8-2018

Essays in network theory applications for transportation planning

Jeremy David Auerbach
University of Tennessee, jauerbac@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Jeremy David Auerbach entitled "Essays in network theory applications for transportation planning." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Geography.

Hyun Kim, Major Professor

We have read this dissertation and recommend its acceptance:

Eugene C. Fitzhugh, Yingjie Hu, Shih-Lung Shaw

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Essays in network theory applications for transportation planning

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Jeremy David Auerbach

August 2018

Acknowledgments

I would like to thank my advisor Hyun Kim (Associate Professor of Geography at the University of Tennessee) for his patience and mentorship and my committee members, Eugene Fitzghugh, Shih-Lung Shaw, and Yingjie Hu, for their valuable input. I would like to thank Alex Zendel (GIS Analyst at the Knoxville-Knox County Metropolitan Planning Commission) for compiling the school network data used for the case study in the second chapter “Optimization of network connectivity: An evaluation of heuristics applied to random networks and a transportation case study”. For the third chapter “Small changes in street connectivity result in big gains for student walking” I would like to thank Ellen Zaviska (Senior Transportation Planner at the Knoxville Regional Transportation Planning Organization) for comments and advice, Alex Zendel (GIS Analyst at the Knoxville-Knox County Metropolitan Planning Commission) for compiling the school network data, and Rickey Grubb (Director of Enrollment and Transportation at Knox County Schools) for relevant information about school busing.

Abstract

Throughout the dissertation, network methods are developed to address pressing issues in transportation science and geography. These methods are applied to case studies to highlight their use for urban planners and social scientists working in transportation, mobility, housing, and health. The first chapter introduces novel network robustness measures for multi-line networks. This work will provide transportation planners a new tool for evaluating the resilience of transportation systems with multiple lines to failures. The second chapter explores optimizing network connectivity to maximize the number of nodes within a given distance to a focal node while minimizing the number and length of additional connections. These methods can be used to identify optimal thoroughfare design around important facilities, such as schools. The third chapter utilizes the network optimization heuristics presented in Chapter 2 to identify the impact of thoroughfare connectivity on student active commuting. Housing developers can incorporate these findings when planning new residential developments around schools. The dissertation concludes with future directions in this research domain.

Table of Contents

List of Tables

List of Figures

List of Algorithms

Introduction

Transportation studies and geography have benefited greatly from network science (Larson and Odoni, 1981, Magnanti and Wong, 1984, Derrible and Kennedy, 2010a) but there are still many questions and issues in the field which need to be addressed. Measuring the resilience of systems that have multiple connections between nodes, such as rail systems with several lines between stations, and optimizing the thoroughfare connectivity around important facilities, such as greenways around schools, are examples of problems still not fully explored. Utilizing network theory approaches to address these questions can help urban planners and social scientists make more informed policy recommendations for improving healthcare, infrastructure, and security.

The dissertation is composed of several chapters, each exploring a different aspect of transportation research and network theory with the emphasis on developing new methodology and written in separate manuscript formats. Each chapter addresses a separate research question under the umbrella of transportation science and network theory. Definitions of terms, literature reviews, methodologies, results, and conclusions are present in every chapter.

Robustness, or the resilience of a system to failures and attacks, is an increasingly important characteristic of networks. Even though there have been proposed several measures of network robustness, there is a lack of analysis of these techniques for networks with multiple connections between the same nodes (multi-line networks). In chapter 1, to address the issue of measuring vulnerability of multi-line network systems, a set of novel global indexes and a local index to measure robustness for these multi-line networks are introduced. These new indexes are designed to uncover the potential vulnerabilities of network components (i.e. hubs, terminals and lines) to possible malfunctions which may not be well identified with traditional connectivity matrix-based methods. Results of these measures are compared with traditional network connectivity and robustness indexes for

simulated networks that vary topologically and are each composed of multiple lines that overlap in connectivity. A set of representative rail systems in the United States are used as a case study to provide insights for planners working to enhance preparedness for network disruptions, and highlight some future directions to further research in this realm.

Network optimization has generally been focused on solving network flow problems, but recently there have been investigations into optimizing network characteristics. Chapter 2 introduces optimizing network connectivity to maximize the number of nodes within a given distance to a focal node and then minimizing the number and length of additional connections has not been as thoroughly explored, yet is important in several domains including transportation planning, telecommunications networks, and geospatial location. We compare several heuristics to explore this network connectivity optimization problem with the use of random networks, including the introduction of two planar random networks that are useful for spatial network simulation research, and a real-world case study from urban planning and public health. We observe significant variation between nodal characteristics and optimal connections across network types. This result along with the computational costs of the search for optimal solutions highlights the difficulty of finding effective heuristics. A novel genetic algorithm is proposed and we find this optimization heuristic outperforms existing techniques and describe how it can be applied to other combinatorial and dynamic problems.

Student active commuting to school is an important component to student achievement and student health. Increasing street and trail connectivity between residential developments and schools is a key to fostering student active commuting. In chapter 3, a cost-benefit analysis of increased connectivity around schools is conducted. Benefits, which include potential cost-savings to a school system if they had fewer students to bus to school, increased student walking, and the reduction in health-care costs of fewer obese students, are compared to the financial costs of the new connections. Advanced network optimization techniques were applied to several urban and suburban schools from a U.S.

school system to locate the optimal new connections that maximize student walking to a school. Results from this representative case study showed that short connections could lead to a large increase of potential student active commuters. This work can inform city planners, housing developers, and school officials on the impact of greater connectivity for student active commuting and residential development.

The dissertation is concluded with a section highlighting the results from each chapter along with providing future avenues of research for these topics.

Chapter 1

Hiking Eulerian Forests: Measuring Robustness and Coverage in Multi-line Networks

Abstract

Robustness, or the resilience of a system to failures and attacks, is an increasingly important characteristic of networks. Even though there have been proposed several measures of network robustness, there is a lack of analysis of these techniques for networks with multiple connections between the same nodes (multi-line networks). To address the issue of measuring vulnerability of multi-line network systems, a set of novel global indexes and a local index to measure robustness for these multi-line networks are introduced. These new indexes are designed to uncover the potential vulnerabilities of network components (i.e. hubs, terminals and lines) to possible malfunctions which may not be well identified with traditional connectivity matrix-based methods. Results of these measures are compared with traditional network connectivity and robustness indexes for simulated networks that vary topologically and are each composed of multiple lines that overlap in connectivity. A set of representative rail systems in the United States are used as a case study to provide insights for planners working to enhance preparedness for network disruptions, and highlight some future directions to further research in this realm.

Table 1.1: List of symbols and their definitions.

Symbol	Definition
A	Traditional (normalized) adjacency matrix
F	Forest adjacency matrix
e	Number of network edges
e_F	Number of Forest edges
ν	Number of network nodes
ν_F	Number of Forest nodes
T	Number of trails in a Forest
$\mathcal{T}_i$	Set of edges for trail i
$\lvert\mathcal{T}_i\rvert$	Length of trail i
p	Number of network subgraphs
α	Network connectivity index (α index) the ratio of existing circuits to the maximum number of circuits
β	Network connectivity index (β index) the average number of connections per node
γ	Network connectivity index (γ index) the ratio of existing routes to all potential routes
λ	Eigenvalue
Λ_F	Eigenvalue robustness index
J_F	Line coverage similarity index
C_F	Muti-line coverage comparison index
ℓ	Transit line
h	Hub ($h \in v_i^H$)
$\lvert h\rvert$	Number of hubs in a given network
$\lvert\ell^t\rvert$	Number of transit lines incident to hub h
$\lvert\ell^m\rvert$	Number of multiple transit lines incident to hub h
$\lvert\ell^h\rvert$	Number of intermediate and end stations incident to hub h
HS_h	Hub-stress index for hub h ($h \in v_i^H$)
HS^Global	Global Hub-stress index
HS^Total	Total Hub-stress index
$\lvert N\rvert$	Number of nodes in a given network
L	Number of transit lines
L_M	Maximum length of a transit line in a system
$\overline{L}$	Average transit line length
D_A^h	Highest degree of nodes in the traditional adjacency matrix
D_A^M	Average degree of nodes in the traditional adjacency matrix
D_F^h	Highest degree of nodes in the Forest adjacency matrix
D_F^M	Average degree of nodes in the Forest adjacency matrix

Introduction

Networks that have multiple connections between the same nodes (i.e. multi-line, multilayer, or multiplex networks) have always existed or evolve in many domains. For example, transportation networks regularly offer several modes of transit or multiple lines of the same type of transit between two locations, and information can be transmitted between the same two individuals through different forms of communication, by phone, email and in-person over the course of the same day. The expectation is that these multi-line networks would be more robust or resilient to failures or disruptions than single-line networks, but measures for robustness of these multi-line networks have yet to be explored or developed (Mattsson and Jenelius, 2015).

Network robustness and resilience, the ability for the network to still maintain connectivity after a disruption, has been traditionally evaluated from network connectivity. This connectivity has been mathematically characterized by adjacency matrices where the number of connections between nodes is normalized (Ellens and Kooij, 2013, Wu et al., 2011). Although this approach has been widely used and applied for network analysis because of its benefit to abstract complex system into a simple form, this adjacency matrix approach does not include the additional information from the existence of multiple connections between the same nodes in multi-line systems (Derrible and Kennedy, 2010b). The traditional global measures of connectivity, such as the α, β, and γ indexes, lose their scale and meaning when applied to multi-line systems. Therefore, the suite of network robustness measures that account for multiple lines should be expanded and this will allow those involved with transportation planning and emergency preparedness to make more informed decisions.

To do this, this paper initiates the process of developing multi-line network robustness indexes. Several novel and evident global and local indexes are presented to measure multi-line network resilience. First, modifications of the α, β, and γ indexes are introduced

for multi-line systems. These are followed by a novel global measure based on spectral graph theory, employing eigenvalues to study systems, which many modern approaches to network analysis rely on and has yet to be done with mulit-line networks. Eigenvalues have been used to study many phenomena in networks: the spread of diseases (Wang et al., 2003), finding the most connected individual (Newman, 2008) or measuring the speed of decision-making in groups (Gavrilets et al., 2016), and ranking webpages by their importance (Bryan and Leise, 2006). Specifically, the largest eigenvalue of a network has been linked to the robustness of the network (Kooij et al., 2008, Wu et al., 2011, Ellens and Kooij, 2013, Lyer et al., 2013, Zhao et al., 2016). By applying spectral graph theory, eigenvalues assess and become a useful barometer of the robustness of multi-line networks, enabling the comparison of resilience among networks regardless of their sizes and topology.

This paper also introduces a global robusteness index utilizing the inclusion-exclusion principle, a line coverage similarity index, which is analogous to network dissimilarity indexes (Alvarez-Socorro et al., 2015). This measure provides a way to view the differences and similarities of connectivity for all pair of lines in a network. The last global network index provided is based on the differences between traditional adjacency matrices and the new multi-line adjacency matrices, which can be used to weigh the impact of shared connectivity from the overlap of the systems multiple lines. This is similar to the road network vulnerability measure developed by Jun-qiang et al. (2017), but applied to multi-line networks.

For local level robustness measures, an index to identify nodes that are significant in the resilience of multi-line systems is proposed. In complex networks, some nodes are more important than others. The more critical the node becomes, the more likely their elimination induces the network's collapse, and identifying them is crucial in many circumstances. Here, Morone and Makse (2015) devised a rigorous method to determine the most influential nodes in random networks by mapping the problem onto optimal percola-

tion and solving the optimization problem with an algorithm that the authors call 'collective influence'. They find that the number of optimal influencers, or hubs, is much smaller, and that low-degree nodes can play a much more important role in the network than previously thought. Not all of the hubs are critical in terms of newtork resilience (Kovacs and Barabási, 2015). A way for identifying the critical ones is introduced in this paper and is based on the number of low-degree stations, such as intermediate stations and/or end stations, which are sequentially but cumulatively linked to a hub.

Given the traditional normalized representation of network connections raises a critical issue when it is applied to rail systems because of the type of stations and the level of connectivity by lines. As proposed by Derrible and Kennedy (2010b) and Derrible (2012), a graphic representation for the transit systems requires the definition of the concepts of nodes and edges in a different manner. Specifically, they defined the differentiated nodes as transfer and end nodes and edges as single and multiple edges. These concepts were designed to investigate the structure of transit systems to see the relationships between directness and complexity. Note that the stations between hubs or end stations are ignored when a network is abstracted with graph form as reflecting them into a matrix is not necessary in computing the proposed indicators. However, intermediate stations are also critical components which affect the level of resilience of a network (Kim et al., 2016a) because they cannot provide any option to reroute until flows are accumulated and they reach hub stations. The more intermediate nodes in a network, the more vulnerable or less resilient the network would become. In addition, the hub incident with more intermediate nodes in a line can have a greater resilience to handle flows compared to hubs with a small number of intermediate nodes.

Hubs, in terms of multi-line networks, serve as transfer stations for passengers when they need switching their routes to reduce transportation cost or find alternative paths. With a handful of hubs, the total network cost is reduced and the hub-and-spoke type systems enhance network efficiencies in operation because of economy of scales of flows

on the network (O'Kelly and Bryan, 1998). It is known that the malfunctioning of these special nodes can dramatically decrease the operation of the network (Kim et al., 2016a, Kim and Ryerson, 2017) and affect network vulnerability and resilience to network failure (O'Kelly, 2015). However, a recent work argues that not all of hubs are critical depending on the network structure or flows among nodes (Kim, 2012, Kovacs and Barabási, 2015), raising a fundamental question on what conditions would be the best determinants or factors to affect network vulnerability and resilience. The paper provides a way for identifying the critical ones based on the number of low-degree stations, such as intermediate stations and/or end stations, which are sequentially but cumulatively linked to a hub. This is particularly crucial in the case of closed-systems with multi-linkages such as heavy rail or subway systems.

Traditionally, transit resilience and robustness have been associated largely with travel time reliability and variability (Levinson, 2005, Scott et al., 2006). It is still an important topic today from quantifying variability itself (Mazloumi et al., 2011, Kieu et al., 2015) or its cost (Benezech and Coulombel, 2013, Sohn, 2006, Cats and Jenelius, 2014, Taylor and Susilawati, 2012), to using reliability and variability as a design criterion (Yao et al., 2014, An and Lo, 2014). Recently, the field of Network Science (Newman, 2010) has emerged as particularly fitted to measure the robustness of a system, notably by studying the impact of cascading failure (Watts, 2002, Crucitti et al., 2004, Kinney et al., 2005). Indeed, as physical networks, metros are composed of stations (nodes) and rail lines (links), and they therefore possess measurable network properties (Derrible and Kennedy, 2010b) that can be used to study their robustness (Berche et al., 2009, von Ferber et al., 2012). Several works have also tried to combine information from both transit operation and network properties to gain insight into the robustness of transit networks (Rodriguez-Nunez and Garcia-Palomares, 2014, Kim et al., 2016a).

A comparison of the results from these new measures with traditional connectivity indexes for US city rail networks (rapid transit systems) is conducted. These transportation

systems are electric railway with the capacity to handle a heavy volume of traffic (Dickens, 2016) and are composed of multiple lines and hubs stations (Derrible and Kennedy, 2010b). These multiple-lines facilitate movement of passengers in the transit systems and the resilience of these transit systems are clearly associated with the placement of hubs for transferring passengers (Kim and O'Kelly, 2009, Li and Kim, 2014). Measuring robustness and assessing the system resilience are critical concerns as the portion of service for ridership is considerably increasing in highly dense urban areas. Although these networks are comparatively safer in operations than other modes of transportation, the number of incidents, injuries, and fatalities by collision and derailments has increased over time and become a serious concern (Nelson and Streit, 2011). A major issue is that the portion of service for ridership is substantially increased in highly dense urban areas, stressing that the role of transfer stations (or hubs) is vital to handle passenger volumes, and the need of measures to evaluate the performance of network structures.

This paper is organized as follows. The next section begins with a re-definition of network components to better describe multi-line systems. Then, a set of standardized resilience indexes for the robustness of networks to be compared directly is described. Section 3 examines the sensitivity of system resilience to different levels of network topology with generated networks. A set of representative rail systems in the United States are also used for this analysis and as a case study to provide insights for planners to enhance preparedness of network emergency. The concluding section highlights some future directions to further research in measuring multi-line network robustness.

