiency (*E*) and the standard deviation of node degree (*S*). For max *E* analysis, networks with the same efficiency can have different standard deviations of node degree (*S*). In order to estimate the feasibility of max *S* analysis, networks with the highest standard deviation are selected for the comparison. For max *S* analysis, networks with the same standard deviation of node degree (*S*) can have different network efficiencies, i.e., different connections among nodes can result in the same standard deviations of node degree (*S*). Thus, the network with the highest efficiency would be selected for the comparison.

Figures 3.16 to 3.18 refer to the optimal networks that have the highest network efficiency with a specific *k* value. Figure 3.16 shows the results of link addition methods for the initial radial network based on two types of analysis. It contains the network graph, the standard deviation of node degree, and network efficiency for each *k* value. For the radial network, the selected networks estimated by two types of analysis are identical with the same *k* value. Figures 3.17 and 3.18 present the results of two types of analysis for the initial sequential network. For the sequential network, the results of the two types of analysis are almost the same. They only differ when *k*=2.

For the same k value, the initial radial network differs from the initial sequential network in S and E, and in most cases, the initial radial network functions more efficiently. It means that the initial network can influence the optimization in the extension of the network. Based on figures 3.16 to 3.18, the radial network has a downtrend for the standard deviation of node degree (S) with the increase of k, while the standard devotion of node degree (S) for the sequential network would rise when k = 0 and then decrease when k = 4. Also, the radial network usually has higher network efficiency than the sequential network with the same k.

Table 3.3. Link removal optimization method

Number of links (m)	Analysis driven by max S links		Analysis driven by max E links	
	Max S	Corresponding E	Corresponding S	Max E
15	0	1	0	1
14	0.5164	0.96667	0.5164	0.96667
13	0.8165	0.9333	0.8165	0.9333
12	1.09545	0.9	1.09545	0.9
11	1.36626	0.86667	1.36626	0.86667
10	1.36626	0.8333	1.36626	0.8333
9	1.41421	0.8	1.41421	0.8
8	1.36626	0.76667	1.36626	0.76667
7	1.21106	0.71111	0.5164	0.73333
6	0.89443	0.65	0.89443	0.67778
5	0.5164	0.58	0.58	0.58

Table 3.4. Link addition optimization method

Number of links (m)	Analysis driven by max S links		Analysis driven by max-E links	
	Max S	Corresponding E	Corresponding S	Max E
5	0.5164	0.58	0.58	0.58
6	0.89443	0.65	0.89443	0.67778
7	1.21106	0.71111	0.5164	0.73333
8	1.36626	0.76667	1.36626	0.76667
9	1.41421	0.8	1.41421	0.8
10	1.36626	0.8333	1.36626	0.8333
11	1.36626	0.86667	1.36626	0.86667
12	1.09545	0.9	1.09545	0.9
13	0.8165	0.9333	0.8165	0.9333
14	0.5164	0.96667	0.5164	0.96667
15	0	1	0	1

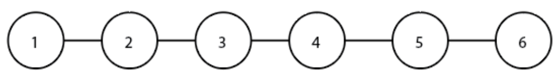


Figure 3.5. Sequential network with n=6 and m=5

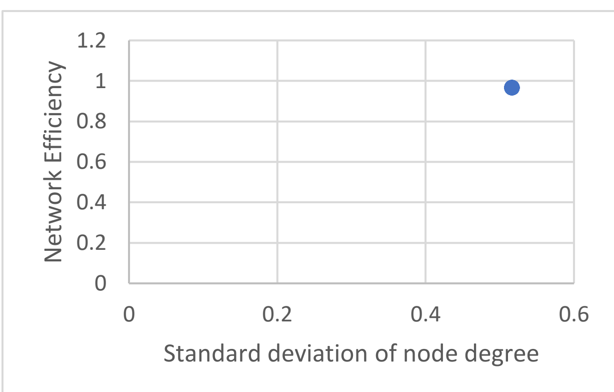


Figure 3.6. The standard deviation of node degree and network efficiency with n=6 and m=14

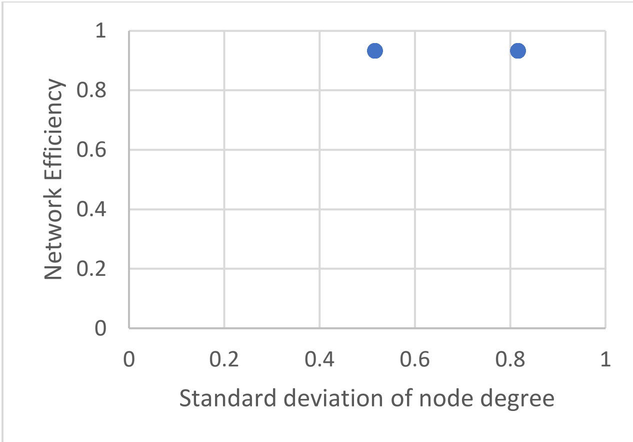


Figure 3.7. The standard deviation of node degree and network efficiency with *n*=6 and *m*=13

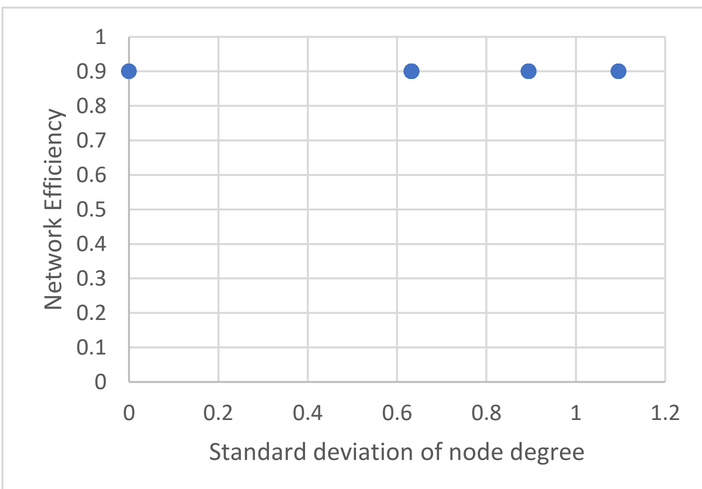


Figure 3.8. The standard deviation of node degree and network efficiency with *n*=6 and *m*=12

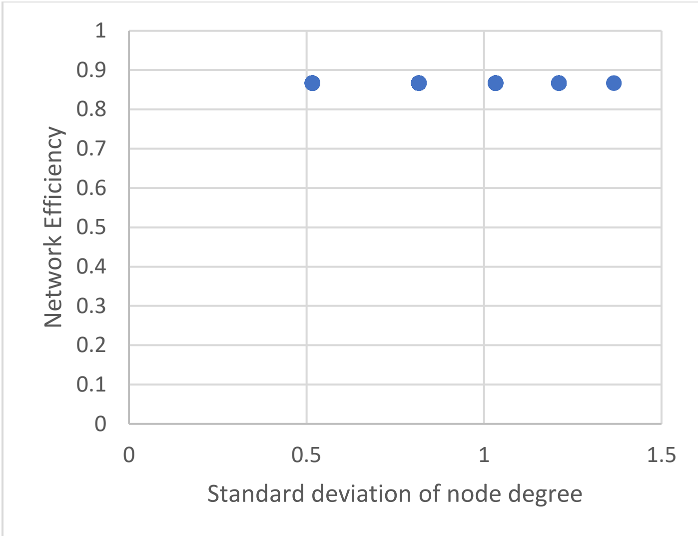


Figure 3.9. The standard deviation of node degree and network efficiency with *n*=6 and *m*=11

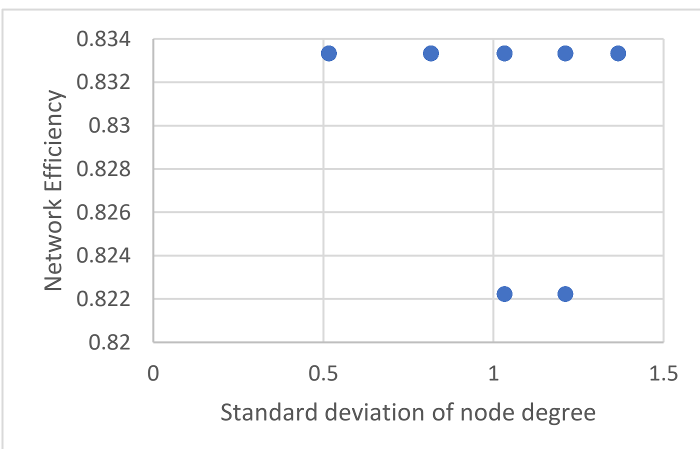


Figure 3.10. The standard deviation of node degree and network efficiency with *n*=6 and *m*=10

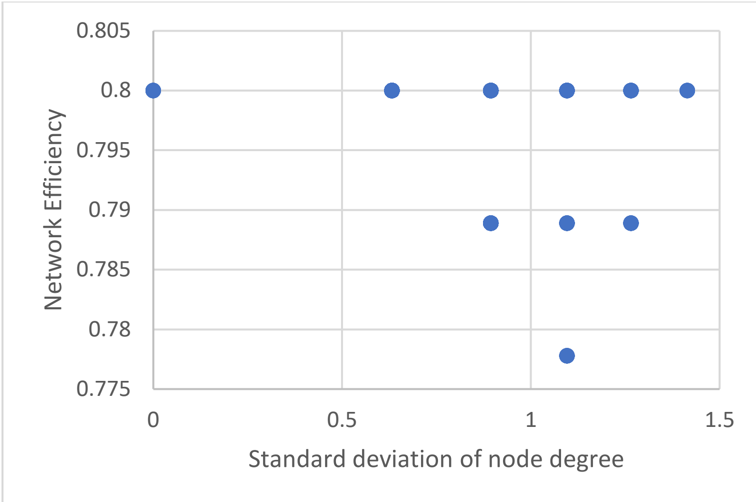


Figure 3.11. The standard deviation of node degree and network efficiency with n=6 and m=9