Index Development

Eulerian Forests

To explore multi-line network robustness the focus of this analysis begins with a subset of these networks called Eulerian Forests. Eulerian Forests are multi-line networks com-

posed of Eulerian trails (see Fig. 1.1). Eulerian trails are paths in a graph that traverse every edge of the path exactly once, and networks composed of separate lines, such as rail networks (Wilson, 1986), can generally be classified as a collection Eulerian trails, i.e. Eulerian Forests. To frame this in terms of transportation networks, a Eulerian trail represents a single transit line within a transportation network, e.g. a line of subway system or an interstate road of a highway network, and the entire transportation system is classified as a Eulerian Forest.

To develop robustness measures for these Forests start with the "adjacency matrix" for Forests with T trails, called an F-matrix. The elements of an F-matrix, using the notation $F_{i,j}$, are equal to the number of trails connecting nodes i and j, therefore $F_{i,j} \in (0, 1, \cdots, T)$. The structure of an F-matrix clearly differentiates it from traditional adjacency matrices which are normalized such that $A_{i,j} = 0$ or 1. The common notation of e for a network edge and ν for a network node will continue to be used for the analysis of these Forests, while $\mathcal{T}_i$ is used to denote the set of connections for trail i and its corresponding trail length will be given by $|\mathcal{T}_i|$. Fig. 1.1 provides a visual description and comparison of three different Eulerian Forests with three trails ($T = 3$) each trail connecting four nodes ($|\mathcal{T}_i| = 3 \,\forall\, i$). Forest (a) has the maximal amount of spatial coverage but is not robust since all nodes are connected by a single line (i.e. a disconnected Forest), (b) offers some geographic coverage, the network property of covering additional nodes, along with resilience for some of the nodes, and (c) is the most robust form of the Forest but minimizes nodal coverage (i.e. an uncut forest). The first row of matrices provides the respective adjacency matrices while the Forest adjacency matrices are on the bottom. Note, networks (b) and (c) are resilient to line failures owing to their multiple lines connecting the same nodes. The novel robustness indexes introduced below are scaled to enable Forests with different compositions (different number and length of trails) to be compared directly.

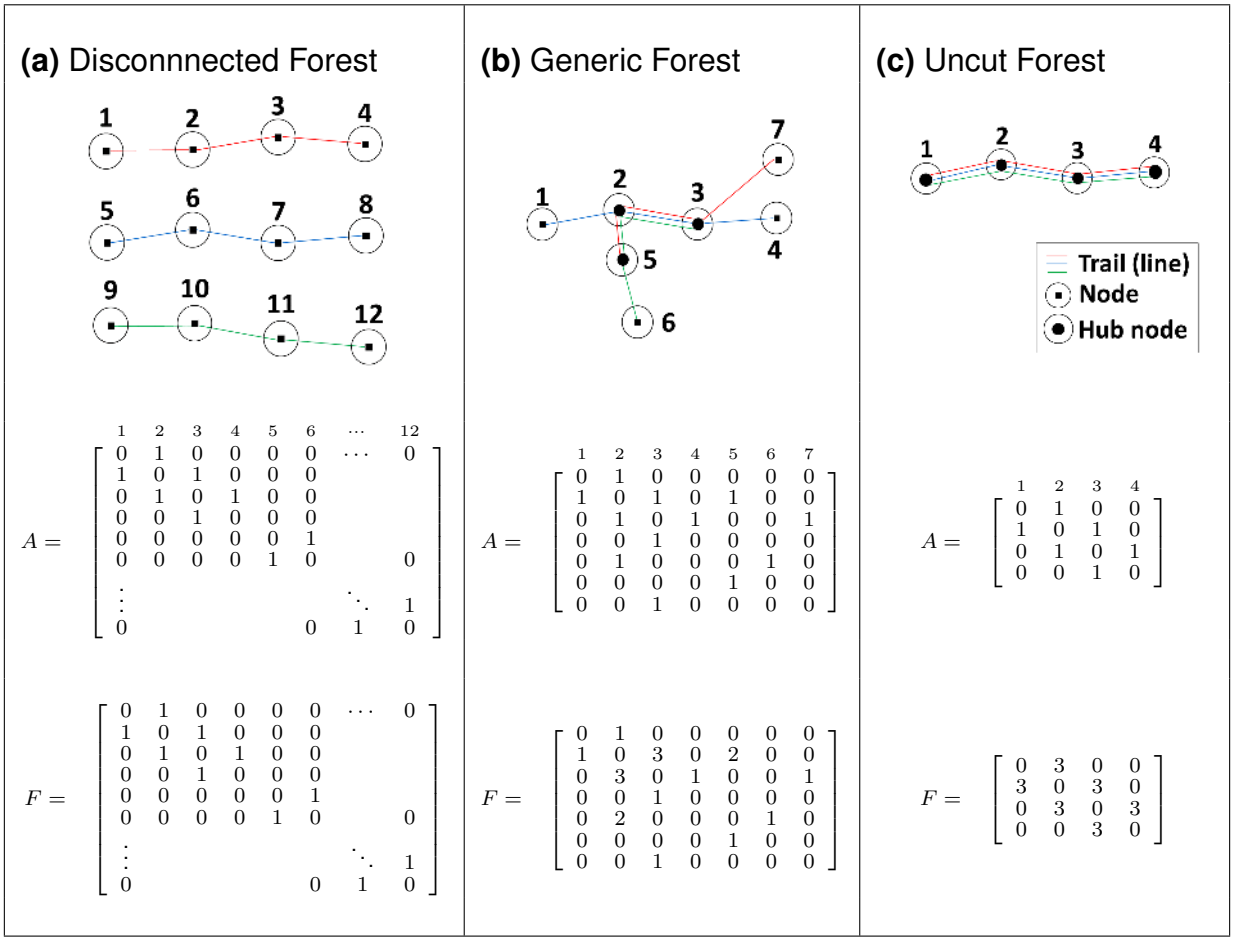


Figure 1.1: Comparison of three Eulerian Forests with the same number and length of trails.

Modified Traditional Global Connectivity Indexes ($\alpha_F, \beta_F, \gamma_F$)

The traditional global indexes of network connectivity, which are used to evaluate nodal network robustness, are the α, β and γ indexes (Garrison and Marble, 1958, Kansky and Danscoine, 1989, Taeffe et al., 1996). The α index (Eq. 1.1), which is commonly interpreted as the ratio of existing circuits to the maximum possible circuits, is given by

$$\alpha = \frac{e - \nu + p}{2\nu - 5} \in [0, 1], \tag{1.1}$$

where p is the number of subgraphs. The lower bound is derived from the most minimally connected graph, where $e = \nu - 1$, and the upper bound is established from the complete graph, where the maximum number of edges $e = \nu(\nu - 1)/2$. The β index (Eq. 1.2),

$$\beta = \frac{e}{\nu} \in [0, (\nu - 1)/2], \tag{1.2}$$

which provides a simple estimate for the average number of connections for a node in the network, with a $\beta = 0$ for networks without any edges and an upper bound which can be greater than 1. The γ index (Eq. 1.3),

$$\gamma = \frac{e}{3(\nu - 2)} \in [0, 1], \tag{1.3}$$

is understood as the amount of existing routes in a network to the number of all potential routes for that given network, and which provides a degree of the networks complexity. In general, β and γ indexes are scaled in such a way that they are useful to compare the complexity of networks with different sizes. Notice that these indexes may represent the complexity of a network, however, they do not capture the robustness of a network. For example, networks (a) and (c) in Fig. 1.1 have the same β value and network (b) has a larger value. There is also little variation in the γ values for these networks.

When these indexes are applied to F-matrices, they no longer have the same range of values (i.e. scale), therefore interpretation and network comparison is lost. These traditional measures are modified to take into account some of the additional information from an F-matrix through rescaling. With the upper bounds of Forests determined by $e = T\,\mathbf{max}\,|\mathcal{T}_i|$ and $\nu = \mathbf{max}\,|\mathcal{T}_i| + 1$, the following modified α connectivity index is proposed (Eq. 1.4):

$$\alpha_F = \left(\frac{e_F - \nu_F + 1}{2\nu_F - 5}\right)\left(\frac{2\ \mathbf{max}\,|\mathcal{T}_i| - 3}{(T-1)\,\mathbf{max}\,|\mathcal{T}_i| + p - 1}\right) \in [0,1], \tag{1.4}$$

the modified β index (Eq. 1.5):

$$\beta_F = \left(\frac{e_F}{\nu_F}\right)\left(\frac{\mathbf{max}\,|\mathcal{T}_i| + 1}{T\,\mathbf{max}\,|\mathcal{T}_i|}\right) \in [0,1], \tag{1.5}$$

note that β_F is now normalized, and the modified γ index (Eq. 1.6):

$$\gamma_F = \left(\frac{e_F}{3(\nu_F - 2)}\right)\left(\frac{3(\mathbf{max}\,|\mathcal{T}_i| - 1)}{T\,\mathbf{max}\,|\mathcal{T}_i|}\right) \in [0,1]. \tag{1.6}$$

As these global indexes approach unity the network is more connected and they now have the same scale regardless of network topology or the number/length of trails. These measures represent the relationship between edges and nodes in the network and the number of and lengths of trails present in the system, but overlapping trails, i.e. trails that share the same nodal connections, do not directly contribute to them. In addition to these global indexes of connectivity, several additional indexes are introduced in the following subsections which account for the additional robustness when trails overlap.

Multi-Line Eigenvalue Robustness Index (Λ_F)

This global index introduced for line network robustness is based on the eigenvalue of a multi-line network. Robustness measures developed from spectral graph theory are currently not applicable to an analysis of F-matrices since they rely on the normalized

adjacency matrices. To compensate for this shortcoming, reliance on some general inferences from linear algebra (see Appendix A.) are used to develop a robustness measure from the F-matrix eigenvalues (Eq. 1.7):

$$\Lambda_F = \frac{\lambda_M - 2}{2(T-1)} \in [0, 1]. \tag{1.7}$$

Since Λ_F is based on λ_M, which is directly related to the degree (i.e. connectivity) of nodes, it provides a measure of Forest robustness. This measure reveals the robustness of multi-line networks in terms of (i) the variance in the connectivity of the nodes and is bounded below by the average number of connections per node and above by the maximum number of connections for a node in the network, and (ii) not only the number of connections for a node but the number of connections for neighboring nodes. When $\Lambda_F \to 0$ there is less variance in the number of connections per node at the same time a lower number of connections for nodes, and more geographic coverage in the Forest at the sacrifice of resilience while $\Lambda_F \to 1$ the Forest is more robust but offers less geographic coverage.

Line Coverage Similarity Index (J_F)

This new global measure of robustness is developed from the idea of the Jaccard index (Jaccard, 1901). The Jaccard index was originally designed to represent the similarity (or dissimilarity) of a pair of sets, and it has modified it to measure the similarity of a pair of trails. This similarity of trails provides an index of trail pair robustness and when averaged over all of the trail combinations for a Forest the following index is produce (Eq. 1.8):

$$J_F = \left[\prod_{i,j>i}^{T} \left(\frac{|\mathcal{T}_i \cap \mathcal{T}_j|}{|\mathcal{T}_i \cup \mathcal{T}_j|} + 1 \right) \right]^{1/\binom{T}{2}} - 1 \ \in [0, 1], \tag{1.8}$$

where $|\mathcal{T}_i \cap \mathcal{T}_j|$ is the number of edges trail i and trail j share in common (i.e. overlap), and $|\mathcal{T}_i \cup \mathcal{T}_j|$ is the total number of edges for both trails i and j. The index provides the geometric mean of these ratios, with the total number of trail combinations for the Forest, given by $\binom{T}{2}$. In case a pair of trails does not share any connections, producing a 0 denominator, a 1 is added to the logarithm and removed from the final measure, which can result in lower predicted values for small networks ($\nu \leq 10$). For this scale, as $J_F \to 0$ lines are less likely to overlap, there is less similarity in line pair coverage and therefore a lower Forest robustness, while $J_F \to 1$ lines overlap more often, there is more similarity between lines and the Forest should be more resilient.

Edge Comparison Index (C_F)

A comparison of the number of edges in the traditional adjacency matrix for a Forest and it's F-matrix can provide another global measure for robustness. This index is a simple representation of the additional connection present in a multi-line network (Eq. 1.9):

$$C_F = \left(\frac{e_F - e}{e_F}\right)\left(\frac{\sum_{i=1}^{T}|\mathcal{T}_i|}{\sum_{i=1}^{T}|\mathcal{T}_i| - \mathbf{max}\,|\mathcal{T}_i|}\right) \in [0,1] \tag{1.9}$$

where

$$\frac{e_F - e}{e_F} = \left(1 - \frac{\sum_{i=1}^{T-1}\sum_{j>i}^{T} A_{i,j}}{\sum_{i=1}^{T-1}\sum_{j>1}^{T} F_{i,j}}\right).$$

The first part of this index represents the fraction of multiple connections in the network, while the second part ensures it is normalized. If a Forest is composed of trails that do not overlap, then $e_F \to e$ and $C_F \to 0$, while a more robust Forest will have a greater distance between e_F and e, and $C_F \to 1$.

Hub-Stress Index (HSI)

The HSI is designed to measure the criticality of a hub station based on the number of intermediate stations and/or end stations which are sequentially but cumulatively linked to a hub. This index is well applicable to the networks where a series of intermediate stations are connected from an end station to a handful of hubs via transit lines and multiple linkages are built among hubs for the enhanced transferability of passengers. The index is calculated using the Eq. 1.10 for individual hub h:

$$HS_h = \sqrt{\frac{\sum_{\ell \in h}(|\ell^h|)^2}{|\ell^t|/(1+|\ell^m|)}}, \tag{1.10}$$

where h is the index of hubs, ℓ is the index of transit line, which is incident to hub h, $|\ell^h|$ is the number of intermediate and end stations in transit line ℓ incident to hub h from adjacent hub in transit lines ℓ, $|\ell^t|$ is the number of transit lines incident to hub h, and $|\ell^m|$ is the number of multiple transit lines incident to hub h.

Notice that the more intermediate stations are linked successively until the terminal intermediate station is incident to the hub station, the greater hub stress is expected because the flows from intermediate stations to the hub are delivered only a single path without any chance of re-routing. The index ranges $[0 < HS_h \leq |N| - 1]$, where the greatest value is found when all stations except a hub station are connected to a hub as single sequence. Fig. 1.2 presents the properties of HSI to the density of connections (Fig. 1.2 a and b), stress-sharing (c and d), and congestion level by multiple transit lines among hubs (e and f). Given the index value, a hub becomes more stressful if the more intermediate and end stations are incident to the hub (b), compared to a smaller number of stations are connected to a hub (a). In case of multiple hubs are connected to each other, the hub stress reduced (c) because the flows among hubs can be transferred to the shared hub links compared to the case that a single hub takes care of all flows from the lines (d). In terms of the number of existence of multiple lines to a hub, the hub stress

increases with more multiple-lines since more flows can transit into the hub (f), rather than a single line serve (e). This individual index is extended to two global network level indexes, Total HSI (HS^{Total}) and Global HSI (HS^{Global}), which is the sum of HS_h and the average stress on hubs in a transit system, respectively.

$$HS^{Total} = \sum_{\ell \in h} HS_h, \tag{1.11}$$

$$HS^{Global} = \sum_{\ell \in h} HS_h/|h|. \tag{1.12}$$

Numerical Experiments

Simulated Forests

To evaluate the novel robustness measures they were first applied to generated Eulerian Forests of different sizes and topologies. Simulated networks were created to observe how the new indexes responded to variation in network structures. Networks were generated with variation in robustness, i.e. multi-line coverage, from systems with no overlapping lines (no resilience) to a topology where all lines overlapped (maximum robustness). As shown in Table 1.2, the global robustness indexes (α, β, and γ) when applied to the generated Forests in the network examples from Fig. 1.1 are not effective at identifying any of the additional connectivity provided by the existence of overlapping lines, which are captured by α_F, β_F, γ_F, Λ_F, J_F, and C_F. It is clearly seen that the α index does not represent any of the variation in the robustness across the networks, and the β index values are inconsistent when presented with multi-line networks, and the γ index captures little of the variation between the Forest network topologies. Whereas these global connectivity indexes did not measure the resilience of having additional connections between the same nodes, the new connectivity indexes (α_F, β_F, and γ_F) and robustness measures

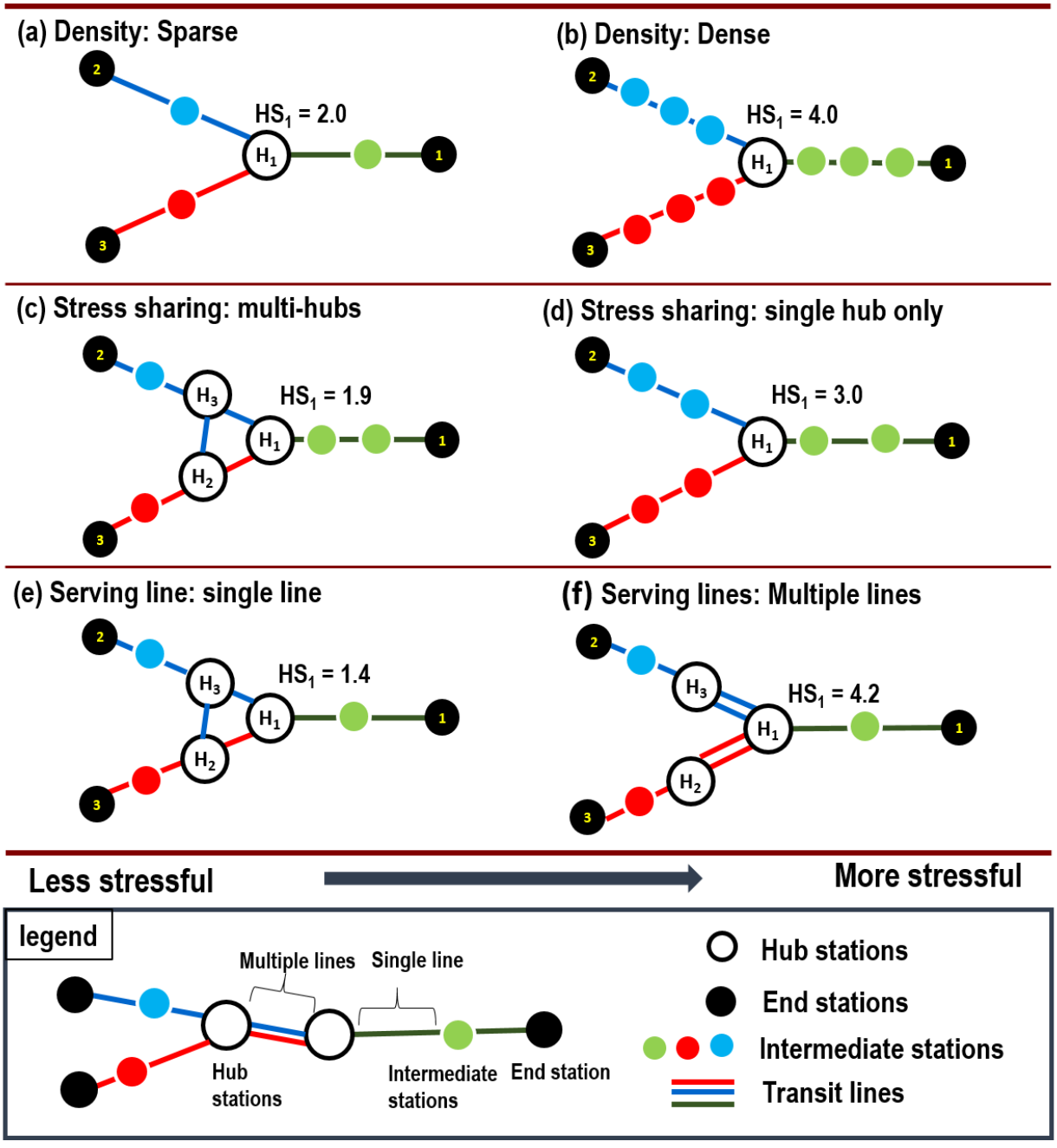


Figure 1.2: Characteristics of the Hub Stress index.

(Λ_F, J_F, and C_F) when applied to the F-matrices are sensitive to the existence of multiple connections in a clear and consistent fashion. There is a clear trend of increasing index value corresponding to increasing robustness in the new measures. For example, the most robust network (Forest C) results in the α_F, β_F, γ_F, Λ_F, J_F, and C_F indexes having a maximum value at or close to unity, while the α, β, and γ indexes do not. Similarly, the least robust network (Forest A) results in the α_F, β_F, γ_F, Λ_F, J_F, and C_F indexes having a minimum value at or close to 0, while the β, and γ indexes do not. Note, the β_F and γ_F index values for the least resilient Forest (Fig. 1.1(a)) is slightly greater than 0 while Λ_F slightly less than unity due for the most resilient Forest (Fig. 1.1(c)) due to sensitivities of these indexes to small network sizes ($\nu \leq 10$).

Table 1.2: Network characteristics and global connectivity and robustness measures for the Forests in Fig. 1.1

					Nodal Indexes						Line Indexes		
Forest	ν	e	e_F	p	α	α_F	β	β_F	γ	γ_F	Λ_F	J_F	C_F
A	12	9	9	3	0	0.000	0.750	0.333	0.300	0.200	0.000	0.000	0.0
B	7	6	9	1	0	0.167	0.857	0.571	0.400	0.600	0.587	0.293	0.5
C	4	3	9	1	0	1.000	0.750	1.000	0.500	1.000	0.814	1.000	1.0

Fig. 1.3 presents the distributions of the index values for the simulated Forest network topologies highlight the differences between the indexes applied to the connectivity and the F-matrices. To generalize the characteristics for these index value distributions, 10000 random network topologies were created for Forests with 5 trails connecting 10 nodes each. Small Forest networks were generated to reduce computational burden from the set of network permutations that grows factorially as the number of nodes increases. The index values are slightly sensitive to Forest size and will be greater for larger Forests networks, i.e. the distributions will shift slightly right as the number of nodes or trails increases. Without loss of generality, the new measures, when applied to the F-matrices respond to the variance in the robustness of the networks. The α_F index is more responsive than the α index (Fig. 1.3(a)), as a large amount of shared connectivity between

lines is necessary for increasing the index value. The β index values are not bounded to the same ranges as the other measures and the index does not vary as much as the β_F measure even though the networks were generated with the minimum and maximum robustness (Fig. 1.3(b)). Similarly, the γ index does not cover the range of resilience present in the networks which the γ_F measure does (Fig. 1.3(c)). The line coverage similarity index (J_F) and the edge comparison index (C_F) similarly respond to the additional information provided by the F-matrices with respect to network robustness (see Fig. 1.3(d)). Notably, the new connectivity measures (α_F, β_F, γ_F, Λ_F, J_F, and C_F) cover the entire range of values, as they should since the networks range from not being robust to having maximum resilience. The α, β, and γ indexes do not share the same range of values, even for the same set of networks, nor do they cover the entire range of possible values for each index.

Fig. 1.4 highlights how the eigenvalue robustness index (Λ_F) when applied to the simulated Forest networks can represent the robustness of the simulated networks as the number of overlapping lines varies. Plots of the largest and second largest eigenvalues for different Forest sizes and trails lengths are provided in Fig. 1.4 to demonstrate the clear relationship between network robustness and eigenvalues. Fig. 1.4(a) and (b) show the results for 5 trails that connect 10 and 20 nodes respectively, Fig. 1.4(c) and (d) display the results for 10 trails that connect 10 and 20 nodes respectively, and Fig. 1.4(e) shows the results for a 5 trail Forest with 10, 15, 20, 25, and 30 trail lengths. Squares represent the most covered Forest and the diamond represents the most robust Forest for the given number and length of trails. Forests with trail overlappings, i.e. multiple trails connecting the same set of nodes, are separated with different colors. Clusters exist from the design of the sampling algorithm to ensure selection of various types of Forest network topologies, otherwise the eigenvalues would be more continuous across the space. Horizontal lines represent the eigenvalues analytic bounds $[2, 2T]$, a diagonal line is given for reference, and 10000 Forests were created for each trail length and number combi-

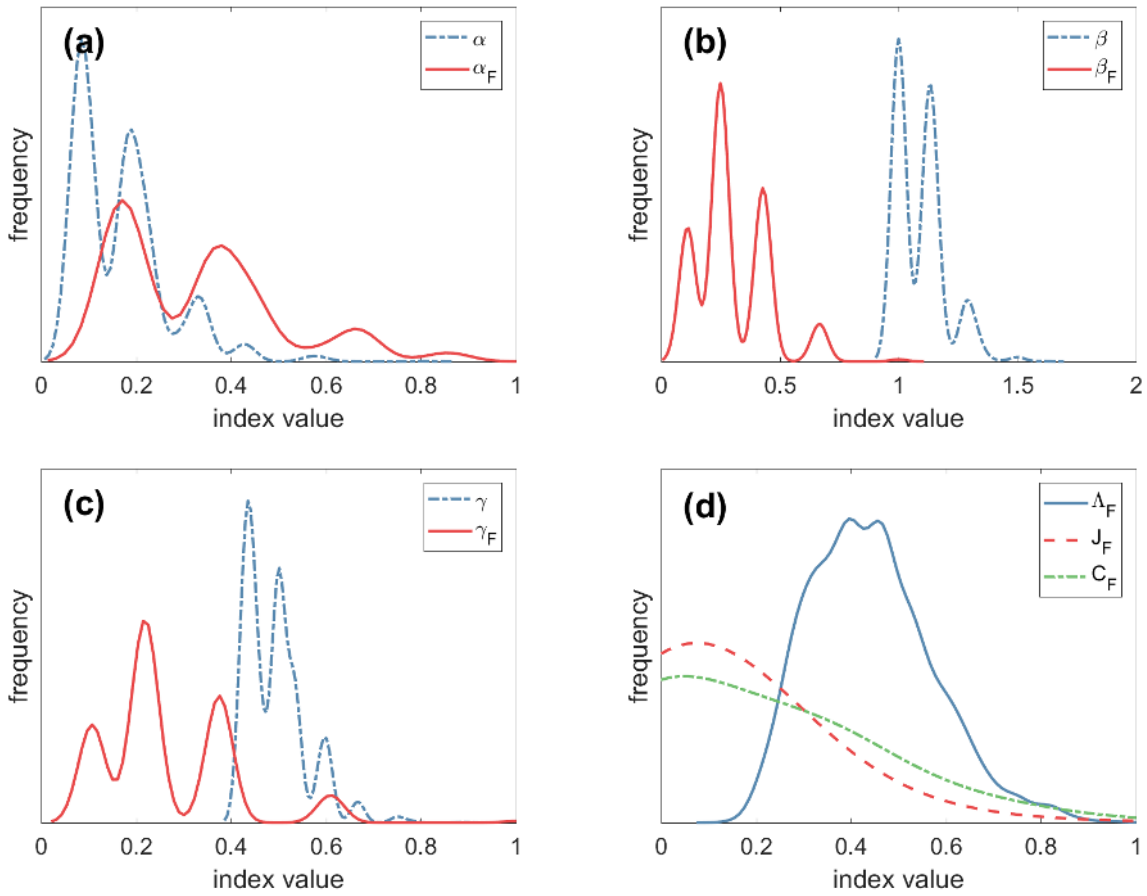


Figure 1.3: Distributions of the various global robustness measures applied to the same set of simulated Forests (10000 random networks).

nation. Even though the second largest eigenvalue is not used in the index it provides a dimension for clarity of pattern trends and additional network systems were explored in the Appendix (Fig. 1.A1). Notice that the eigenvalues are bounded by $[2, 2T]$ regardless of network size, length and number of trails, and topology. As the robustness of the network increases the largest eigenvalue approaches to $2T$, which translates to $\Lambda_F = 1$ and when the network is less resilient the largest eigenvalue approaches 2, i.e. $\Lambda_F = 0$. It is evident that when there are additional trail overlaps (e.g. two to three trails connecting the same set of nodes), the eigenvalue clearly step to a larger value.

Table 1.3 summarizes the correlations between the new indexes and the old measures for the simulated networks. Notice that when the new indexes (α_F, β_F, and γ_F) are applied

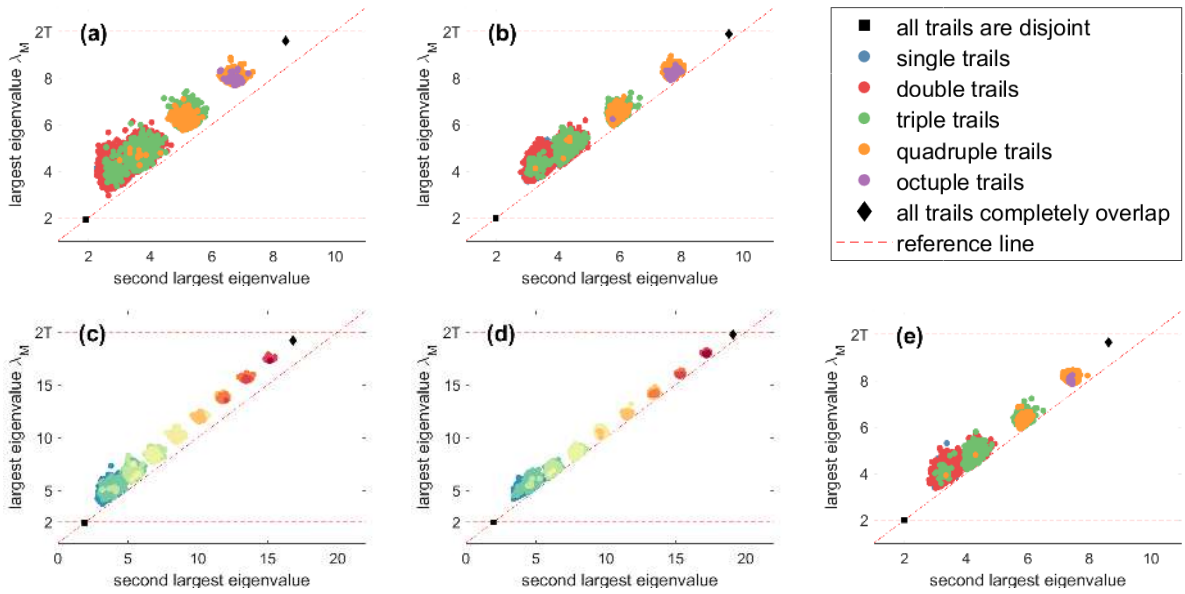


Figure 1.4: Plots of the largest and second largest eigenvalues for different Forest sizes and trails lengths.

to the F-matrices for the same set of networks, they are more strongly correlated than the α, β, and γ indexes (upper-left and middle boxes). The J_F and C_F indexes are also highly correlated with each other, which makes sense as they both measure pairwise line overlap (lower-right box). Most importantly, the α, β, and γ indexes correlate weakly with the α_F, β_F, and γ_F measures and weakly to not at all with the line robustness Λ_F, J_F, and C_F indexes. These new indexes are not inclusive to the traditional measures but inclusive within the family of measures.

Case Study with US City Rail Systems

A comparison of the results from these new measures for US city rail networks (rapid transit systems) is conducted as a case study of transportation systems composed of multiple lines and hubs stations (Derrible and Kennedy, 2010b). The set of the twelve largest rail systems, by ridership, with multiple lines in the United States (Dickens, 2016) was used for the real-world case study which provides variation in line numbers, line lengths, and network topology. Network characteristics and index values for these rail systems are

Table 1.3: Correlations between the global measures for the simulated Forests (10000 random networks).