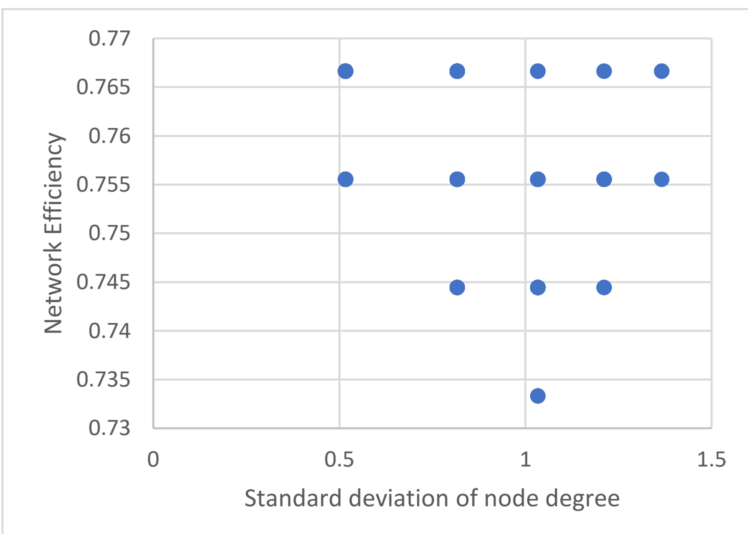


Figure 3.12. The standard deviation of node degree and network efficiency with n=6 and m=8

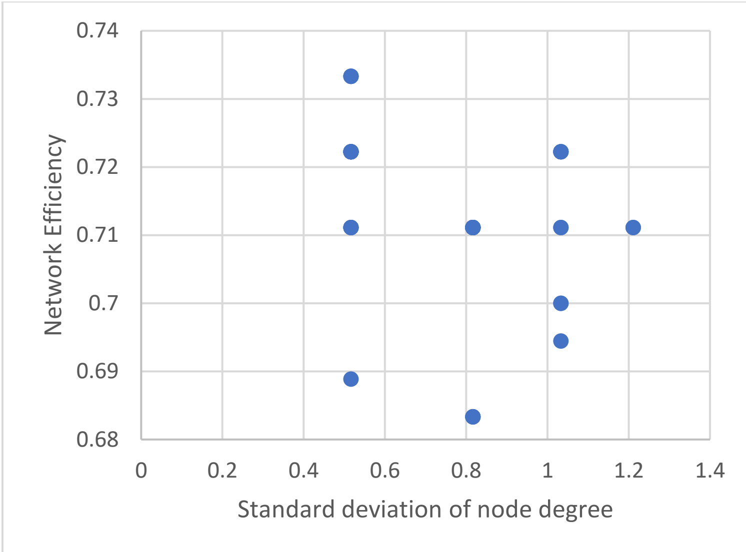


Figure 3.13. The standard deviation of node degree and network efficiency with n=6 and m=7

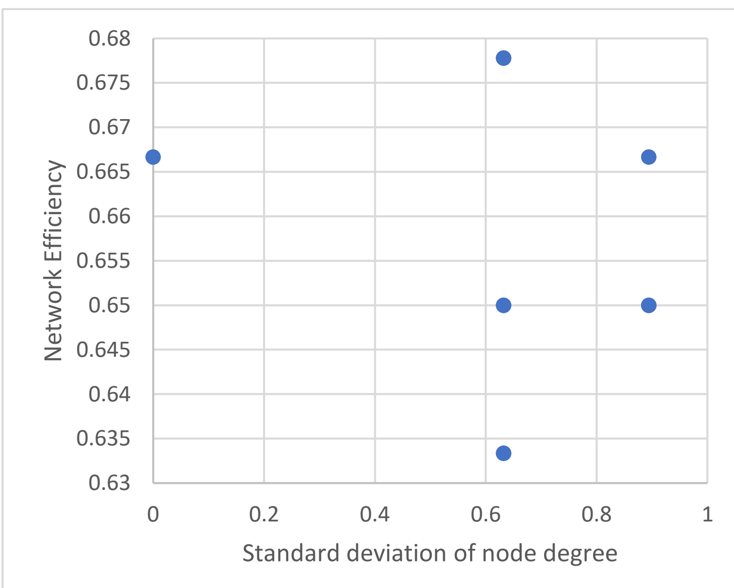


Figure 3.14. The standard deviation of node degree and network efficiency with n=6 and m=6

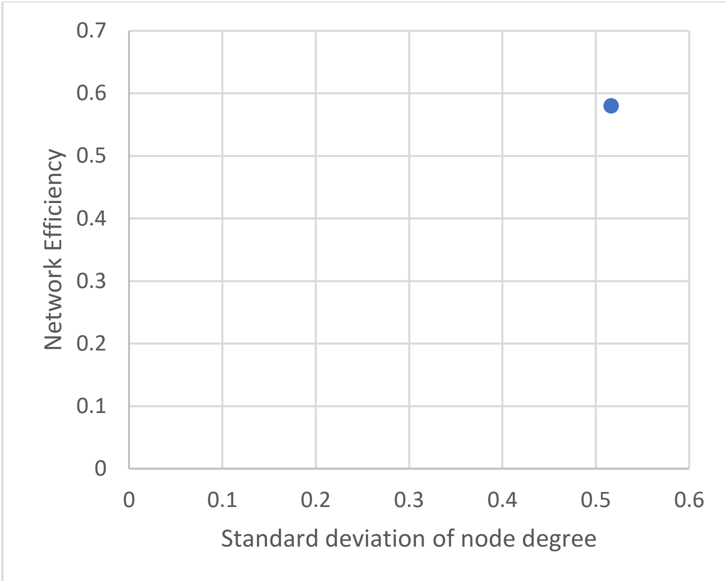


Figure 3.15. The standard deviation of node degree and network efficiency with *n*=6 and *m*=5