	α	β	γ	α_F	β_F	γ_F	Λ_F	J_F	C_F
α		0.985	0.960	0.808	0.815	0.807	0.457	-0.069	-0.070
β	0.985		0.904	0.709	0.724	0.706	0.348	-0.209	-0.209
γ	0.960	0.904		0.940	0.941	0.938	0.649	0.205	0.210
α_F	0.808	0.709	0.940		0.995	1.000	0.808	0.514	0.518
β_F	0.815	0.724	0.941	0.995		0.993	0.820	0.509	0.516
γ_F	0.807	0.706	0.938	1.000	0.993		0.805	0.515	0.518
Λ_F	0.457	0.348	0.649	0.808	0.820	0.805		0.745	0.747
J_F	-0.069	-0.209	0.205	0.514	0.509	0.515	0.745		0.981
C_F	-0.070	-0.209	0.210	0.518	0.516	0.518	0.747	0.981	

given in Table 1.4. The α index is consistently not sensitive to the robustness of the systems whereas the α_F is. Even though there is considerable variation in connectivity in the rail networks there is little variation in the β and γ indexes ($[0.974, 1.143]$ and $[0.336, 0.424]$, respectively), while the β_F and γ_F indexes do vary according to this range of network connectivity ($[0.132, 0.549]$ and $[0.114, 0.504]$, respectively).

The correlation of the network characteristics and novel robustness measures for the twelve largest United States city rail systems is provided in Table 1.5. It should be noted that the new global measures are all inversely related to the size of the network, whether it be the number of nodes, edges, or multi-line edges, confirming the assumpiton that as larger networks are built for geographic coverage they sacrifice resilience. Similar to the simulated networks, the J_F and C_F indexes are highly correlated for the rail systems. Table 1.5 also summarizes the result of hub stress indexes for 12 heavy rail networks. Notice that the HS^{Total} increases with the number of hubs in a network (r=0.907). However, it does not necessarily indicates that hub stress at individual level in a large network also greater than that of a small network (r=0.054). For example, NYMTA is the largest network among 12 networks, but its HS^{Global} is much smaller than other mid-size networks, ranked 10th. In contrast, the smallest network, SEPTA shows the 2nd highest HS^{Global}

Table 1.4: Network characteristics and robustness index values for the twelve largest United States city rail systems with multiple lines.

	PATH	MARTA	MDT	BART	RTA	SEPTA	LACMTA	BTA	MBTA	WMATA	CTA	NYMTA
ν	13	38	42	47	50	53	81	107	109	141	158	419
e	14	37	42	47	49	52	80	108	108	140	162	479
e_F	20	57	59	99	63	52	85	134	109	153	169	529
L	4	4	5	4	3	2	6	6	8	6	8	9
L_M	7	18	21	24	23	27	20	34	19	33	33	108
$\overline{L}$	5.00	14.25	12.40	21.00	22.00	26.50	15.00	22.33	14.50	26.50	22.00	59.78
D_A^h	4	4	4	3	3	3	4	4	4	10	6	8
D_F^h	4	8	7	8	6	3	6	8	4	10	8	9
D_A^M	2.153	1.947	2.000	2.000	1.960	1.925	1.938	2.019	1.963	2.17	2.051	2.286
D_F^M	3.077	3.000	1.982	4.213	2.520	1.925	2.086	2.505	1.981	2.170	2.133	2.525
α	0.095	0.000	0.013	0.011	0.000	0.000	0.000	0.010	0.000	0.000	0.016	0.073
α_F	0.200	0.172	0.106	0.372	0.138	0.000	0.012	0.051	0.001	0.018	0.011	0.033
β	1.080	0.974	1.000	1.000	0.980	0.981	0.988	1.010	1.000	0.993	1.025	1.143
β_F	0.440	0.400	0.294	0.549	0.438	0.509	0.184	0.215	0.132	0.186	0.138	0.142
γ	0.424	0.343	0.350	0.348	0.340	0.340	0.338	0.343	0.340	0.336	0.346	0.383
γ_F	0.325	0.352	0.270	0.504	0.400	0.472	0.161	0.200	0.114	0.172	0.127	0.138
λ_M	3.605	4.609	4.112	7.715	4.909	2.108	4.034	5.686	2.998	3.556	4.213	4.537
Λ_F	0.067	0.035	0.026	0.062	0.030	0.001	0.013	0.017	0.005	0.006	0.007	0.003
J_F	0.000	0.125	0.055	0.187	0.120	0.000	0.019	0.010	0.001	0.000	0.009	0.009
C_F	0.000	0.385	0.288	0.525	0.222	0.000	0.071	0.021	0.009	0.000	0.041	0.095
$\|h\|$	11	21	22	29	13	3	9	25	7	34	13	126
HS^{Total}	29.00	53.71	82.37	95.17	62.68	32.77	55.94	125.17	42.51	132.37	202.36	404.38
HS^{Global}	2.64	2.56	3.74	3.28	4.82	10.92	6.22	5.69	6.07	4.14	15.57	3.21
max HS_h	3.54	6.48	10.07	14.88	15.59	16.16	14.46	34.05	12.87	25.62	32.6	12.79
min HS_h	1.41	1.41	1.41	1.41	1.41	6.27	2.12	1.41	1.73	1.41	2.45	0.71

Table 1.5: Correlations between the global measures for the twelve US rail networks.

	ν	e	e_F	α	β	γ	α_F	β_F	γ_F	Λ_F	J_F	C_F	$\|h\|$	HS^T	HS^G
ν		0.999	0.990	0.361	0.728	0.155	-0.404	-0.605	-0.560	-0.526	-0.346	-0.266	0.895	0.951	0.045
e	0.999		0.994	0.393	0.752	0.190	-0.375	-0.574	-0.532	-0.495	-0.328	-0.244	0.911	0.954	0.017
e_F	0.990	0.994		0.399	0.755	0.195	-0.279	-0.518	-0.472	-0.423	-0.238	-0.151	0.942	0.965	-0.043
α	0.361	0.393	0.399		0.896	0.975	0.187	0.009	-0.098	0.376	-0.272	-0.208	0.491	0.406	-0.279
β	0.728	0.752	0.755	0.896		0.779	-0.035	-0.282	-0.337	0.039	-0.348	-0.257	0.772	0.743	-0.171
γ	0.155	0.190	0.195	0.975	0.779		0.280	0.144	0.017	0.510	-0.215	-0.168	0.315	0.199	-0.325
α_F	-0.404	-0.375	-0.279	0.187	-0.035	0.280		0.720	0.688	0.917	0.822	0.798	-0.050	-0.226	-0.508
β_F	-0.605	-0.574	-0.518	0.009	-0.282	0.144	0.720		0.987	0.671	0.621	0.532	-0.316	-0.502	-0.229
γ_F	-0.560	-0.532	-0.472	-0.098	-0.337	0.017	0.688	0.987		0.584	0.663	0.576	-0.282	-0.456	-0.187
Λ_F	-0.526	-0.495	-0.423	0.376	0.039	0.510	0.917	0.671	0.584		0.597	0.564	-0.196	-0.366	-0.510
J_F	-0.345	-0.328	-0.238	-0.272	-0.348	-0.215	0.822	0.621	0.663	0.597		0.949	-0.063	-0.195	-0.371
C_F	-0.266	-0.244	-0.151	-0.208	-0.257	-0.168	0.798	0.532	0.576	0.564	0.949		0.047	-0.099	-0.413
$\|h\|$	0.895	0.911	0.942	0.491	0.772	0.315	-0.050	-0.316	-0.282	-0.196	-0.063	0.047		0.907	-0.323
HS^T	0.951	0.954	0.965	0.406	0.743	0.199	-0.226	-0.502	-0.456	-0.366	-0.195	-0.098	0.907		0.055
HS^G	0.0450	0.017	-0.043	-0.279	-0.171	-0.325	-0.508	-0.229	-0.187	-0.510	-0.371	-0.413	-0.323	0.055	

among them. The SEPTA is characterized as a typical network having denser intermediate stations, little stress-sharing but single-line based structure, resulting in placing in the 2nd rank.

As Fig. 1.5 displays, heavy rail networks can be classified into four groups regarding hub stress levels using HS^{Total} (X-axis) and HS^{Global} (Y-axis). In this Figure, the median values of each measure are used as a reference to draw the quadrats. Given this classification, some observations are worthy to note. First, obviously, the NYMTA is the most stressful network with the largest value of HS^{Total} because it is the largest heavy rail with many hub stations. However, it does not necessarily imply that the hubs are also highly stressful because its HS^{Global} indicates its average hub stress falls into the class of low (Q4). Rather, such networks as the CTA and the BTA placed in Q1 should be highlighted as the most stressful networks. In the Q2, the SEPTA is noticeable because of its large HS^{Global}. Based upon both hub stress indexes, arguably, the networks in the Q2 are the most vulnerable when their hubs face any malfunction because of their great network dependency on hubs. In contrast, the networks in the Q3 such as the MARTA and the PATH would be least vulnerable because their network size is relatively smaller than other networks and its hub dependency from intermediate stations is low.

Combined with Fig. 1.5, Fig. 1.6 shows hub stress on the networks at the individual station level from a different dimension using box-plots. Note that the upper outliers in the Box-plot represent the extremely high hub stress stations, the more outliers the network has, the more dependable the network operates via hubs. Given this indication, the BTA and the WMATA are considered vulnerable if a few of “highly” stressful hubs are in malfunction, though they are positioned on the marginal line of Q1 and Q4, respectively. The decent size of rail systems, for example, the RTA, BART, MBTA, and MDT are resilient for most of the random malfunctions. Identifying the hub stations with extreme HS index is the key to protect and improve the robustness of the network.

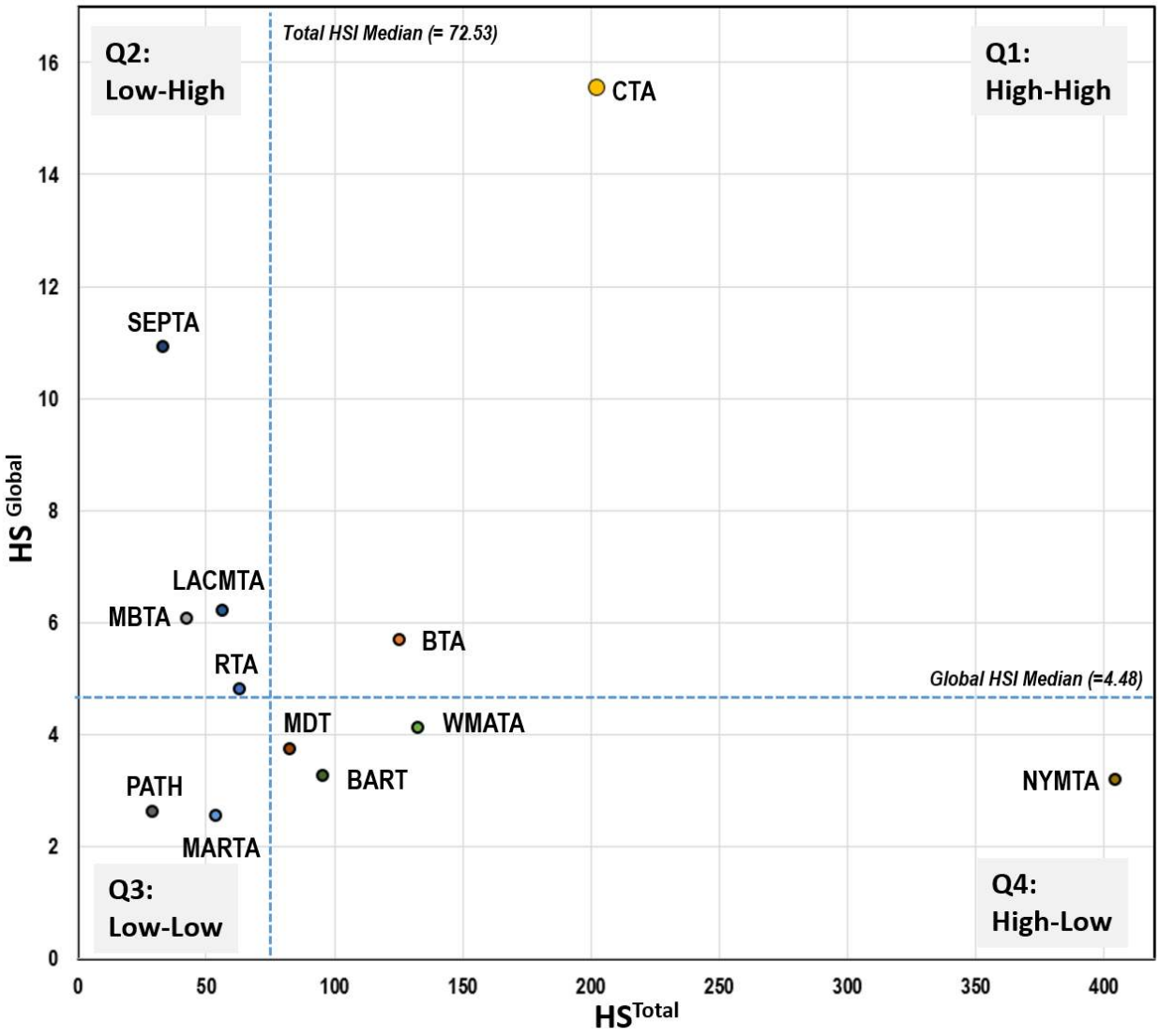


Figure 1.5: Comparison of Hub Stress index values for the twelve heavy rail networks.

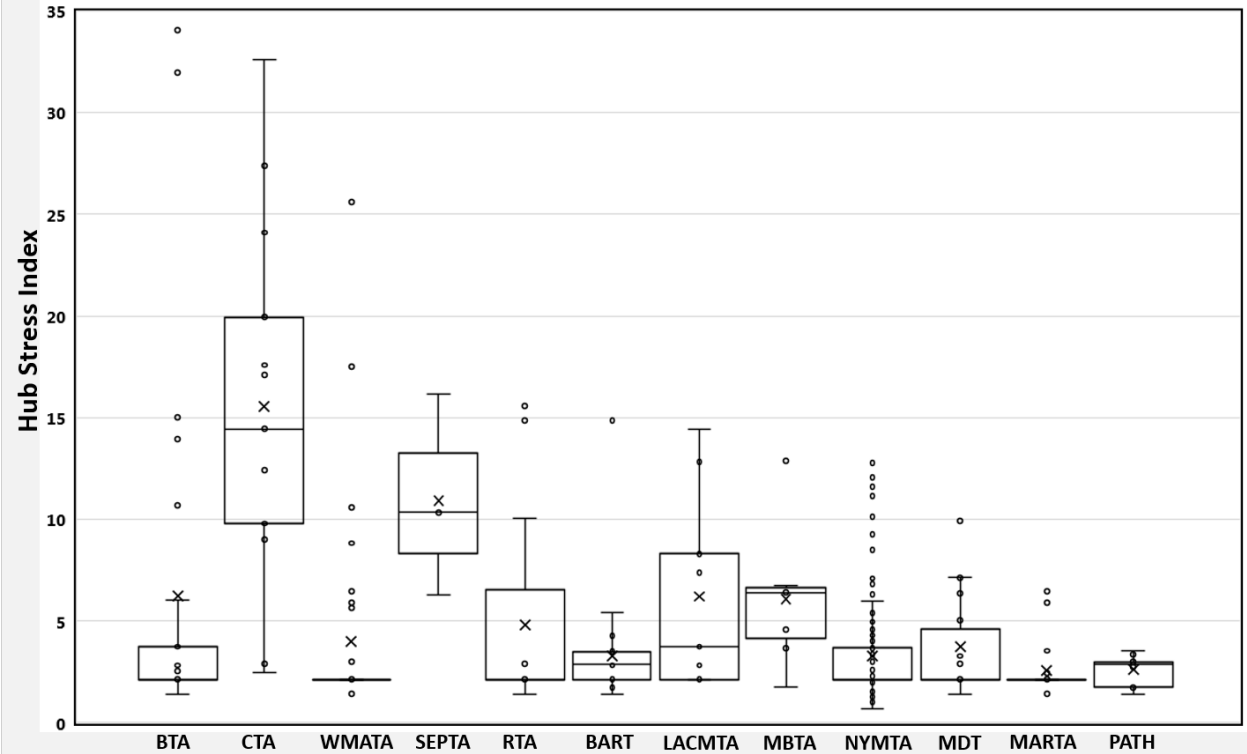


Figure 1.6: The ranges of Hub Stress index values for the hubs of each of the twelve rail networks.

Conclusion and Future Work

Traditional measures appear to not be enough to evaluate the robustness of transit networks and other multi-line systems. By redefining network components and beginning the development of new indexes we can begin to capture some of the information previously lost in these multi-line systems and identify differences in network criticality. These new measures appear to be consistent in their response and highly correlated to the existence of additional lines in a network and provide additional information about the system's tradeoff between a resilient structure and one that attempts to cover the geography of demands. These new indexes can be applied to categorizing networks and have the same range of values, both important characteristics for measuring the robustness of transportation systems.

Rail systems are recognized as the major public transit networks of US major metro areas as they provide high-speed mobility for passengers with separate rights-of-way from which all other vehicular and foot traffic are excluded. Measuring robustness and assessing the system resilience are critical concerns as the portion of service for ridership is considerably increasing in highly dense urban areas. By using our measures, we can help decision makers and transportation planners prioritize the protection or maintenance of stations, not based on a simple but monotonic traditional approach, but our comprehensive approach.

There are several avenues for future work in this area including disruption scenarios and their impact on the indexes. Extending these global measures for systems with directed graphs and circuits should also be considered. Once the robustness measures mentioned in Ellens and Kooij (2013) and established Laplacian techniques are applied to multi-line matrices and their F-matrices, results can be compared with the findings presented in this work. Another extension is developing additional local measures for robustness, centrality, and connectivity in multi-line networks using network spatial partitioning (Ji and Geroliminis, 2012). Finally, these new indexes can also provide novel ways to look at distance based matrices, resilience simulations, and longitudinal analysis.

Appendix

Eigenvalue Index Development

Since the F-matrices are positive and square, the spectrum of Forests is contained to $(0, 2T)$ according to the Gershgorin circle theorem (Varga, 2002). In order to find the magnitude of the largest eigenvalue (λ) for a set of F matrices with the same number of trails, start with the forest that has the $F_{i,j}$ and largest degree, i.e. the Forest with all trails overlaying each other ($|\mathcal{T}_k| = c \ \forall \, k$ and see Fig. 1.1 (c) for an example). This is also the most robust Forest structure and call it an uncut Forest. The structure of an uncut F-

matrix is a pseudo-Toeplitz matrix and Kulkarni et al. (1999) provided an analytic solution for the eigenvalues of such matrices

$$\lambda_i = 2T \cos\left(i\pi / \left(|\mathcal{T}_k| + 1\right)\right), \tag{1.A1}$$

where the largest eigenvalue converges to $2T$. Numerical results agree that the largest two eigenvalues (λ_M and λ_{M-1}) converge to $2T$ as the length of trails increases and this convergence is quicker for larger Forests (see Fig. 1.4 and Appendix Fig. 1.A1).

For the lower bound of the maximum eigenvalues the analysis begins with the other extreme of the Forest topological structures, is found when each trail is disjoint from the others and which contains the most subgraphs (see Fig. 1.1 (a)). These Forests are the least resilient yet cover the most nodes, or cover the most geographic space for transportation networks. Treating the F-matrix as a block matrix composed of T Toeplitz matrices on the diagonal and 0 blocks on the off diagonal, the following result from the Schur complement is used (Horn and Zhang, 2005)

$$\det(F - \lambda I) = \Pi_{k=1}^{T} \det(F(\mathcal{T}_k) - \lambda I). \tag{1.A2}$$

Therefore the eigenvalues of each $F(\mathcal{T}_k)$ are eigenvalues of the F-matrix and the solution for these Teoplitz matrices,

$$\lambda_i = 2 \cos\left(i\pi / \left(|\mathcal{T}_k| + 1\right)\right), \tag{1.A3}$$

gives a maximum eigenvalue magnitude around 2 for these disconnected Forests. Fig. 1.4 presents numerical results which support this lower bound, $\lambda_M \approx 2$.

Comparing the eigenvalues for Forests with the same number of lines but different structures is a non-trivial problem since Forests with the same number of trails can have F-matrices with different sizes and structures, but numerical results place the largest eigenvalue for all possible Forest variations, with the same number and lengths of trails,

on the line $[2, 2T]$ (see Fig. 1.4 and Fig. 1.A1). It is important to note this may not be the case for traditional adjacency matrices of different sizes. These bounds are also independent of variation in trail lengths for a Forest (see Fig. 1.4 (e)) and as the trail length increases the eigenvalues get closer to the bounds (see Fig. 1.4 (a) - (d)). This is calibrated to create the index:

$$\Lambda_F = \frac{\lambda_M - 2}{2(T-1)} \in (0, 1). \tag{1.A4}$$

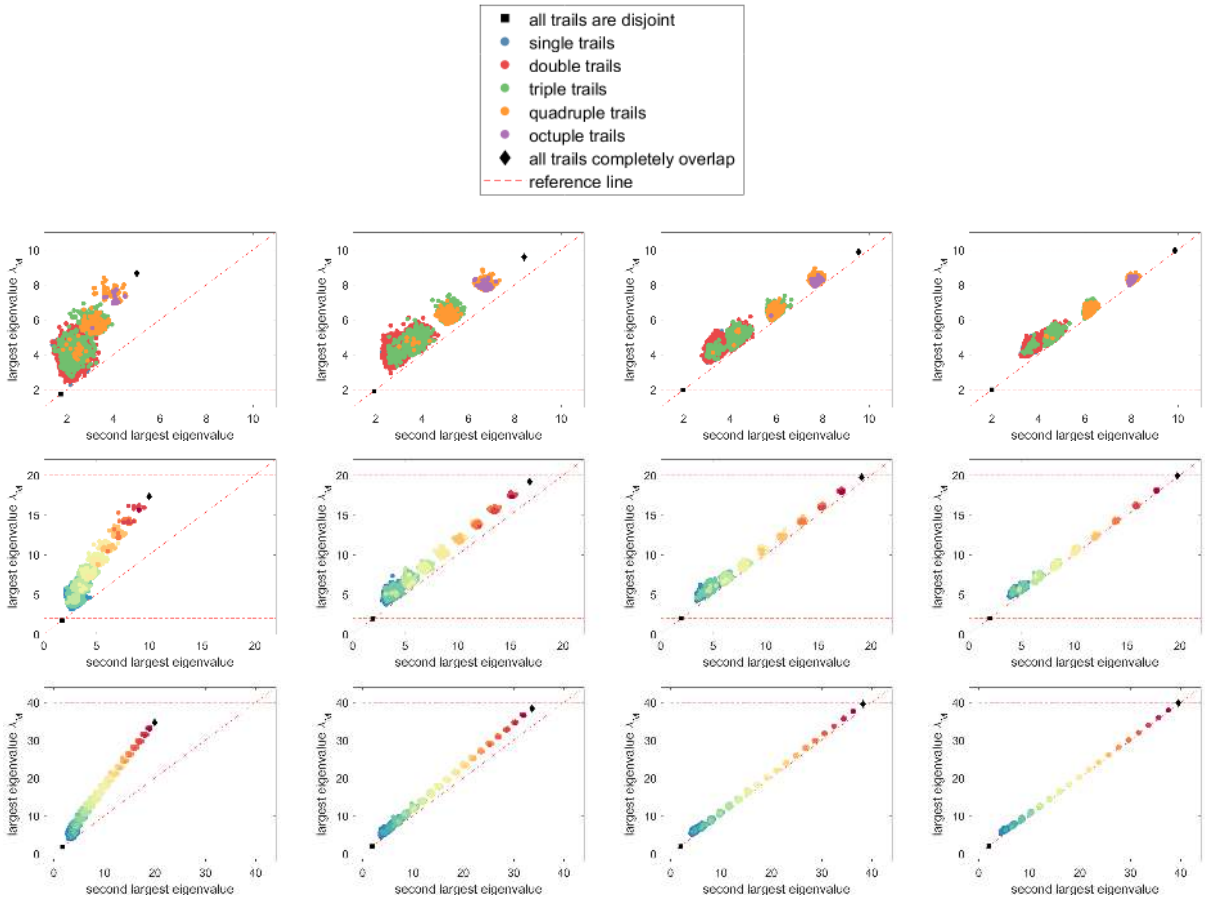


Figure 1.A1: Numerical simulations to explore the eigenvalues of multi-line networks. The plots show the largest and second largest eigenvalues for different Forest sizes and trails lengths. The first row contains the results for 5 trails, the second row for 10 trails, and the third row for 20 trails. The first column contains the results for trail length of 5, the second column for trail length of 10, the third column for trail length of 20, and the fourth column for trail length of 20. Squares represent the most covered Forest and the diamond represents the most robust Forest for the parameter combination. Forests with trail overlays, multiple trails connecting the same set of nodes, are separated with shades, the lighter shade representing more overlays. A diagonal line is given for reference and 10000 Forests were created for each parameter combination.

Chapter 2

Network Connectivity Optimization: An Evaluation of Heuristics Applied to Complex Networks and a Transportation Case Study

Abstract

Network optimization has generally been focused on solving network flow problems, but recently there have been investigations into optimizing network characteristics. Optimizing network connectivity to maximize the number of nodes within a given distance to a focal node and then minimizing the number and length of additional connections has not been as thoroughly explored, yet is important in several domains including transportation planning, telecommunications networks, and geospatial location. We compare several heuristics to explore this network connectivity optimization problem with the use of random networks, including the introduction of two planar random networks that are useful for spatial network simulation research, and a real-world case study from urban planning and public health. We observe significant variation between nodal characteristics and optimal connections across network types. This result along with the computational costs of the search for optimal solutions highlights the difficulty of finding effective heuristics. A novel genetic algorithm is proposed and we find this optimization heuristic outperforms existing techniques and describe how it can be applied to other combinatorial and dynamic problems.

Table 2.1: List of symbols and their definitions.

Symbol	Definition
ν	Network node
e	Network edge
N	Number of nodes in a given network, $N = \sum_i \nu_i$
A	Network adjacency matrix
a_{ij}	Adjacency matrix element ij
F	Focal node
$d(i,j)$	Network distance between nodes i and j
D	Threshold distance from focal node
N_C	Set of close nodes, $N_C \subset N$
N_D	Set of distant nodes, $N_D \subset N$
N'_C	Set of nodes that are now close after a new connection
L_F	Average path length to the focal node
$C(i,j)$	Cost of the new connection
$B(i,j)$	Benefit of the new connection
α	Cost weight
β	Benefit weight
t	Optimization iteration
O^t	Optimal solution for iteration t
O^*	Optimal solution
M	Set of long-term memory solutions
N_i^C	Set of neighboring close nodes for node i
N_j^D	Set of neighboring distant nodes for node j
C_i^D	Degree centrality of node i
C_i^C	Closeness centrality of node i
σ_{ij}	Shortest path between nodes i and j
$\sigma_{jk}(i)$	Shortest path between nodes j and k that includes node i
C_i^B	Betweeness centrality of node i
λ	Eigenvalue
x_i	Eigenvector
C_i^E	Eigenvector centrality of node i
α_P	Attenuation factor
C_i^P	Pagerank centrality of node i
η	Variable neighborhood size
μ	Genetic algorithm mutation rate
s	Genetic algorithm selection coefficient
P	Population of solutions for the genetic algorithm
$f(i,j)$	Genetic algorithm fitness function
ϵ_B	Benefit error from heuristic
ϵ_C	Cost deviation from heuristic
p	Connection probability (Erdös-Rényi graphs and Klemm and Eguílez graphs)
p_W	Rewiring probability (Watts-Strogatz graphs)
k_L	Initial node degree (Watts-Strogatz graphs)
m_0	Initial network size (Barabási and Albert graphs and Klemm and Eguílez graphs)
m	Degree of new nodes (Barabási and Albert graphs)
p_S	node selection probability (Klemm and Eguílez graphs)
p_R	Edge removal probability (Delaunay and Voronoi random graphs)
$\overline{C_D}$	Mean degree of a network
$\overline{L}$	Average path length of a network
c_i^w	Weighted clustering coefficient for node i
w_{ij}	Weight of connection between nodes i and j
$\overline{C}$	Weighted clustering coefficient of a network
$\overline{C_r}$	Weighted clustering coefficient of a completely random network
$\overline{L_r}$	Average path length of a completely random network
γ	Power law exponent
$P(n)$	Degree distribution
E	Efficiency of a network
E_r	Efficiency of a completely random network
E_G	Global efficiency of a network
K	Number of clusters

Introduction

Network optimization has generally been focused on solving the following classes of problems (i) finding the shortest path between nodes, (ii) maximizing the flow of information across a network, (iii) minimizing the cost of the flow of information across a network, and (iv) the problems dealing with multiple types of information flows across the network (Schrijver, 2002, Wu et al., 2004). One problem, optimizing network connectivity around a specific node with the introduction of new edges, has not been thoroughly explored and yet is important in several domains. Optimizing the network connectivity of additional edges attempts to maximize the number of nodes within a given distance to a focal node to be connected and minimizing the number and length of additional connections is essential in network layout planning for telecommunications and computer systems (Resende and Pardalos, 2006, Donoso and Fabregat, 2007), the spread of information or diseases in social networks (Eubank et al., 2004), and the development of neural networks (Whitley et al., 1990). This network connectivity problem is particularly important with transportation planning in urban environments, where the weights of the network edges can be physical distances or riderships and future street connections or transportation lines can impact flow to established facilities. For example, residential developers could optimize thoroughfare connectivity around existing schools to foster student active commuting and reduce busing costs when planning new developments (Linehan et al., 1995), and evaluating accessibility and patient travel time to health care facilities (Branas et al., 2005).

Optimization approaches have been applied to several network problems: the search for new edges that minimize the average shortest path distance in a network (Meyerson and Tagiku, 2009); the minimization of the diameter of the network, i.e. minimizing the the maximal distance between a pair of nodes (Demaine and Zadimoghaddam, 2010); and maximization of betweenness centrality (Jiang et al., 2011). However, the search for new edges, or shortcuts, that maximize connectivity to a focal node and minimize the length of

these new edges is less understood and as with the above mentioned graph optimization problems, the search for optimal solutions can become costly when networks are large and complex. This work compares a set of heuristics for this task drawn from a review of combinatorial heuristics (Mladenović et al., 2007) and from methods used for location models as this problem has many applications where space is an essential component (Brimberg and Hodgson, 2011). In this study, we show that optimization heuristics are preferred for the analysis and practice due to the nonlinearity of the solution space and the optimal soultion's dependence on nodal characteristics, such as distance to the focal node.

In connectivity optimization, network nodes are first segmented and assigned to 'close' and 'distant' sets by a specified weighted network distance from the network's focal node. An exhaustive search, where all possible edges from distant to close nodes for a network are evaluated to identify the optimal connections and as a benchmark for the time to find these solutions. This approach ensures that the optimal edges are found. However, as the number of nodes increases and therefore the number of possible connections between close and distant nodes increases, it can become computationally expensive and timely to implement. When the exhaustive search routine was applied to random networks and a real-world street networks, we also discovered that the optimal solution is nonlinearly related to nodal characteristics. To counter this, several heuristics are explored to find the optimal connection utilizing nodal characteristics and possibly in a quicker and less computationally expensive manner: hill climbing with random restart (Russell and Norvig, 2004); stochastic hill climbing (Greiner, 1992); hill climbing with a variable neighborhood search (Mladenović and Hansen, 1997); simulated annealing, which has a history of applications in graph problems (Kirkpatrick et al., 1983, Johnson et al., 1989, Kirkpatrick, 1984); and genetic algorithms, which has been successfully used for combinatorial optimization (Anderson and Ferris, 1994). A Tabu heuristic was not employed as it has been observed to not be an effective method for multi-objective optimization problems, whereas

simulated annealing and genetic algorithms have shown to be effective (Golden and Skiscim, 1986, Kim et al., 2016b). Among these methods, the genetic algorithm presented here introduces a novel chromosome formulation where the genes are not properties of a specific variable but weights for the probability to move in a given direction in the solution space. This allows the method to dynamically change what solution characteristics to explore while possibly reducing the size of the local neighborhood search.