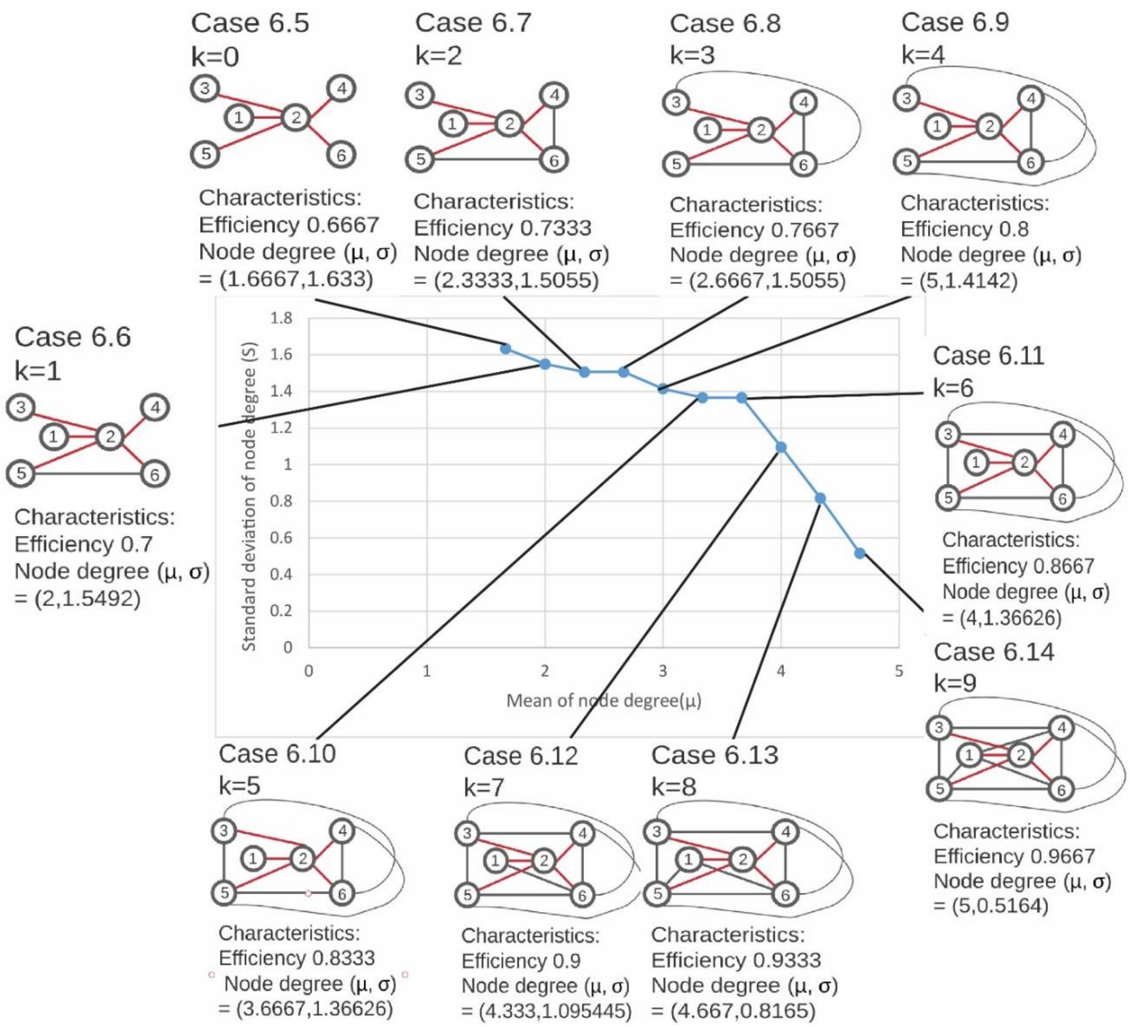


Figure 3.16. The results of two types of analysis for an initial radial network

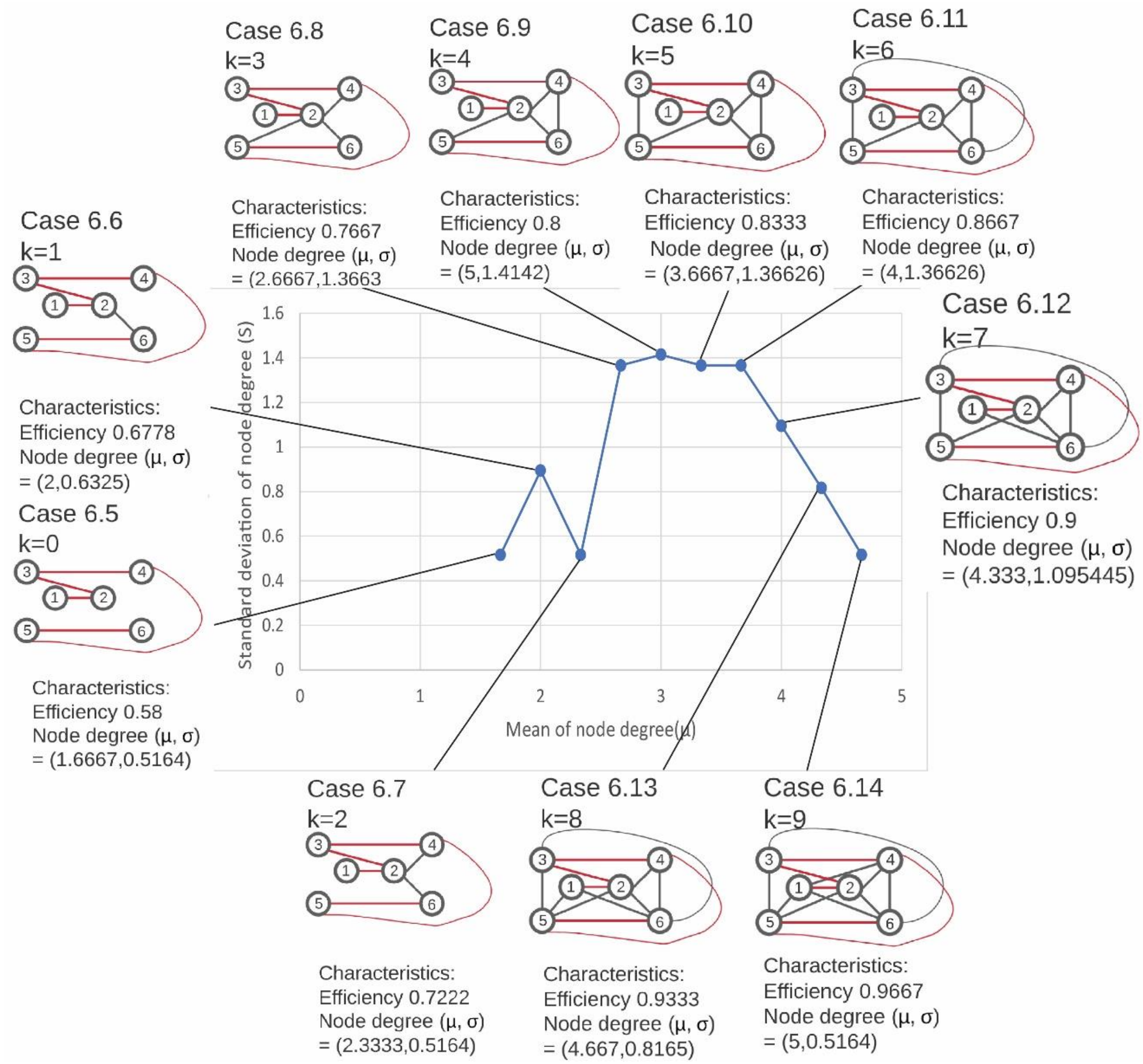


Figure 3.17. Networks with maximum efficiencies and corresponding standard deviations for an initial sequential network

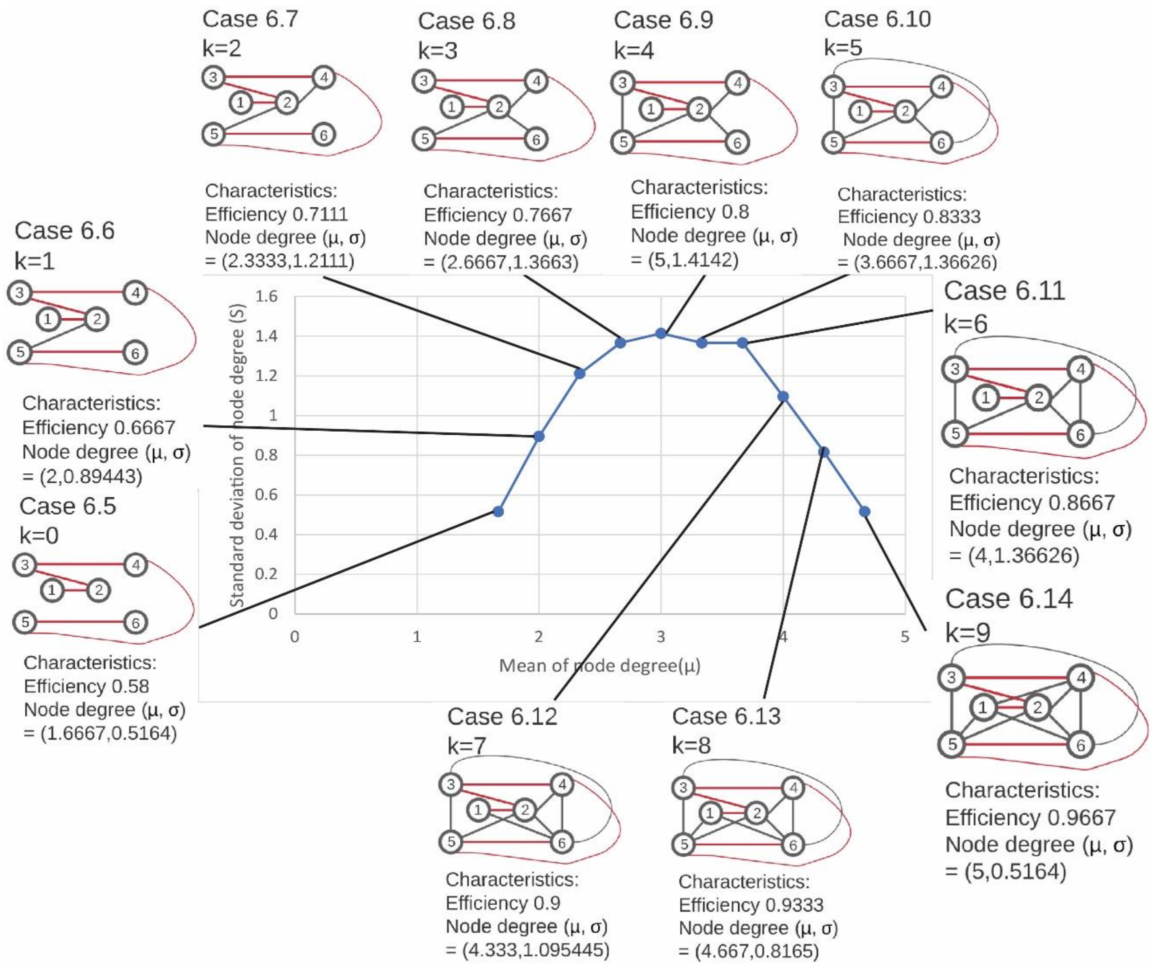


Figure 3.18. Networks with maximum standard deviations and corresponding efficiencies for an initial sequential network

Chapter 4. Case Study: Class I Rail Network

4.1. Background

More than 600 freight railroads operate 140,000-mile U.S. freight rail network and seven Class I railroads provide around 68% of the rail mileage (Association of American Railroads 2020). Moreover, the Class I railroads pass through thousands of cities in North America. The network's topology analysis would estimate a large number of nodes and links if all cities are counted as nodes and the rail between each pair of nodes representing a link. The number of nodes and links would be in the thousands. The calculation for the extensive network is inconvenient and difficult. The topology analysis for the whole network is complex and might not offer the insights necessary to inform decisions.

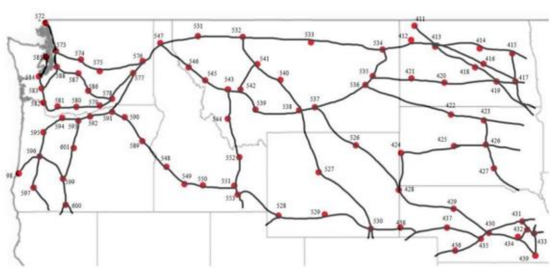
Cao (2020) reduced the Class I rail network's size by identifying and using 638 nodes and 860 links. He reduced the network size based on three criteria: (1) The node density in each state is constant or has little variation. (2) All transfer nodes should be kept. (3) In order to reduce the impact of links' length, longer rails have more nodes than shorter ones. Primarily, he maintained node density and kept all transfer nodes. He also assigned more nodes to long links to account appropriately for links' lengths.

Cao (2020) reduced the size of the Class I rail network; however, the computational efficiency was enhanced. In his network, the long links contain several nodes. Cao (2020) kept these nodes in order to account for links' lengths, although, alternatively, links can have weights to reflect the influence of length, and the further study of the weighted network can estimate the influence of weights on the network topology analysis. This study, however, focuses on the unweighted network. Thus, the network can be further simplified to improve the efficiency of calculation for the purpose of investigating topology optimization.

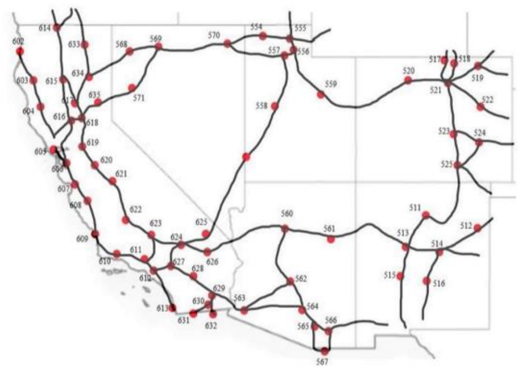
In this study, in order to differentiate among networks cited, the following three distinctions are used: (1) real network (Map store 2017), (2) reduced network (Cao 2020), and (3) simplified network (current work). These distinctions are used to describe the network of different sizes. The real network reflects full Class I freight network with no reduction in size. The reduced network is the network suggested by Cao (2020). The simplified network is the network after the simplification of reduced network. Figure 4.1 presents the real freight network. Figures 4.2 (a) to (f) show the reduced network, and Figure 4.3 describes the simplified network.

For simplification, some transfer nodes that are close to each other are simplified by removing some of them while appropriately connecting of the remaining node. If one remaining node has access to another node passing through the removed node, these two nodes should have a link. In this study, the same node numbering as in the Cao's reduced (2020) network was used to compare analysis results. Figure 4.3 shows the simplified Class I freight railroad network used in this study, including nodes and links. The simplified network contains 82 nodes and 147 links, compared to 638 nodes and 860 links according to Cao's reduced network. The coordinates of nodes are based on Google Maps, and the location of the node is selected based on its proximity to city.