These optimization heuristics are then applied to randomly generated networks that vary in complexity and size to evauluate their efficacy in finding the optimal connection. Several types of random graph networks were generated to analyze the efficacy of the optimization heuristics for systems with different topologies which are generally representative of naturally occuring and built systems: (1) Erdös-Rényi networks, (2) Watts-Strogatz networks, (3) Barabási and Albert networks, (4) Klemm and Eguílez networks, (5) Delaunay triangulation networks, and (6) Voronoi diagrams. Erdös-Rényi random networks are constructed by randomly creating connection between pairs of nodes with a probability (Erdös and Rényi, 1959). These networks, even though they have random connections, consistently have short average path lengths and irregular connections, both of which are well found in natural systems. The Watts-Strogatz networks also have random connections but the networks also form clusters, another feature commonly found in real-world networks (Watts and Strogatz, 1998). The Barabási-Albert model produces random structures with a small number of highly connected nodes, 'hubs', which are observed in numerous types of networks (Barabási and Albert, 1999, Albert and Barabási, 2002). Klemm and Eguílez networks have random connections, clusters, and hubs (Klemm and Eguílez, 2002).

We also introduce two novel types of random planar network versions of Voronoi diagrams and Delaunay triangulations. The reasons these were considered was that planarity is particularly important in many fields and networks generated from Voronoi diagrams and Delaunay triangles have been used in spatial health epidemiology (John-

son, 2007), transportation flow problems (Steffen and Seyfried, 2010, Pablo-Martì and Sánchez, 2017), terrain surface modeling (Floriani et al., 1985), telecommunications (Meguerdichian et al., 2001), computer networks design (Liebeherr and Nahas, 2001), hazard avoidance systems in autonomous vehicles (Anderson et al., 2012). Delaunay triangulation maximizes the minimum angles between three nodes, to generate planar graphs with consistent network characteristics (Delaunay, 1934). Voronoi diagrams, the dual of a Delaunay triangulation, are composed of points and cells such that each cell is closer to its point than any other point. When edges are randomly removed from the connected Delaunay network or Voronoi network, with weights given by node distance from a focal node, we show that these networks display some of the properties similarly found in the networks mentioned above, such as complexity and randomness, but with the added component of being planar and having edge weights that can be framed as physical distances.

To complement the random network analysis, the network connectivity optimization methods are applied to a study of urban transportation planning. We use the network connectivity optimization methods to evaluate the potential costs and benefits of increased thoroughfare connectivity for student active commuting to school. It is assumed that expanding this connectivity around a school would allow for more households, and students, to be included within the walking distance to the school. If more students actively commute to school, this reduces the busing costs for the school system and increases the health and academic achievement of the students (Centers for Disease Control and Prevention, 2010). The combinatorial optimization techniques employed here to identify and evaluate new street connections can complement the optimization approaches used for other transportation planning problems, such as greenway planning (Linehan et al., 1995), bus stop locations (Ibeas et al., 2010, Delmelle et al., 2012), and health care accessibility (Gu et al., 2010).

The following section describes the formulation of the connectivity problem in more

detail, the local search methodology, and the optimization heuristics (see Appendix A for the specific pseudocode of the optimization algorithms). The section also details the data used for the study including descriptions of the random networks and the street networks around schools that are used for the transportation study. Results of the heuristics applied to both the random networks and the case study data are also presented in that section. This is followed by a detailed discussion of the heuristic results, the further implications of these techniques for urban transportation planning, and future work for this avenue of research.

Methods and Results

Network Connectivity Optimization

For the description of the optimization methodology the following nomenclature will be used. The number of nodes is N, and nodes are separated into two sets based on their shortest network paths, $d(i,j)$, where i and j are nodes, to the focal node F. The nodes that are outside the path distance under consideration, D, are assigned to the 'distant' set, $N_D \subset N$, i.e. $i \in N_D$ if $d(i,F) > D$. The nodes that are within this distance are assigned to the 'close' set, $N_C \subset N$, i.e. $i \in N_C$ if $d(i,F) \leq D$ and $F \in N_C$ (see Figure 2.1 (A)). Node neighborhoods are assigned to the sets N_i^D and N_j^C for the distant and close nodes of i and j, respectively by the nodal characteristics described in the following subsection. For a network of size N the number of new connections to evaluate is $\leq N^2/4$.

When a new connection is evaluated, any distant nodes that are now within the path distance D to the focal node are assigned to the new set N'_C. For example, if a new connection is established between distant node i and close node j, then $k \in N'_C$ if $k \in N_D$ and $d(k,i) + d(i,j) + d(j,F) \leq D$ (see Figure 2.1 (C)). The number of nodes in N'_C set is considered the benefit of this new connection, $B(i,j) = |N'_C|$. The cost of the new connection is denoted by $C(i,j)$ and for simplicity and this analysis $C(i,j) = d(i,j)$. The

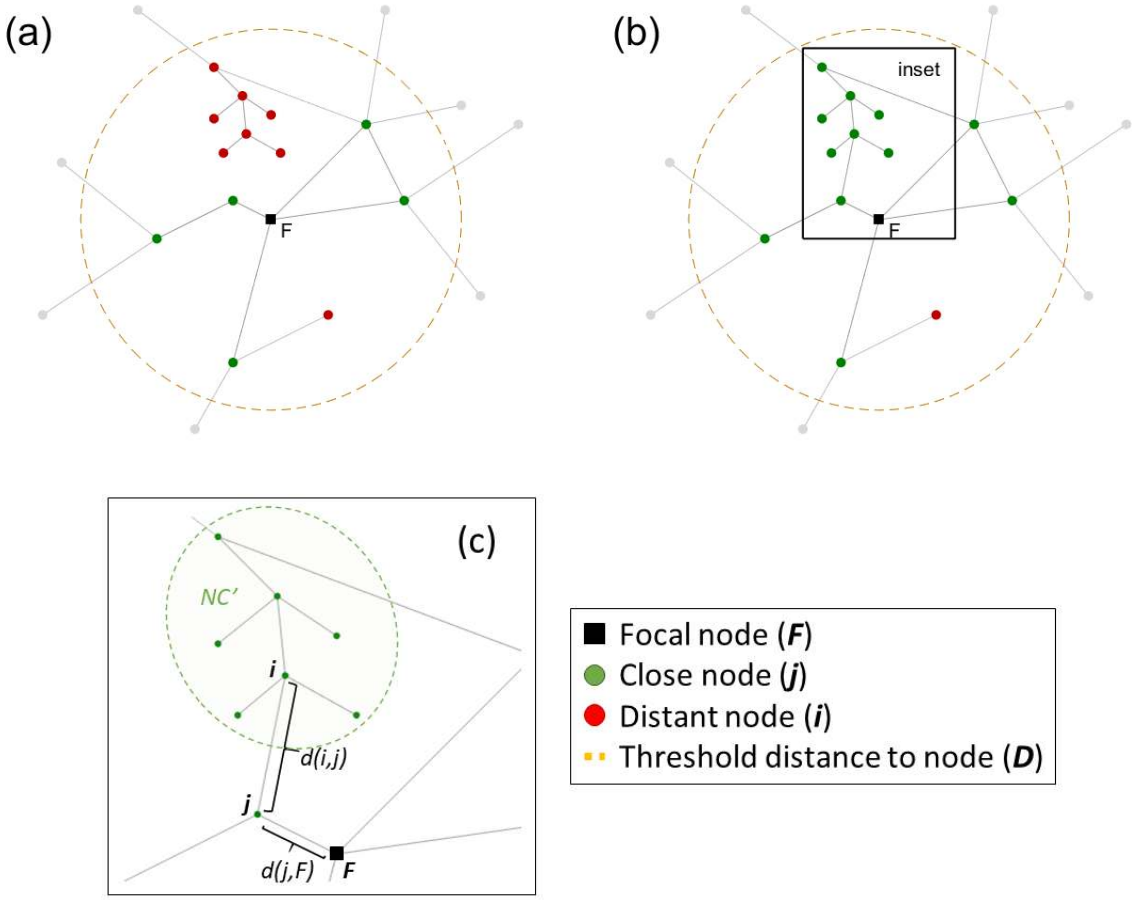


Figure 2.1: Diagram of the network connectivity optimization problem. The close nodes that are within a threshold network distance (orange dashed circle) from the focal node (black square) are colored green, distant nodes that could be within the threshold network distance with additional edges are colored red, and distant nodes that could not be within this distance regardless of additional edges are gray. Figure (a) is an example graph, (b) shows the same graph with the optimal new connection that maximizes the number of additional nodes within the threshold network distance and minimizes the length of the new connection, and the inset (c) highlights this optimal connection, between nodes i and j, with the methodological terminology presented in Section 2.

optimal solution is the solution with the greatest benefit, or number of new nodes now within the distance to the focal node which can be expressed as the bi-objective function

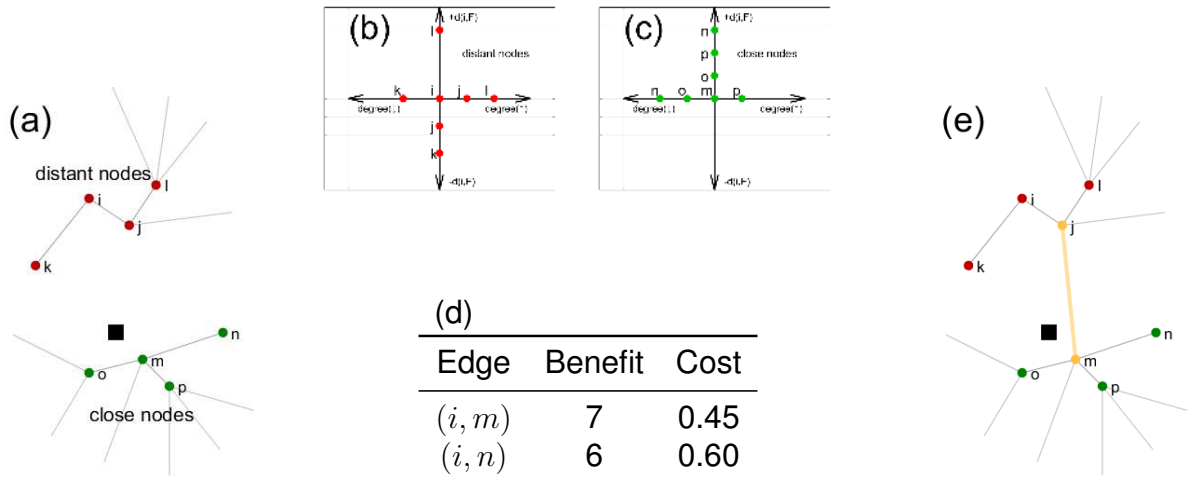
$$O^* = \max_{(i,j)}(\alpha B(i,j) + \beta C(i,j)),$$

where α and β are the weights for the benefits and costs, respectively and for this study $\alpha = 1$ and $\beta = \infty$. If $\beta = \infty$, then the objective is only to minimize costs for the same benefit. Some of the heuristics are also dependent on the number of iterations (t) and terminate when the solutions converge, $O^t = O^{t-1}$, or the solution does not improve, $O^{t-1} > O^t$.

Local Search Methodology

The selection of neighboring nodes to improve solutions begin with on several evident nodal characteristics (see Figure 2.2 and Appendix Table 2.A2). These nodal characteristics are explored to find the critical network properties for connectivity optimization and their impact on the performance in finding the optimal solution. Nodes are ranked by these characteristics and this creates a multidimensional solution space. A two-level selection process is used, with the following nodal level characteristics: (i) distance to the focal node, (ii) degree centrality, (iii) closeness centrality, (iv) betweenness centrality, (v) eigenvector centrality, (vi) pagerank centrality, (vii) weighted clustering coefficient; and the the following clustering of the characteristics: (i) hierarchical clusters, (ii) network-constrained clusters, and (iii) network modularity. Multiple nodes in a network can have the same degree or assigned to the same cluster, therefore the local searches include a random shuffling routine to evaluate nodes with the same values.

Nodes are ranked by their distance to the focal node and moving in this solution dimension may result in lower connectivity length costs but may not maximize the number of nodes ultimately connected to the focal node. Ranking and selecting nodes by their



(d)

Edge	Benefit	Cost
(i, m)	7	0.45
(i, n)	6	0.60
⋮		
(j, m)	8	0.35
⋮		
(ℓ, p)	5	0.85

Figure 2.2: Diagram of the local search methodology. Figure (a) shows a generated network with the focal node represented by a black square, the close nodes are colored green and the distant nodes red. Figure (b) shows part of the search space neighborhood for distant node i by increasing and decreasing degree and distance to the focal node, $d(i, F)$ (the other nodal characteristics and clusters are not shown for simplicity). Figure (c) similarly represents the space for the close node m. The number of neighbors selected at each iteration of the optimization routine is heuristic dependent. Table (d) gives the costs and benefits for selected connections from the local search and the optimal connection (j, m) for this iteration is shown in (e) an an orange edge.

centrality, i.e. the importance of the node, could result in maximizing the number of nodes within the specified distance to the focal node but with the possibility of higher connectivity length costs compared to selecting nodes by other characteristics. Several commonly used measures of centrality are explore: degree, the number of edges incident to a node; closeness centrality, the average length of the shortest path between the node and all other nodes in the network (Bavelas, 1950); betweenness centrality, the frequency of a node included in the shortest paths between all other node pairs (Freeman, 1977); eigenvector centrality, which is a relative ranking of nodes such that nodes with high values are connected to other nodes with high values (Newman, 2008); and pagerank centrality, a

variant of eigenvector centrality that ranks nodes based on their probability of being connected to a randomly selected node and which is commonly used in web-page rankings (Brin and Page, 1998).

Nodes are clustered using weighted clustering coefficients, hierarchical clustering, network-constrained clustering, and network modularity. The weighted clustering coefficient of a node is the count of the triplets in the neighborhood of the node and accounts for the weights of the edges times the maximum possible number of triplets that could occur (Barrat et al., 2004). Nodes are also clustered by their characteristics with hierarchical clustering utilizing Ward's method and the gap criterion. Hierarchical clustering with Ward's method attempts to minimize variance within clusters and maximize variance between clusters (Ward, 1963). The gap criteria is used to identify the optimal number of hierarchical clusters by maximizing the distance between the within-cluster variation and the expected within-cluster variation found from bootstrapping (Tibshirani et al., 2001). The network-constrained clustering method utilizes the shortest paths between nodes and thereby capturing the network neighbors of each node (Yamada and Thill, 2006). Network modularity attempts to cluster nodes by maximizing the number of connections within a cluster and minimizing the number of connections between the clusters. Network modularity accomplishes this by comparing the probability that an edge is in a cluster with the probability a random edge is in the module, i.e. an edge is present in a random graph with the same node degree distribution (Newman, 2006).

Network Connectivity Optimization Heuristics

The following techniques were selected for the network connectivity optimization study from their extensive use in optimization (see the Appendix for the algorithms). Parameter selection was simplified for easy comparison of the methods. Random restart, randomly selecting initial nodes to avoid local optima and running the routine until the optimal solution is found, was used for each method to ensure the methods did not converge on

suboptimal solutions due to the initial starting values. Six heuristics are employed as below.

Exhaustive Search (ES). The exhaustive search optimization routine creates an edge for every combination of distant and close nodes (see Algorithm 1 in the Appendix). Because the results by ES is the optimal, the solution times and the objective values are used to benchmark the solutions by other methods.

Hill Climbing (HC). The solution space was observed to be hilly from the exhaustive search results, so several modifications were introduced to the hill climbing technique to address this (Algorithm 2). A stochastic hill climbing **(HCS)**, an advanced search method based on HC, routine is also explored where the selection of nodes for the next iteration is randomly picked with

$$\text{probability}(i,j) = \frac{\alpha C(i,j) + \beta B(i,j)}{\sum_{(m,n)}(\alpha C(m,n) + \beta B(m,n))},$$

which terminates when a better solution is no longer found (Algorithm 3). A hill climbing algorithm is coupled with a variable neighborhood **(HCVN)** where the size of the neighborhood starts with the nearest neighbors ($\eta = 1$) and is updated as follows:

$$\eta = \begin{cases} 1 & \text{if } O^t > O^{t-1} \\ \eta + 1 & \text{if } O^t \leq O^{t-1} \end{cases},$$

and the HCVN method terminates after n_{max} is reached (Algorithm 4).

Simulated Annealing (SA). As a meta-heuristic approach, the simulated annealing method randomly selects an initial solution from the solution space to avoid entrapment in a local optima. At each iteration, the heuristic evaluates the neighboring solutions and if it does not find an improved solution, it moves to a new solution with the following probability

$$\text{probability}(i,j) = \exp\left(-\frac{O^{t-1} - O(i,j)}{t}\right),$$

to obtain an improvement of the solution. The distance of the move decreases with the number of iterations until a better solution is no longer found (Algorithm 6).