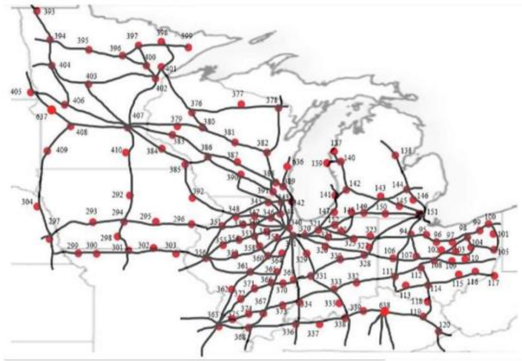
Figure 4.1.Real freight network (Map store 2017)

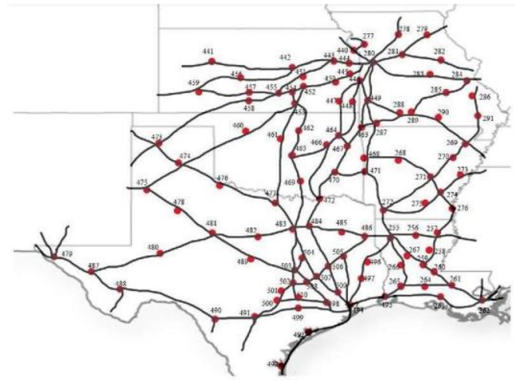


(a) Northwest area



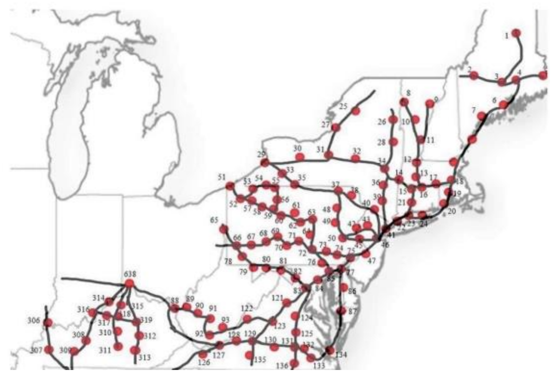
(b) Southwest area





(c) Great Lakes area

(d) Central area



(e) Southeast area

(f) Northeast area

Figure 4.2. Reduced network (Cao 2020)

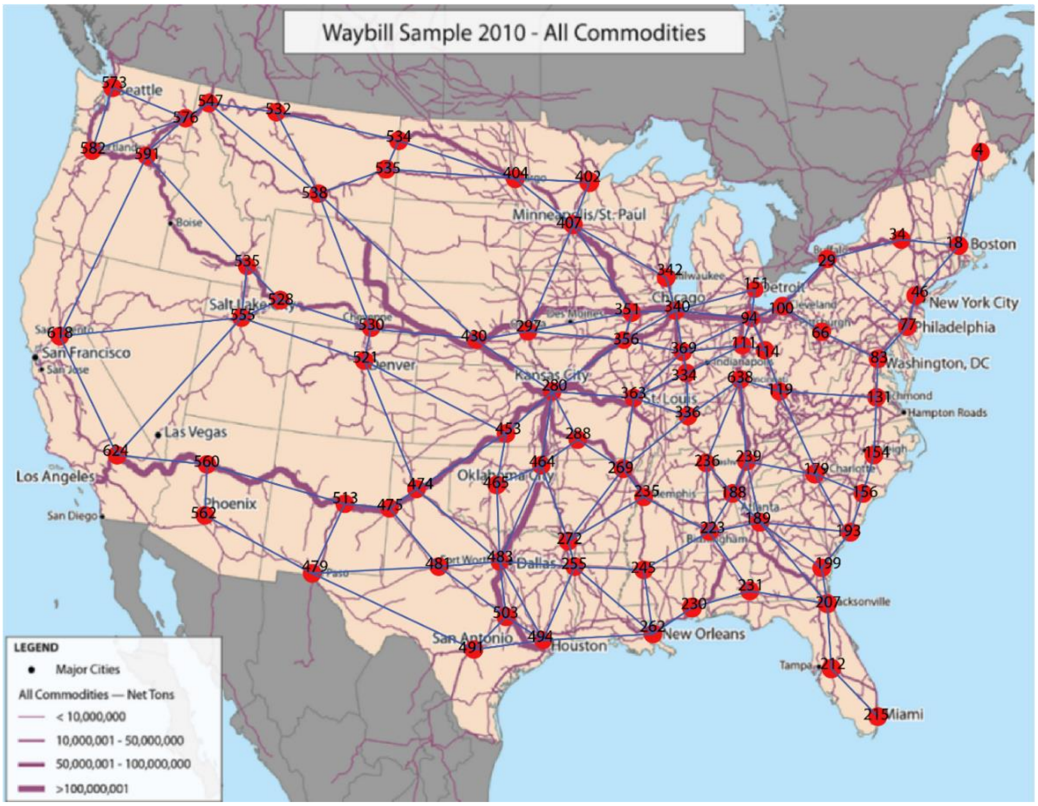


Figure 4.3. Simplified network (Federal Railroad Administration 2020)

4.2. Unweighted Network

4.2.1. Characteristics of Network

Table 4.1 presents the characteristic of the simplified Class I railroad network and reduced network. The simplified network contains 82 nodes and 148 links. The average node degree is 3.61. The characteristic path length is 5.41. The network's diameter, also known as the most extended path length between two nodes, is 13. The clustering coefficient is 0.2405. The network efficiency is 0.2535. Compared to the simplified network, the reduced network contains 638 nodes and 830 links. The average node degree is 2.696. The characteristic path length is 16.726. The diameter of the network is 46. The clustering coefficient and network efficiency are 0.033 and 0.085.

The reduced network has a higher characteristic path length and network diameter, while the simplified network gets higher network efficiency and average node degree. Table 4.1 shows a greater network efficiency for the simplified network compared to the reduced network. The primary reasons are:

- The reduced network contains more nodes than the simplified network. The calculation of network efficiency takes the number of links for all possible pairs of nodes into account. Ideally, with the same matrix of the shortest path length, the more nodes for a network lead to low efficiency.
- The simplified network has a smaller characteristic path length than the reduced network. The simplified network has smaller values in the matrix of shortest path length, and the calculation of network efficiency uses the inverse of these values, which makes the efficiency of the simplified network is higher than the reduced network.

4.2.2. Node-based Network Efficiency Impact

The node-based network efficiency impact is estimated based on the change in network efficiency by disabling one node at a time. When one node is disabled, the analysis removes all links connected to the node, i.e., the number of nodes in the network is not changed. After that, the new network efficiency is compared to the original network efficiency, and the comparison could estimate how critical the node is to the network. Figure 4.4 maps the critical nodes for the simplified network and the reduced network. In the figure, three distinctions are used: (1) the blue nodes, (2) the red nodes, and (3) the yellow nodes. The blue nodes are the critical nodes for the reduced network. The red nodes are critical for the simplified network. The yellow nodes are critical for both the reduced network and the simplified network.

Table 4.2 presents the ten highest node-based network efficiency impact of the simplified network. Based on the table, the node with the highest efficiency impact when disabled is node 280. The efficiency impact of node 280 is 0.404. Kansas City is the nearest city.

Table 4.3 shows the reduced network's results of the node-based network efficiency impact with the nodes in the ten critical nodes of the simplified network. The rank and the efficiency impact of the critical nodes changed after the removal of nodes. The most critical node of the reduced network is node 29, and the second critical node is 280. However, node 280 is more critical than node 29 in the simplified network.

4.2.3. Link-based Network Efficiency Impact

The link-based network efficiency impact is estimated based on the change in network efficiency by removing one link at a time. The removal of one link could reduce network efficiency. The link-based network efficiency impact estimates the influence of the removal of a link on the network efficiency. The critical links have more considerable efficiency impacts than the other nodes. In the analysis, removing a link will not change the number of nodes in the network.

Table 4.4 presents the ten highest link-based network efficiency impact of the unweighted simplified network. The link with the highest efficiency impact is the link between nodes 207 and 212. The efficiency impact is 0.733758. The link is in southeast America. Figure 4.5 maps the most critical links.

Table 4.5 shows the reduced network's results of the link-based network efficiency impact. Because the number of nodes and the connectivity of nodes is changed for the simplified network, this study does not compare the Link-based network efficiency impact between the reduced network and the simplified network.

4.2.4. Recovery Strategies and Resilience Loss

In this study, a network's disruption may cause a node to be disabled, and all links connected to the node are impacted and treated as removed from the network topology. The recovery for the network means recovers all removed links. One link will be recovered at a time, so the recovery strategy is the recovered links' order (Henry and Ramirez-Marquez 2012). The network efficiency changes with each link restored. Thus, the efficiency changes vary for each recovery strategy, and the resilience loss estimates how well the strategy is. The concept of the resilience triangle is first introduced by Bruneau et al. (2003) to assess the loss of resilience for a network with disruption. Then Bocchini and Frangopol (2012) introduced a metric for the analysis of resilience as follows:

$$R_e = \frac{\int_{t_0}^{t_0+t_h} Q(t)dt}{t_h Q_0} \tag{4.1}$$

where R_e is resilience index; $Q(t)$ is the disrupted performance of the system; t_0 is the time when the disruption happens; t_h is the time to recover the disrupted network to the original network.

Zhang et al. (2018) used the resilience index to estimate the Shanghai metro network. They changed Eq. (4.1) in order to put network efficiency into consideration. The resilience index can then be calculated as:

$$R_e = \frac{\int_{t_0}^{t_0+t_h}[E_f(t)]dt}{t_h E_{f0}} \tag{4.2}$$

where $E_f(t)$ is the network efficiency at time *t*; E_{f0} is the original network efficiency.

Based on Eq. (4.2), the sum of the resilience loss and resilience index should be one. The strategy with less loss is better than other strategies because it leads to an effective use of network.

Node 280 has the highest efficiency impact. Thus, node 280 was analyzed for the purpose of recovery needed in the resilience analysis in order to determine the optimal strategy with the lowest loss. Table 4.6 shows the ten recovery strategies with lower resilience loss. The sequence of the optimal strategy is (280, 363)-(280, 430)-(280, 464)-(280, 453)-(280, 356)-(280, 288) and its resilience loss is 0.01716943. In this strategy, the first link recovered is the link between nodes 280 and 363. The second recovered link is the link between nodes 280 and 430. Finally, six links are recovered in this strategy in total.

Table 4.1. Characteristic of unweighted Class I railroad networks: simplified and reduced

No.	Characteristics of network	Computational notes and models	Simplified network	Reduced network (Cao 2020)
1	Number of nodes	The number of cites or waypoints	82	638
2	Number of links	The number of edges between nodes	148	860
3	Average node degree, $\overline{K}$	$\overline{K} = \sum_{i=1}^{n} K_i/\text{n}$, where K_i is the node degree of node *i*; *n* is the number of nodes.	3.61	2.696
4	Characteristic path length, *L*	$L = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij}$, where d_{ij} is the shortest path length between node *i* and node *j*.	5.41	16.726
5	Diameter of network, *D*	The longest path length in link count among all network possible path lengths	13	46
6	Clustering coefficient, $\bar{C}$	$\bar{C} = \frac{1}{n} \sum_{i=1}^{n} C_i$, where $C_i = \frac{2e_{ni}}{K_i(K_i - 1)}$; e_{ni} is the number of links between neighbors of node *i*; K_i is the node degree of node *i*.	0.2405	0.033
7	Network efficiency, E_G	$E_G = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}}$, where d_{ij} is the shortest path length between node *i* and node *j*.	0.2535	0.085

Table 4.2. Node-based network efficiency impact for the simplified network

No.	State	Node numbering	Node degree	Node-based efficiency impact
1	Missouri	280	6	0.064968
2	Washington	576	4	0.052737
3	Florida	207	4	0.051905
4	Texas	483	7	0.046306
5	Montana	538	5	0.044967
6	Alabama	223	6	0.044437
7	Nebraska	430	4	0.044428
8	Ohio	100	4	0.042459
9	Ohio	119	4	0.040472
10	Illinois	340	6	0.039131

Table 4.3. Node-based network efficiency impact for the reduced network

No.	State	Node numbering	Node degree	Node-based efficiency impact
1	New York	29	5	0.0295
2	Missouri	280	7	0.0285
3	Illinois	341	6	0.0201
4	Alabama	222	5	0.018
5	Florida	207	6	0.0163
6	Tennessee	235	5	0.0151
7	Minnesota	407	8	0.0142
8	Michigan	151	6	0.0122
9	Illinois	363	7	0.012
10	New York	31	3	0.0119

Table 4.4. Link-based network efficiency impact for the simplified network

No.	Area	Starting node	Ending node	Link-based efficiency impact
1	South East (SE)	207	212	0.030687
2	North East (NE)	4	18	0.014874
3	North East (NE)	100	29	0.013738
4	Central South (CS)	280	430	0.012691
5	North west (NW)	547	576	0.010983
6	Great Lakes (GL)	336	638	0.010982
7	Great Lakes (GL)	114	119	0.010153
8	Central South (CS)	453	280	0.010007
9	North East (NE)	131	83	0.009685
10	South East (SE)	245	223	0.009462

Table 4.5. Link-based network efficiency impact for the reduced network

No.	Area	Starting node	Ending node	Link-based efficiency impact
1	South East (SE)	181	244	0.0115
2	South East (SE)	170	242	0.0113
3	North East (NE)	31	32	0.0109
4	North East (NE)	7	18	0.0099
5	North East (NE)	35	37	0.0097
6	North East (NE)	13	16	0.0094
7	North East (NE)	124	125	0.0089
8	Great Lakes (GL)	282	363	0.0088
9	North East (NE)	6	7	0.0081
10	North East (NE)	37	38	0.0074

Table 4.6. Recovery strategies and resilience loss for node 280

Ranking	Recovery sequence	Resilience loss
1	[(280, 363)-(280, 430)-(280, 464)-(280, 453)-(280, 356)-(280, 288)]	0.01716943
2	[(280, 363)-(280, 453)-(280, 430)-(280, 464)-(280, 356)-(280, 288)]	0.01732198
3	[(280, 430)-(280, 363)-(280, 464)-(280, 453)-(280, 356)-(280, 288)]	0.01737279
4	[(280, 363)-(280, 430)-(280, 453)-(280, 464)-(280, 356)-(280, 288)]	0.01737463
5	[(280, 464)-(280, 430)-(280, 363)-(280, 453)-(280, 356)-(280, 288)]	0.01750242
6	[(280, 430)-(280, 363)-(280, 453)-(280, 464)-(280, 356)-(280, 288)]	0.01757799
7	[(280, 430)-(280, 464)-(280, 363)-(280, 453)-(280, 356)-(280, 288)]	0.01758207
8	[(280, 453)-(280, 363)-(280, 430)-(280, 464)-(280, 356)-(280, 288)]	0.01764332
9	[(280, 363)-(280, 430)-(280, 464)-(280, 356)-(280, 453)-(280, 288)]	0.01771981
10	[(280, 363)-(280, 430)-(280, 464)-(280, 453)-(280, 288)-(280, 356)]	0.01787827