Genetic Algorithm (GA). The genetic algorithm begins with a population of P randomly selected solutions with a set of chromosomes composed of genes which represent the weights of selecting a neighbor and are all initialized to unity (Algorithm 7). During each iteration of the method, solution scores (fitnesses) are computed by

$$f(i,j) = \frac{O(i,j)}{\sum_{(m,n)} O(m,n)},$$

and a new generation of solutions are selected based on the following probability condition

$$\text{probability}(i,j) = \frac{s * f(i,j) + (1-s)}{\sum_{(m,n)}(s * f(m,n) + (1-s))},$$

where s is the selection coefficient. Weak selection, $s \ll 1$, is used to ensure the that random mutations impact solution frequency. Crossover is conducted by alternating the weights for the offspring from each parent, also known as cycle crossover (Oliver et al., 1987). Mutations are introduced at a low rate $\mu \ll 1$ for each gene and increase the nodal characteristic or cluster neighbor selection weight by one. The probability that characteristic or cluster m is used to find a neighbor for node i is given by

$$\text{probability(characteristic or cluster)} = \frac{\text{gene}(i,m)}{\sum_k \text{gene}(i,k)/K},$$

where K is the total number of nodal characteristics and clusters. This formulation ensures that the nodal characteristics or clusters that improve the solution increase in weight, results in a greater probability they will be selected for neighborhood exploration, and reduces the size of the neighborhood search.

Simulated Data: Complex Random Networks

Complex Random Networks. Several types of random networks were used to evaluate the effectiveness of each hueristic in identifying the optimal new connection. The following types of random undirected networks were generate: (1) Erdös-Rényi (**ER**) networks, (2) Watts-Strogatz (**WS**) networks, (3) Barabási and Albert (**BA**) networks, (4) Klemm and Eguílez (**KE**) networks (see Appendix Figure 2.A1 (A) – (D) and for the algorithms used to generate the networks see Prettejohn et al. (2011)).

Complex Planar Networks. Two novel types of random networks are created here, random Delaunay triangulation (**DT**) (Appendix Figure 2.A1 (E)) and random Voronoi diagrams (**VD**) (Appendix Figure 2.A1 (F)). These networks are inherently planar and edges are removed from network nodes randomly based on their distance from the focal node with probability

$$p_R \cdot \frac{\max(d(i,F), d(j,F))}{\max_k d(k,F)},$$

where p_R is the removal probability and weighted by the normalized edge distance from the focal node.

Parameter Selection. To compare the efficacy of different optimization methods for different network topologies, identifying the best set of parameters are critical. Parameters values were selected for each type of random network to ensure network complexity (Appendix Table 2.A5 summarizes the parameters which were used in the analysis). Variation in network size was also explored and the most connected node in each network was selected as the focal node. Uniformly randomly generated edge weights in [0,1] were used for the network distances and the threshold distance was set to ensure that half of the nodes were initially within the distance to the focal node. The costs and benefits were normalized using the ranges from the exhaustive search routine as a benchmark to compare the results from the different optimization methods.

Empirical Data: Street Networks Around Schools

Networks composed of street edges and residence nodes around several schools from a US school system were used for the analysis. Ten suburban and rural schools from Knox County, TN, were selected for the analysis, including seven elementary and three middle, that would benefit the most from increased thoroughfare connectivity, i.e. had the most students within the Euclidean walking distance but not the network distance to the school. Urban schools were not used since the street connectivity around the schools was significantly high and the from additional thoroughfares would be low. Residential nodes were placed on the street networks. The residences within 1 mile and 1.5 miles, for the elementary schools and the middle schools respectively, are considered close nodes while the nodes outside of these distances were classified as distant nodes (see Figure 2.3). The school networks do not generally display the characteristics of complex network, they had low average degree, large path lengths, and were not efficient, yet have a few intersections (nodes) with a large number of street connections (see Appendix Table 2.A4). The networks were evaluated with each optimization method to maximize the number or close residences connected to the school and minimize the distance of the new thoroughfares. The costs and benefits of these street connections were normalized using the ranges from the exhaustive search routine as a benchmark to compare the results from the different optimization methods.

Results

Several finding are worthy to note. First, there were consistent nonlinear relationships between the nodal characteristics and the quality of the solutions for each type of random network and the school networks (see Figure 2.4). Nonlinear Pareto frontiers were also observed as was significant variation for which nodal characteristics were correlated with the quality of the solution across networks (see Table 2.1). Among those, the distance

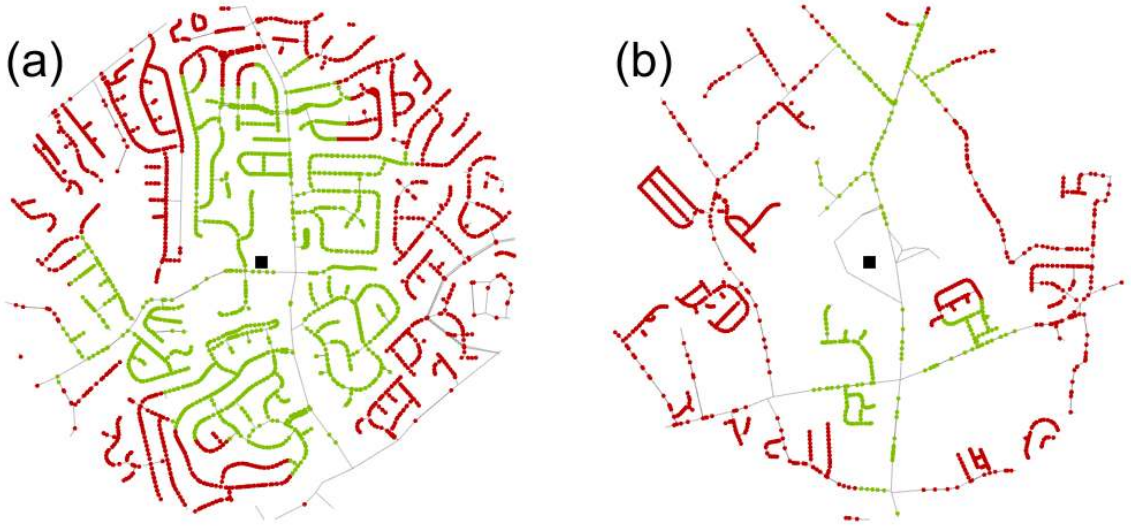


Figure 2.3: Examples of the street networks used for the analysis: (a) a suburban elementary school and (b) a rural elementary school. The red nodes represent the distant residences, i.e. the residences within the 1-mile Euclidean walking distance but not the 1-mile network walking distance, the green nodes are the close residences within the network walking distance, and the black square represents the school.

between the close node and the focal node and the distance between the distant node and the close node were most often highly correlated with the quality of the solution across networks. The centrality measures were inconsistently related to the solution quality for the random networks. The clustering methods were consistently unrelated to the quality of the solutions for the random networks, while the network modularity for the distant node was correlated with spatial networks and the school networks.

Results of the termination times and the optimal solutions deviations from the optimization heuristics applied to the random networks are summarized in Figure 2.5 and Appendix Figures 2.A2 and 2.A3. The hill climbing method was consistently faster for all of the networks, yet had the largest cost and benefit deviations. Simulated annealing and the genetic algorithm had similar termination times, but the genetic algorithm was consistently superior to all of the other methods in approaching the optimal solution. The results from the application of the optimization heuristics applied to the ten school networks are shown in Figure 2.5. The times to termination for each hueristic according to network size

consistently followed the following pattern: ES>SA>HCVN>GA>HCS>HC and network topology played a significant role in these times (e.g. the dips in 2.5 (e)). The genetic algorithm clearly outperformed the other heuristics, followed by simulated annealing, in terms of cost and benefit deviations (see Figure 2.5 (b), (d), and (f)).

Discussion

The network connectivity problem introduced in this study is relevant to a wide range of applications and is nontrivial as the number of solutions can become large even for small systems. This type of combinatorial optimization problem highlights the difficulty in determining local search routines a priori. The nodal characteristics were nonlinearly related to the solutions while different characteristics varied in their correlation with solution quality for different networks making it difficult to exclude specific characteristics for network connectivity optimization. Distance to the focal node was consistently related to the quality of the solution as this lowers connectivity length costs, while centrality was intermittently correlated with solution quality it provides greater benefit through more connections. Clustering nodal characteristics did not provide additional useful information from the nodal characteristics for the random networks. This could arise from the following issues: the curse of dimensionality, i.e. large sparse subspaces in the solution space; the nodal characteristics are highly correlated with each other; outliers; finding the appropriate influential nodal characteristics is not possible a priori; and the influence of specific characteristics is dynamic as the heuristics converge. For the school networks: the clustering coefficient was a poor measure due to the lack of triplets in the networks; the network modularity also had poor results possibly due to the measure's inability to account for the spatial component of the nodes; and the network-constrained clusters were also poor in explaining the solutions, due to the complexity of the network topology.

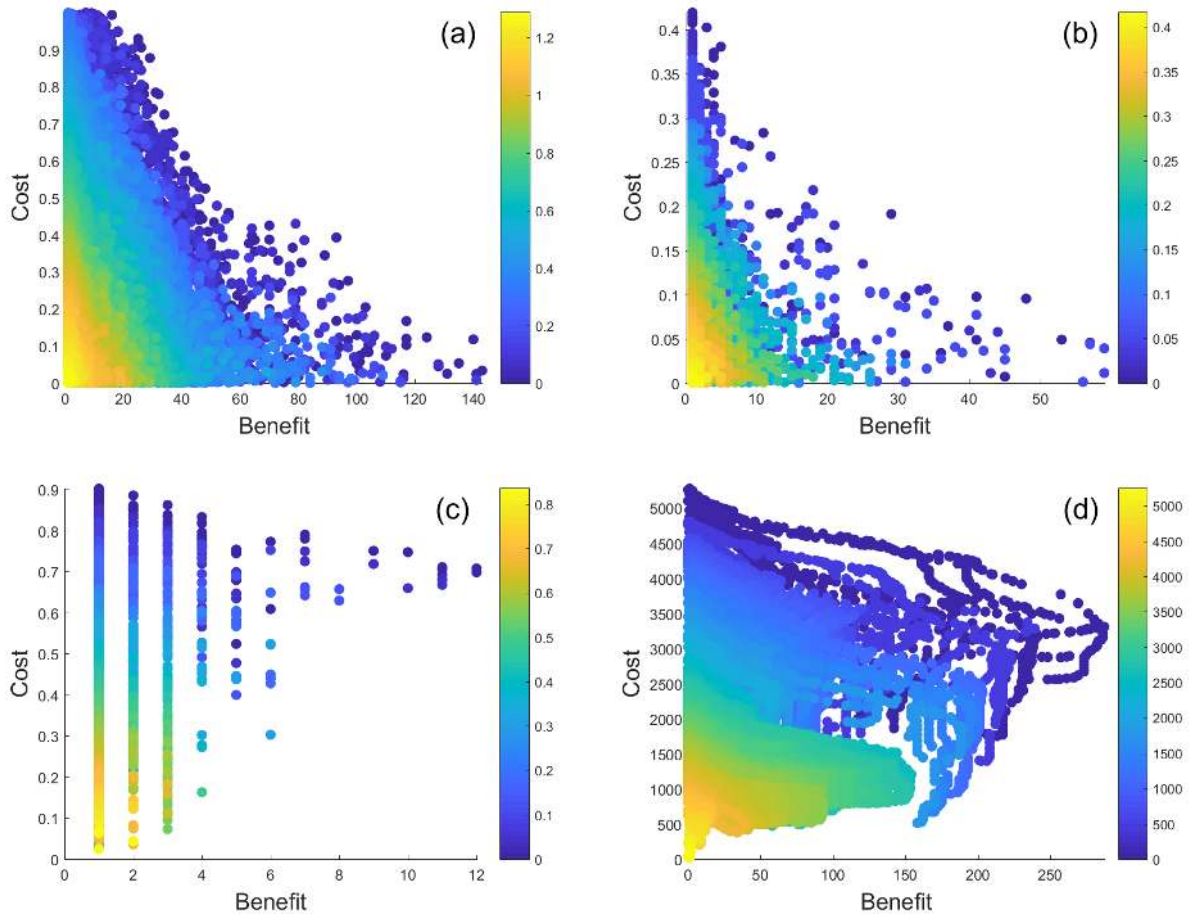


Figure 2.4: The relationship between the distances of the close node and the focal node with the costs and benefits for each solution for different networks. Figure (a) shows the relationship for a Watts-Strogatz network with $N = 500$, (b) a Barabási and Albert network with $N = 500$, (c) a Delaunay network with $N = 500$, and (d) a suburban school network ($N \approx 4000$). Each point represents a connection between a distant and close node, where the cost is the length of the connection and the benefit is the number of new nodes within the distance to the focal node or school.

Table 2.1: Average correlation coefficients for the random networks (1000 graphs for each network type with $N = 1000$) and the ten school networks. The three coefficients with the largest magnitude are highlighted for each network type. There was no variation in clustering coefficients as triplets were not common in the street networks.

		Network						
		ER	WS	BA	KS	DT	VD	Schools
Nodal characteristics	close node							
	$d(j, F)$	-0.08	-0.52	-0.37	-0.48	-0.28	-0.41	-0.30
	C_j^D	0.14	0.08	0.05	0.09	-0.01	0.02	0.02
	C_j^C	0.10	0.26	0.05	0.15	0.15	0.15	0.18
	C_j^B	0.14	0.15	0.05	0.04	-0.02	0.03	0.06
	C_j^E	0.12	0.11	0.05	0.09	0.01	0.08	-0.01
	C_j^P	0.14	0.07	0.05	0.10	-0.03	-0.01	-0.00
	c_j^w	0.00	-0.03	0.02	0.00	0.14	0.03	*
	HC_j	0.01	-0.01	-0.02	-0.01	0.03	0.00	-0.00
	NM_j	-0.01	0.03	0.07	0.06	-0.06	0.02	0.22
	distant node							
	$d(i, F)$	-0.46	-0.18	-0.12	-0.09	-0.07	-0.08	-0.04
	C_i^D	0.17	0.03	0.08	0.05	0.29	0.30	0.08
	C_i^C	0.15	-0.01	0.07	0.00	0.35	0.21	0.06
	C_i^B	0.19	0.03	0.07	0.03	0.20	0.20	0.07
	C_i^E	0.21	0.02	0.07	0.06	-0.16	-0.01	-0.03
	C_i^P	0.17	0.01	0.08	0.07	0.30	0.22	0.01
	c_i^w	0.08	-0.01	0.03	0.03	0.12	0.12	*
	HC_i	0.00	0.00	0.02	0.00	0.05	-0.03	-0.01
	NM_i	0.12	-0.07	0.02	-0.01	-0.30	-0.19	0.17