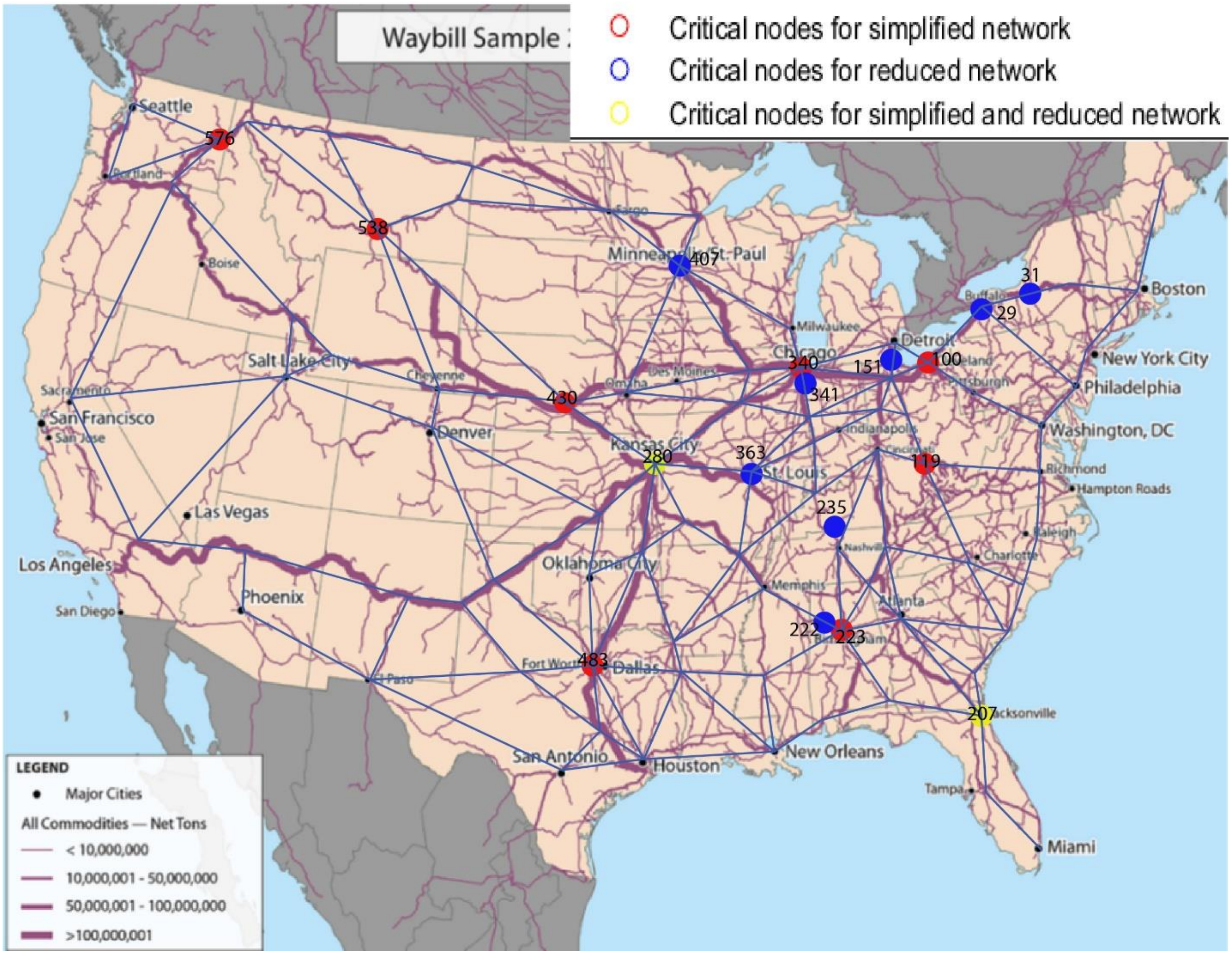


Figure 4.4. Critical nodes for the simplified and reduced network

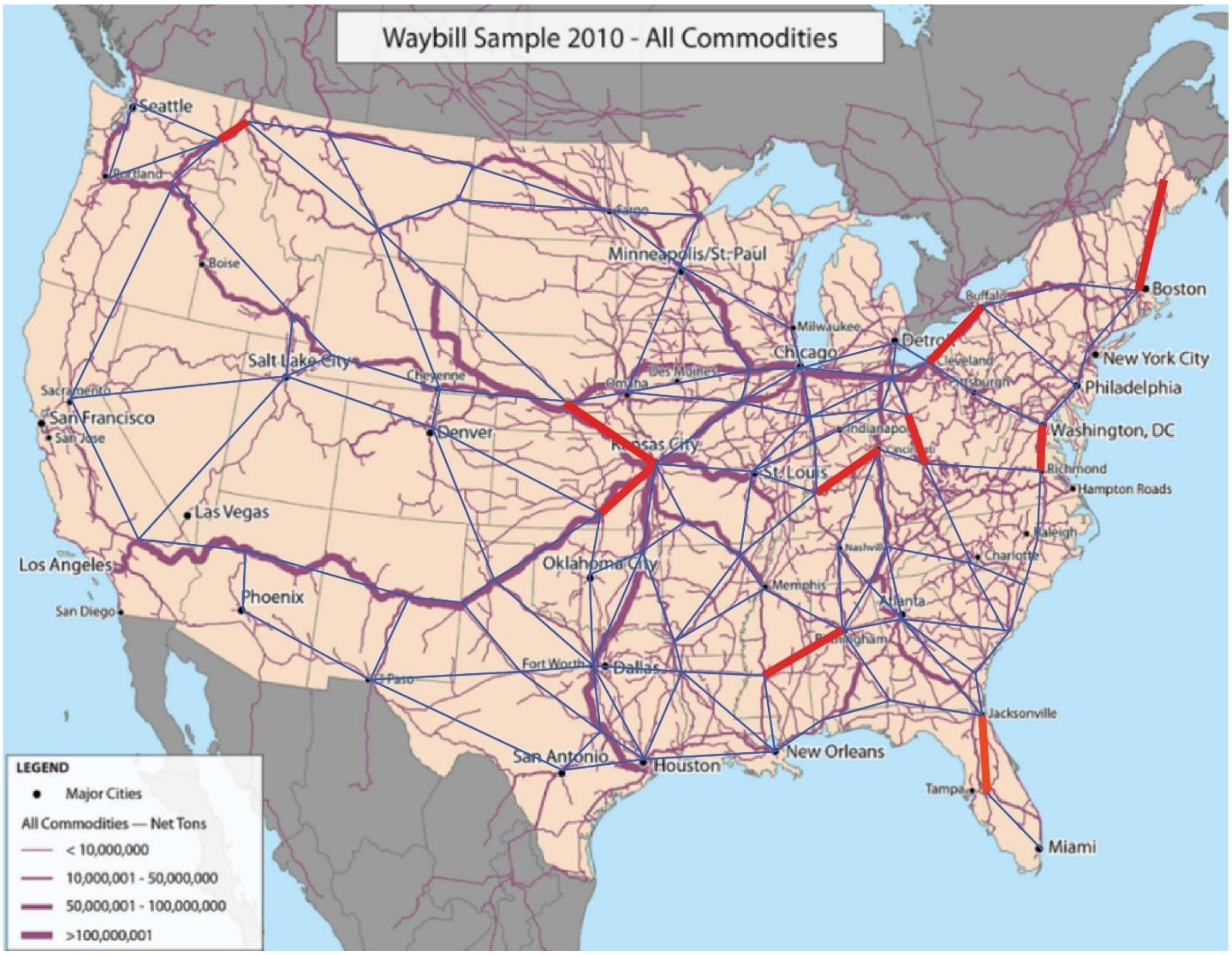


Figure 4.5. Ten critical links with higher link-based efficiency impact of simplified network

4.3. Network Optimization

Researchers have developed several optimization and solution methods. Different optimization types vary in optimization objectives that determine the optimization method used in the analysis. For example, optimization methods and their objectives can be:

- Ant colony optimization method to solve the traveling salesman problem (Dorigo and Gambardella 1997).
- Harmony search to find better solutions with fewer iterations (Zong et al. 2001).
- Simulated Annealing Algorithm (SAA) to reduce the operation cost for a network (Feng et al. 2018)

Dorigo and Gambardella (1997) used the accumulated information to find the shortest route passing through each network node. Zong et al. (2001) has used harmony search to solve the traveling salesman problem. They set an initial network and upgraded it sequentially to get the shortest route that connects all nodes.

In this study, the objective is to maximize network efficiency by adding links to a given initial network. The methods by Dorigo and Gambardella (1997), Zong et al. (2001), and Feng et al. (2018) are not suitable for this purpose. Thus, a brute force method is proposed to assess the optimization's feasibility based on the proposed use of the standard deviation of node degree (*S*). The analysis compares all cases to add links to the network to find a case that can maximize network efficiency.

For the simplified network, the fully connected network, i.e., each node is connected directly to all other nodes, contains 3,321 links. Thus, 3,173 links can be added to the network to increase the network efficiency with initially 148 links according to the simplified network. In order to optimize the growth of the simplified network, the link addition method would be used. The consideration of all the potential additional links requires a very long execution time. Thus, this study estimates the extended networks by k = 1 to 79 additional links added incrementally one at a time.

Max S (standard deviation of node degree) and max E (network efficiency) analysis are used in this case study. The results of the two types of analysis would be used to verify the feasibility of max S analysis to optimize the network in order to enhance computational efficiency. Figures 4.6 and 4.7 present the network efficiency and standard deviation of node degree for the max S and max E analysis using k = 1 to 79. Figure 4.6 shows that the two analyses have a slight difference with k less than 20, while the difference reduces later. The normalized efficiency difference (d) between the efficiencies of two types of analysis is calculated in this study as follow:

$$d = (E_{maxE} - E_{maxS})/E_{maxE} \times 100\% \tag{4.3}$$

where d is the normalized efficiency difference between two types of analysis; E_{maxE} is the network efficiency of max E analysis; E_{maxS} is the max S analysis's network efficiency.

The normalized efficiency difference between the two types of analysis efficiencies is used to estimate the feasibility of using mas S analysis to optimize the network instead of max E analysis. Figure 4.8 describes the normalized efficiency difference between max S and max E analyses in network efficiency. When k is less than 20, the normalized efficiency difference fluctuates between 0 and 1.4%. When k is larger than 20, the fluctuations tend to flatten out. Generally, the normalized efficiency difference decreases with the increase of k and is finally less than 0.1%. The slight normalized efficiency difference shows the feasibility of the standard deviation of node degree to optimize the network.

The extended network with $k = 1$ is used as an example to take a detailed understanding of max S and max E analysis. Figure 4.9 presents the extended network when $k = 1$. The red link between nodes 189 and 538 could maximize the network efficiency, which is 0.26592 and given by max E analysis, while the blue link between nodes 94 and 483 could maximize the standard deviation of node degree, and the new network efficiency is 0.26412. The normalized efficiency difference between the two efficiencies is slight. However, the analysis is about the unweighted network, so the links shown in Figure 4.14 are practically long. Thus, in order to get a more practical solution, the filtration of links is needed. The filtration is based on the length of links computed by the coordinates of nodes. Five hundred shortest links are used for the practical solution.

The practical solution used five hundred links to extend the network, and the extended networks with $k = 1$ to 79 are analyzed. Figures 4.10 and 4.11 present the efficiency comparison for practical solutions. The two figures show that max S can be used with fewer links and the results are also feasible. Figure 4.12 describes the link's location added by two types of analysis, which is different from the results without filtration of links. The blue link between nodes 340 and 223 is added based on the max S analysis, and the red link between nodes 100 and 194 is added based on the max E analysis. The efficiencies for max E and max S analysis are 0.26294 and 0.26207. This study also estimates 200, 150, and 100 shortest links for practical solutions. The results are similar to the solution with 500 shortest links. The normalized efficiency difference between max *S* and max *E* analysis decreases with more links added to the network.

The solution does not mean a new railroad. It could also be the existing railroads that are not included in the Class I rail network. Thus, the decision-makers can improve the existing railroads into the Class I rail network to optimize it.

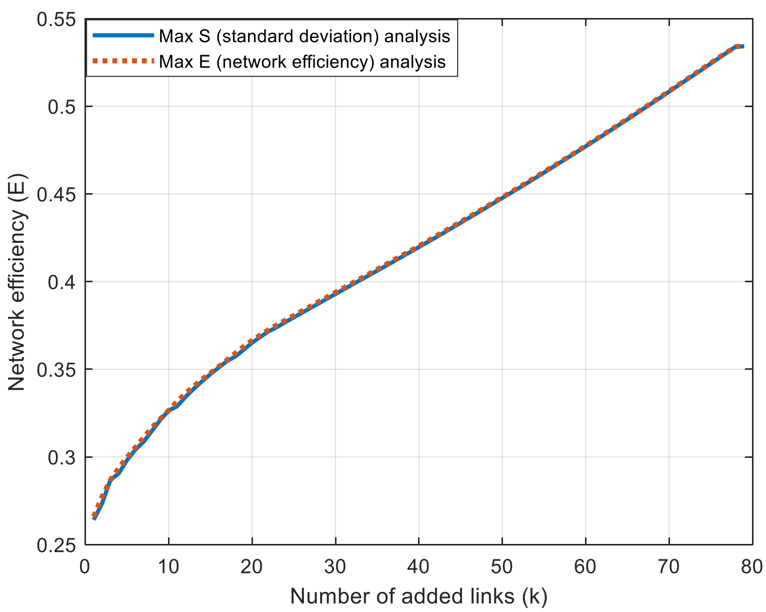


Figure 4.6.Network efficiency using max *S* and max *E* analysis for the additional links

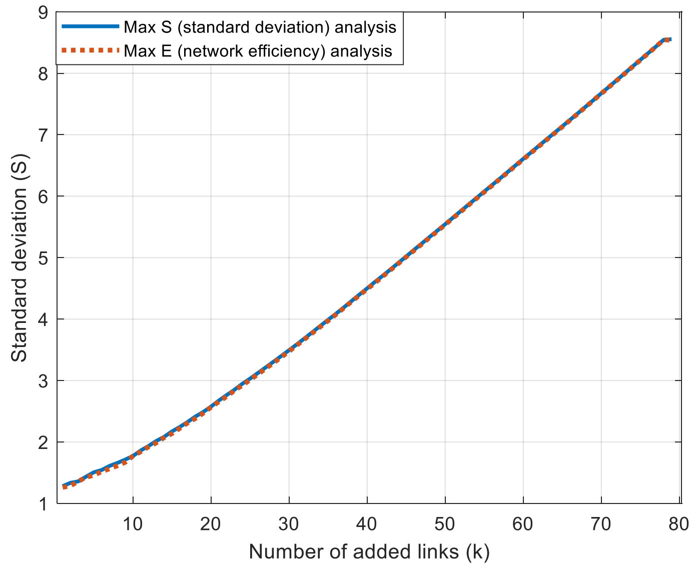


Figure 4.7. Standard deviation of node degree using max S and max E analysis for the additional links

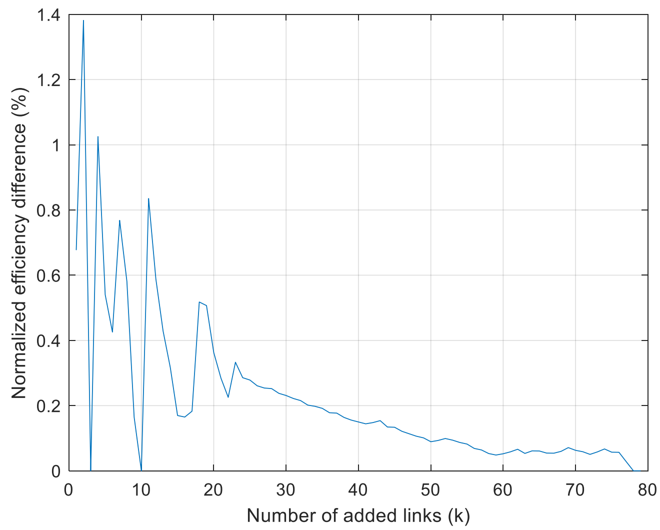


Figure 4.8. Normalized efficiency difference between max S and max E analysis in efficiency

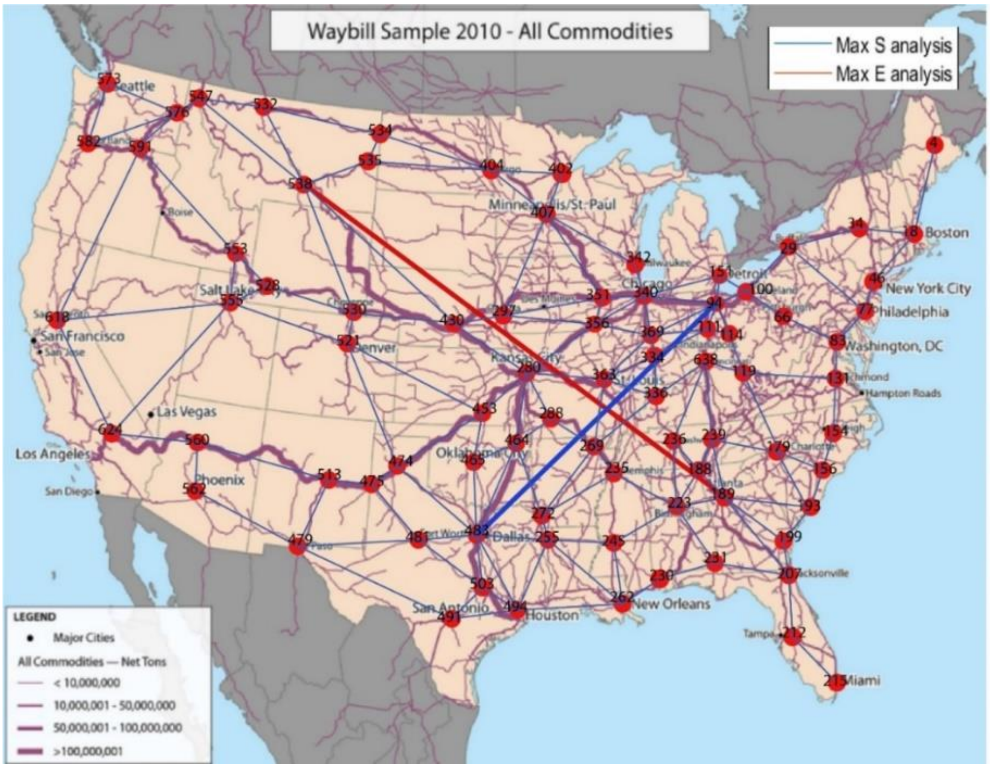


Figure 4.9. Optimal network of two types of analysis when adding one link

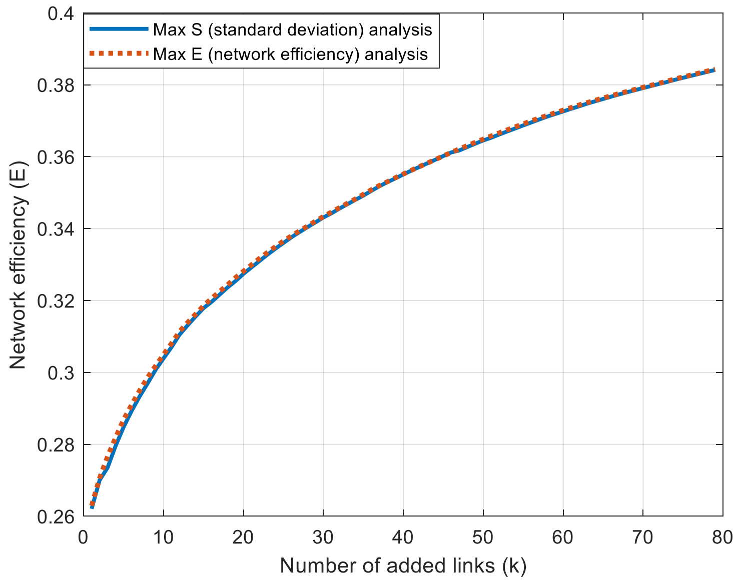


Figure 4.10. Network efficiency of practical solution using max *S* and max *E* analysis for the additional links

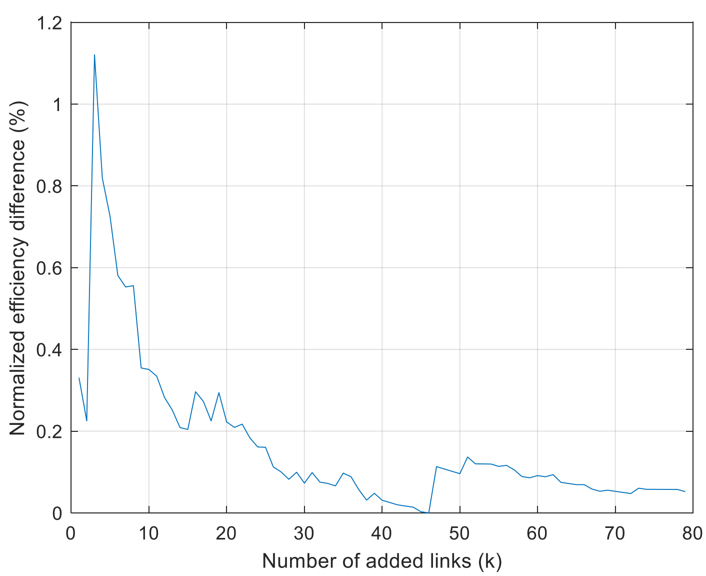


Figure 4.11. Normalized efficiency difference between max *S* and max *E* analysis in efficiency for the practical solution

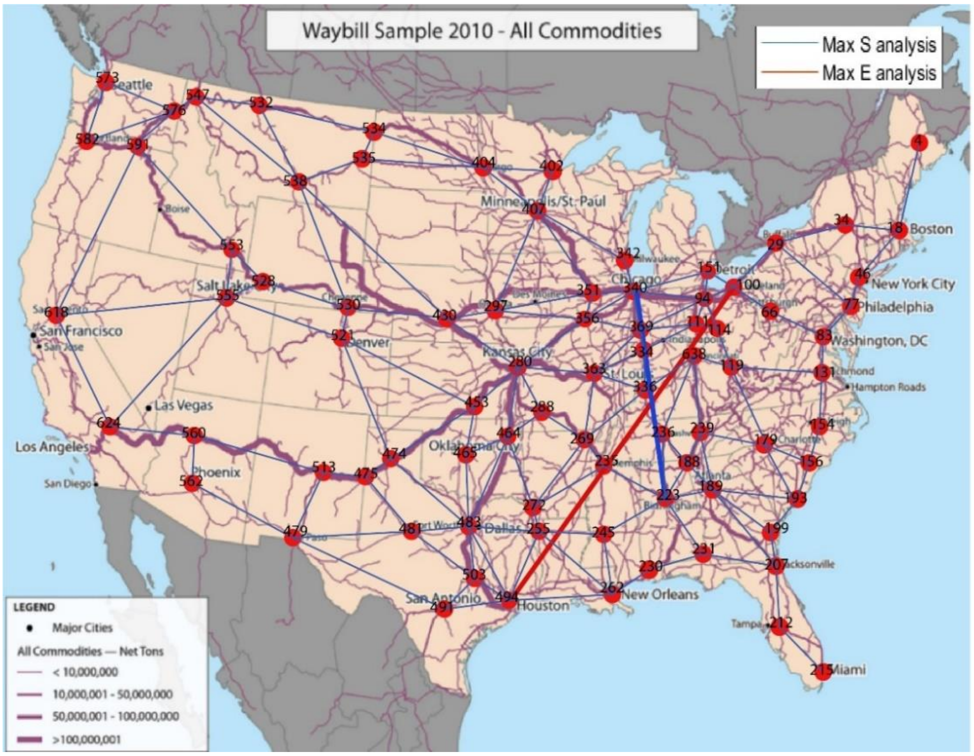


Figure 4.12. Optimal network of two types of analysis when adding one link for the practical solution

Chapter 5. Conclusion and Contribution

In the United States, rail networks play a critical role in freight and passengers' transportation, with the highest proportion among all transportation modes (Bureau of Transportation Statistics 2010). Optimization of railroads can have significant economic and other benefits. Although rails are well established over many decades of development and enhancement, the ever-changing market needs and societal demands always require an evolving railroad network to meet changing needs and demands.

Since the railroad network is mature, enhancement can be identified and undertaken by reclassing, repurposing, and acquiring abandoned or new railroads. The method developed in this thesis provides planners a way to assess and optimize a particular rail network for enhancing network efficiency. Topological properties depicting the relationship among nodes (including cities and waypoints) and links (as railroads) are used to assess the network. Node degree, characteristic length, and network efficiency describe the attributes of the network. Efficiency impacts resulting in disabling nodes or links point out the critical ones, respectively. Recovery strategies for a particular node or link can be based on reducing the resilience loss during the recovery period.

Network efficiency offers a basis for quantifying other network attributes and is a primary focus in this thesis for network enhancement. The expansion of a network can be achieved by adding new links or revising the existing links by reclassing or acquisition, or repurposing them to enhance network efficiency. The notion of adding or removing a link should be understood in this context. Different potential locations of new links are identified to optimize network efficiency. However, when the size of the network is large, the method of using network efficiency to find the optimal solution can be computationally inefficient. Thus, this study introduces the standard deviation of the node degree as the criterion for optimization in order to enhance the computational efficiency.

Simplified networks with six nodes were probed to illustrate that the network with a higher standard deviation of node degree generally has superior network efficiency. The initial or starting networks for expansion are separated into types that are minimally connected: (1) a sequential network and (2) a radial network. One link at a time is added to the initial network until it is fully connected, i.e., all pairs of nodes are connected, to examine the devised method to optimize the network based on its efficiency. For an initial radial network, the optimal networks provided by the two optimization methods corresponding to maximizing efficiency and the standard deviation of the node degree have the same maximum network efficiency given the same number of links. For an initial sequential network, the optimal solutions of the two methods are the same in most cases. The analysis offers insights that can be helpful to rail analysts and managers.

In this thesis, a case study of the Class I rail network is introduced to examine the two optimization methods. A simplified network consisting of 82 nodes and 148 links is used for the analysis. The simplified network comes from the reduced network suggested by Cao (2020) by removing redundant nodes. This study starts with placing all potential links, that is, 3173 links, into consideration. The results show that the method based on the standard deviation of node degree can be used to find an optimal solution. This method makes the computation more efficient with an appropriate approximation with a normalized difference of less than 0.2% based on adding 40 links compared to the direct use of network efficiency. Additionally, the difference decreases for cases with more links added to the network. In the analysis, considering 3173 potential links is computationally taxing because the long links are not practical to be added to the network for optimization. In order to obtain a suitable solution for planners or analysts to reclass, acquire, or repurpose railroads, five hundred shortest links were suggested as the potential links for the enhancement of network efficiency. In conclusion, the solutions based on this method increase network efficiency and offer a basis for reclassing, acquiring, and repurposing railroads.

The primary contribution in the thesis is the proposed method to optimize the network reliably with computational efficiency based on the standard deviation of the node degree. This method reduces the optimization's complexity compared to the direct use of network efficiency. The normalized difference of solutions between the two optimization methods is minimal. The standard deviation of the node degree method offers a practical basis to obtain the network expansion's optimal solutions. These solutions become the starting point for economic and legal consideration for reclassing, acquiring, and repurposing railroads or building new railroads.

A primary area for future pursuits based on the thesis's limitation of using an unweighted network is to add weights to links and nodes, such as reflect the distance between each node, track class, etc. Future work can focus on the optimization for weighted networks.

Distance, the volume of goods, passenger flow, and so on can be analyzed as weights could enhance the solution set obtained from the optimization.

References

Abandonedrails 2021. “Abandoned rails,” <https:// http://www.abandonedrails.com/> (Accessed 15 Feb 2021)

American rails 2021. “ Class I Railroads (USA),” <https://www.american-rails.com/class.html> (Accessed 05 Jan 2021)

Association of American Railroads 2013. “Class I Railroad Statistics,” <https://web.archive.org/web/20131103072236/https://www.aar.org/StatisticsAndPublications/Documents/AAR-Stats-2013-04-17.pdf > (Accessed 05 Jan 2021)

Association of American Railroads 2019. "The Economic Impact of America's Freight Railroads," <https://web.archive.org/web/20131103071634/https://www.aar.org/StatisticsAndPublications/Documents/AAR-Stats-2013-02-07.pdf.> (Accessed 05 Jan 2021)

Association of American Railroads 2020. “Overview of America’s Freight Railroads,” <https://www.aar.org/wp-content/uploads/2020/08/AAR-Railroad-101-Freight-Railroads-Fact-Sheet.pdf.> (accessed 25 Dec 2021).

Association of American Railroads 2020a. “Chronology of America’s Freight Railroads,” < https://www.aar.org/wp-content/uploads/2020/07/AAR-Chronology-Americas-Freight-Railroads-Fact-Sheet.pdf > (accessed 16 Feb 2021)

Bai, Y., Huang, N., and Sun, L. 2018. “A key heterogeneous structure of fractal networks based on inverse renormalization scheme,” Physica a: Statistical Mechanics and Its Applications, 499, 67–74

Brooks, P. 2018. “Rail Freight Shipping: Class I Railroads in America,” <https://arcb.com/blog/rail-freight-shipping-class-i-railroads-in-america> (Accessed 05 Jan 2021)

Bureau of Transportation Statistics 2018. “National Transportation Statistics”, <https://www.bts.gov/topics/national-transportation-statistics>, (accessed 03 Mar 2021)

Bocchini, P. and Frangopol, D.M. 2012. “Restoration of bridge networks after an earthquake: multicriteria intervention optimization,” Earthq. Spectra 28, 427–455.

Cao, S. 2020. “Topological Analysis for Vulnerability Enhancement and Recovery Strategy after Disruptions of Rail Networks in the United States,” M.S. Thesis. University of Maryland, College Park, MD.

Chakraborty, A., and Ikeda, Y. 2020. “Testing ‘efficient supply chain propositions’ using topological characterization of the global supply chain network,” PLoS ONE, 15(10), 1–18.

Crucitti, P., Latora, V., Marchiori, M., and Rapisarda, A. 2004. “Error and attack tolerance of complex networks,” Physica a: Statistical Mechanics and Its Applications, 340(1), 388–394.

Delaware Valley Regional Planning Commission 1991. “Potential Reuse of Inactive Rail Lines,” < https://www.dvrpc.org/Reports/91040.pdf > (accessed 15 Feb 2021).

Di Ruocco, G., Sicignano, E., Fiore, P., and D’Andria, E. 2017. “Sustainable reuse of disused railway,” Procedia Engineering, 180, 1643–1652.

Dorigo, M., and Gambardella, L. M. 1997. "Ant colonies for the travelling salesman problem," Biosystems, 43(2), 73–81.

Fan, W., and Machemehl, R. B. 2006. "Optimal transit route network design problem with variable transit demand: genetic algorithm approach," Journal of Transportation Engineering, 132(1), 40–51.

Federal Railroad Administration. 2010. "National Rail Plan Progress Report," https://railroads.dot.gov/elibrary/national-rail-plan-progress-report (accessed 02 Fed 2021).

Federal Railroad Administration 2020."Freight Rail Overview," < https://railroads.dot.gov/rail-network-development/freight-rail/freight-rail-overview-0> (accessed 02 Fed 2021).

Feng, X., Ruan, Z., Zhu, X., and Zhang, L. 2018. "Multi-objective transport network design with a reversible simulated annealing algorithm," Advances in Production Engineering and Management, 13(4), 479–491.

Garrison, W.L. and Marble, D.F. 1962. The structure of transportation networks, No. TR-62-11, Northwestern Univ Evanston Il.

Henry, D., and Ramirez-Marquez, J.E. 2012. "Generic metrics and quantitative approaches for system resilience as a function of time." Reliab. Eng. Syst. Saf. 99, 114–122.

Jiang, J., Chen, J., and Wang, C. 2020. "Multi-granularity hybrid parallel network simplex algorithm for minimum-cost flow problems," The Journal of

Supercomputing : An International Journal of High-Performance Computer Design, Analysis, and Use, 76(12), 9800–9826.

Johnson, C. A., Reilly, A. C., Flage, R., and Guikema, S. D. 2020. “Characterizing the robustness of power-law networks that experience spatially-correlated failures,” Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 1748006–2097447.

Latora, V., and Marchiori, M. 2001. “Efficient behavior of small-world networks,” Physical Review Letters, 87(19), 198701–198701.

Majumder, S., Saha, B., Anand, P., Kar, S., and Pal, T. 2018. “Uncertainty based genetic algorithm with varying population for random fuzzy maximum flow problem,” Expert Systems, 35(4).

Map store 2017. US Rail Network Map (AI), <https://store.mapsofworld.com/digital-maps/us-maps-1-2/usa-rail-network-digital-map.> (accessed 25 Dec 2020).

Musso, A., and Vuchic, V. 1988. "Characteristics of metro networks and methodology for their evaluation," TRANSPORTATION RESEARCH RECORD.

Murray-Tuite, P. M., and Mahmassani, H. S. 2004. “Methodology for determining vulnerable links in a transportation network,” Transportation Research Record: Journal of the Transportation Research Board, 1882(1), 88–96.

Nayeem, M. A., Rahman, M. K., and Rahman, M. S. 2017. “Transit network design by genetic algorithm with elitism,” Transportation Research Part C: Emerging Technologies, 46, 30–45.

Newman, M. 2018. Network, second edition, Oxford University Press.

Nikolić, M., and Teodorović, D. 2013. Transit network design by bee colony optimization. Expert Systems with Applications, 40(15), 5945–5955.

Saadat, Y., Ayyub, B. M., Zhang, Y., Zhang, D., and Huang, H. 2019. “Resilience of metrorail networks: quantification with washington, dc as a case study,” Asce-Asme J Risk and Uncert in Engrg Sys Part B Mech Engrg, 5(4).

Saadat, Y., Ayyub, B., Zhang, Y., Zhang, D., and Huang, H. 2020. "Resilience-Based Strategies for Topology Enhancement and Recovery of Metrorail Transit Networks," ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 6(2), 04020017.

Seman, K., Puspita, F. M., Taib, B. M., and Shafii, Z. 2012. “An Improved Optimization Model of Internet Charging Scheme in Multi Service Networks,” Telkomnika,10(3), 592–598.

Statista 2020. “Total operating revenues of U.S. Class I rail from 1990 to 2019,” <https://www.statista.com/statistics/187655/operating-revenues-of-us-class-i-rail-since-1990/> (Accessed 05 Jan 2021)

Statista 2020a. “North American railroads by revenue 2019” <https://www.statista.com/statistics/271613/leading-north-american-railroad-companies-based-on-revenue/> (Accessed 05 Jan 2021)

Yuan, Z., Zhao, C., Di, Z., Wang, W. X., and Lai, Y. C. 2013. “Exact controllability of complex networks,” Nature Communications, 4, 2447–2447.

Zong, W. G., Joong, H. K., and Loganathan, G. V. 2001. “A new heuristic optimization algorithm: harmony search,” Simulation. 76(2), 60–68.

Zhang, D.-ming, Du, F., Huang, H., Zhang, F., Ayyub, B. M., and Beer, M. 2018. Resiliency assessment of urban rail transit networks: shanghai metro as an example. SafetyScience, 106, 230–243.

Zhang, C., Dai, S., and Xia, H. 2020. Reuse of abandoned railways leads to urban regeneration: a tale from a rust track to a green corridor in zhangjiakou. Urban Rail Transit, 6(2), 104–115.

Zhao, L., Cheng, Z., Li, H., and Hu, Q. 2019. "Evolution of the China Railway Express Consolidation Network and Optimization of Consolidation Routes," Journal of Advanced Transportation, 1–17.