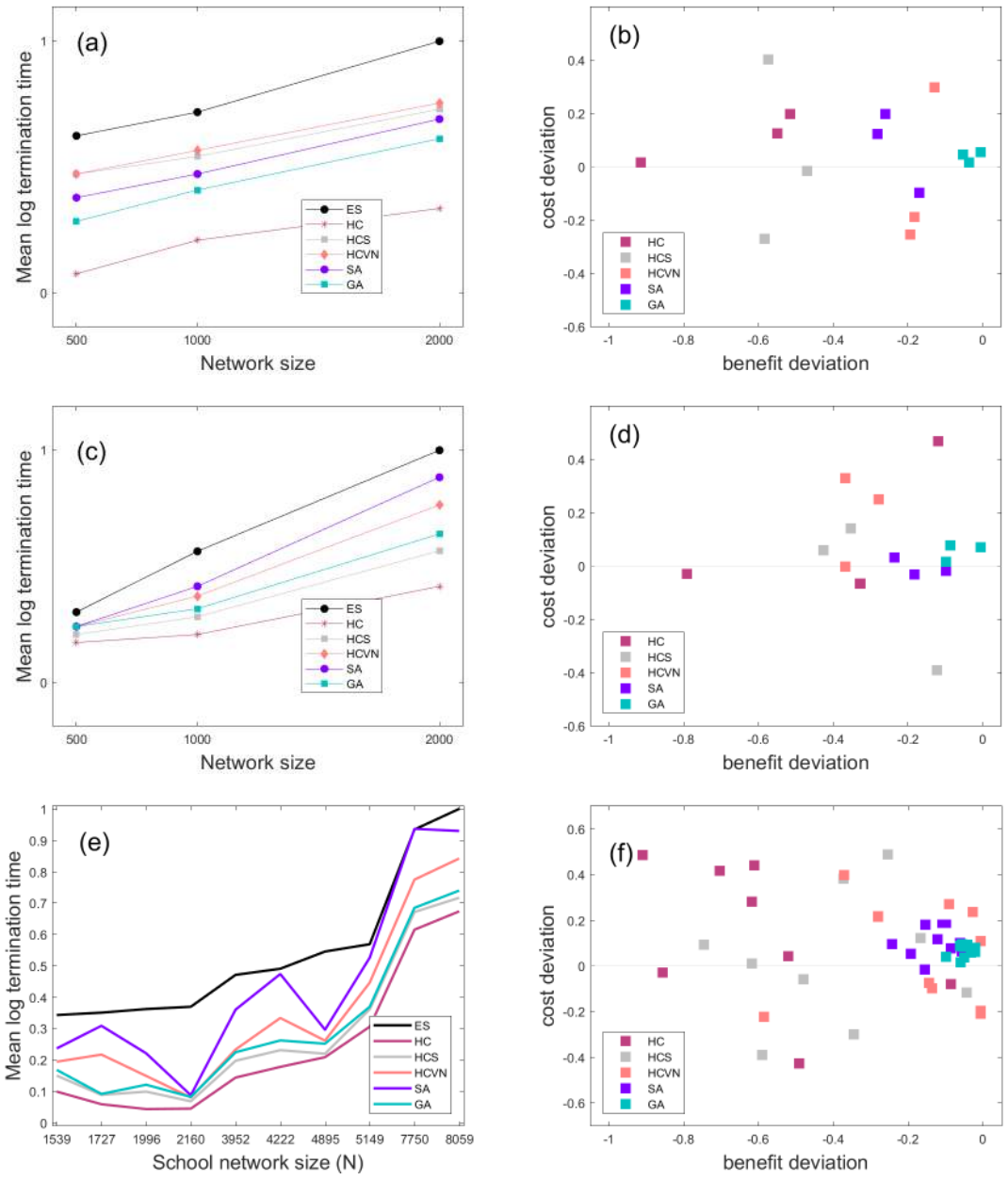


Figure 2.5: The termination times for the heuristics applied to the random networks and school networks: (a) and (b) Erdös-Rényi networks, (c) and (d) Delaunay networks, and (e) and (f) the ten school networks. These are average times for 1000 random restarts for optimization applied to 1000 random network of each type and size. The times are scaled by the exhaustive search time and log transformed for easier interpretation. The benefits were scaled by the results from the exhaustive search, where a longer connection length is a positive cost deviation and a shorter connection is a negative cost deviation.

The optimization heuristics save computational time but vary considerably in their ability to find a solution near the optimal. The stochastic hill climbing search was not effective due to the large neighborhood search space explored. In our experiment, the number of solutions checked at each iteration is > 300 and resulted in a skewed probability distribution of objective values favoring the selection of low values. This degraded the efficiency of the method resulting in the selection of poor solutions. The variable neighborhood search method was similarly not reliable because of the significantly large neighborhood search space (the number of possible solutions explored at a given iteration could be $> 5,000$), and had intermediate results with cost and benefit deviations. The simulated annealing heuristic consistently took longer to converge than the other optimization methods from the exploration of suboptimal solutions prior to moving towards better solutions, yet it was able to converge to values close to the optimal solution.

The computational costs and the variance in the importance of nodal characteristics for the random networks and real-world systems highlights the need for a heuristic that is able to quickly and effectively explore the solution space. The genetic algorithm provided in this work offers a solution to this issue, exploring the nonlinear Pareto frontier and outperformed the other algorithms in terms of the consistently higher solution precision and accuracy. The genetic algorithm is able to dynamically reduce the size of the neighborhood search space and what variables to analyze. This reduction in the local solution search space allows the genetic algorithm presented here to converge on solutions near the optimal in a timely fashion. This shows the power of biologically inspired algorithms to effectively explore multidimensional spaces (commonly found in natural systems) and their potential use in a wide variety of disciplines, including specific applications for planning and health care.

Application of these methods and heuristics to multi-level networks, such as telecommunication systems, higher dimensional real-world networks (transportation networks with elevation), directed networks, and additional planar random networks (e.g. Gabriel

graphs) should be conducted. Different distance measures to the focal node, such as the Hamming distance, could also be evaluated for different applications, and other real world examples should be used for analysis. The methods presented here do not evaluate whether the new connections intersect existing edges and attempts to incorporate such a feature resulted in unrealistic computational times. Optimizing this feature is currently being developed as is a tool for ArcGIS and Python for planners and researchers to utilize.

Appendix

Additional Figures and Tables

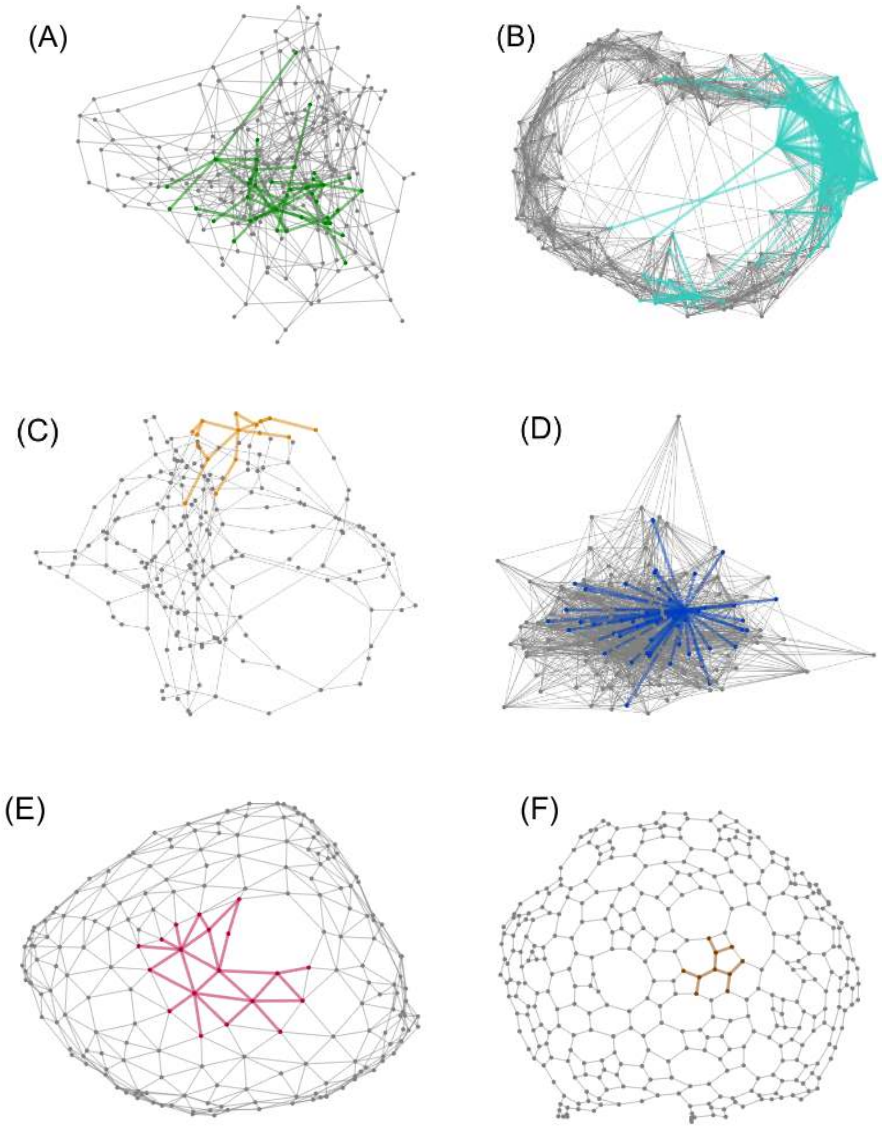


Figure 2.A1: Examples of the random graphs ($N = 250$) used for the analysis: (A) Erdös-Rényi network, (B) Watts-Strogatz network, (C) Barabási and Albert network, and (D) Klemm and Eguílez network, (E) Delaunay network, and (F) Voronoi network. The colored subgraph represents the close nodes to a random focal node with high connectivity and the edge weights (distances) are not to scale for a clearer visualization.

Table 2.A1: Node characteristics used for the neighborhood search and their formulations.

	Local search selection criteria	
Nodal characteristics	distance from focal node $d(i, F)$	degree centrality $C_i^D = \sum_j A(i,j)$
	closeness centrality $C_i^C = 1/\sum_j d(i,j)$	betweenness centrality $C_i^B = \sum_{j \neq i \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}}$
	eigenvector centrality $C_i^E = \frac{1}{\lambda} \sum_j A(i,j) x_j$	pagerank centrality $C_i^P = \alpha_P \sum_j A(i,j) \frac{x_j}{\sum_i A(i,j)} + \frac{1-\alpha_P}{N}$
	weighted clustering coefficient $c_i^w = \frac{1}{(C_i^D - 1)\sum_j a_{ij} w_{ij}} \sum_{j,h} \frac{(w_{ij} + w_{ih})}{2} a_{ij} a_{ih} a_{jh}$	
Clusters	hierarchical clusters network-constrained clusters network modularity	

Measures of Network Complexity. To evaluate the complexity of the random Voronoi networks and random Delaunay networks we use the following measures that capture the random, scale-free, and small-world features of complex networks: average degree, average path length, weighted clustering coefficient, proximity ratio, global efficiency, and power law (see Table 2.A2 for the mathematical formulations of these characteristics). General measures of connectivity and network topology (for connected graphs) include the mean degree and the average path length, which refers to the average number of edges within the shortest path for all pairs of nodes in a network (Albert and Barabási, 2002). The weighted clustering coefficient of a node is the ratio of the node degree and the total number of possible edges for a node in the network (Luce and Perry, 1949). The global version of this measure is the average of the node weighted clustering coefficients and provides an estimate of small-world-ness (Watts and Strogatz, 1998). The proximity ratio is the ratio of the following ratios: (i) the average weighted clustering coefficient and the average path length and (ii) the average weighted clustering coefficient for a

completely random network of the same size and the average path length for that random network of the same size. This provides a measure of the small-world-ness of networks, with $S = 1$ for random networks and $S \gg 1$ for small-world networks (Walsh, 1999). Another measure of small-world-ness is average efficiency which is the average of the inverses of the network's shortest paths and captures the network's ability to exchange information between nodes (Latora and Marchiori, 2001). If the degree distribution of the nodes follows a power law, then the network is said to be scale-free, i.e. random with some highly connected nodes (Barabási and Albert, 1999).

Table 2.A2: The measures used to evaluate network complexity.

Measures of network complexity	
mean degree $\overline{C_D} = \frac{\sum_i C_i^D}{N}$	average path length $L = \frac{1}{N(N-1)} \sum_i \sum_{j>i} d(i,j)$
weighted clustering coefficient $\overline{C} = \frac{1}{N} \sum_i c_i^w$	proximity ratio $S = \frac{\overline{C}/L}{\overline{C_r}/L_r}$
power law $P(n) \approx n^{-\gamma}$	global efficiency $E_G = \frac{E}{E_r}, E = \frac{1}{N(N-1)} \sum_i \sum_{j>i} \frac{1}{d(i,j)}$

Table 2.A3: The average characteristics of 1000 random Delaunay and Voronoi networks.

	Delaunay			Voronoi		
N	500	1000	2000	500	1000	2000
$\overline{C_D}$	4.708	5.392	5.208	2.840	2.900	2.938
L	9	12	17	21	29	40
$\overline{C}$	0.003	0.003	0.003	0.006	0.003	0.003
S	1.005	0.991	0.980	0.845	0.091	0.089
γ	1.667	1.800	3.000	837	1784	1907
E_G	0.0724	0.0523	0.0385	0.0328	0.0238	0.0177

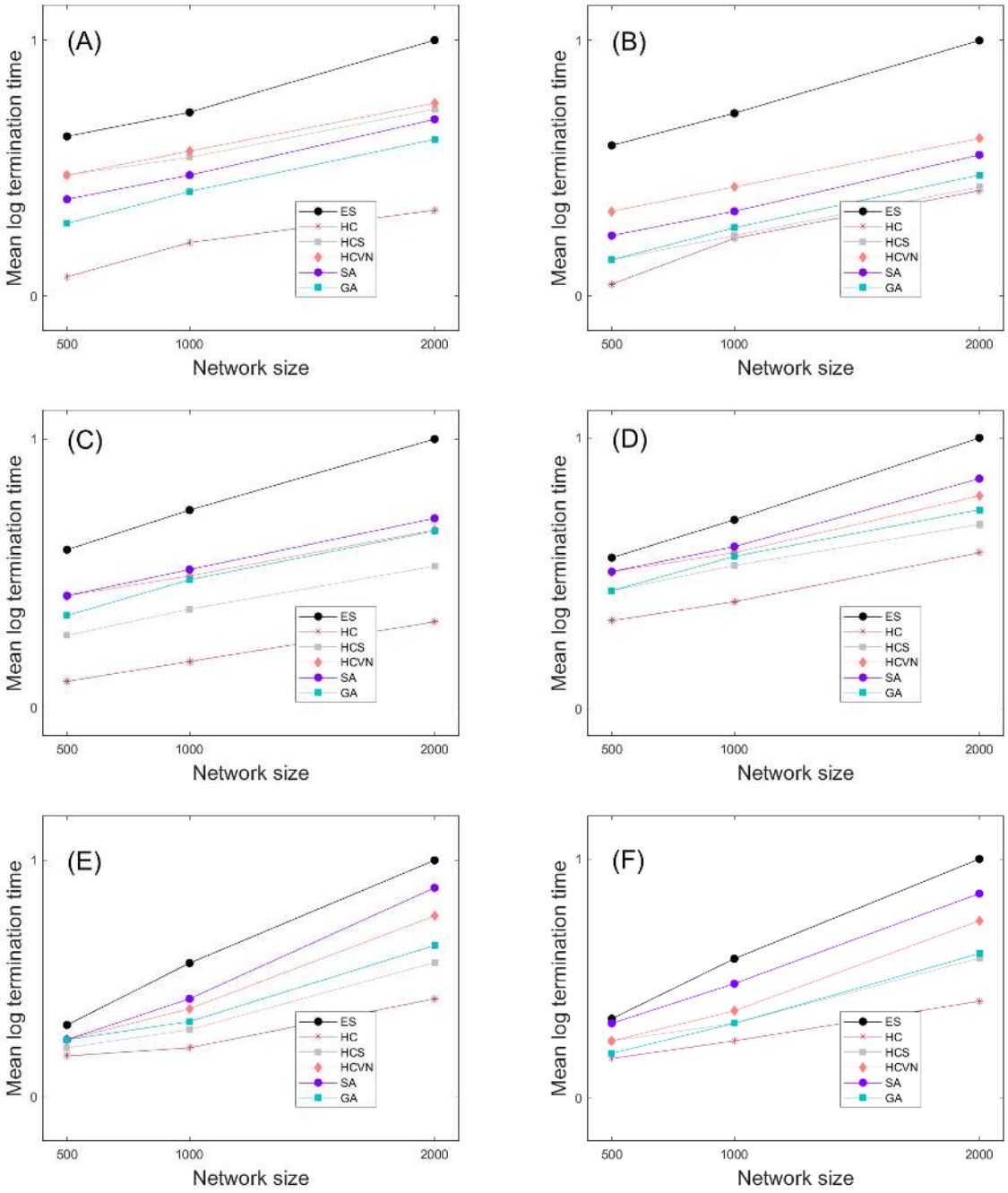


Figure 2.A2: The termination times for the heuristics applied to the random networks: (a) Erdös-Rényi networks, (b) Watts-Strogatz networks, (c) Barabási and Albert networks, and (d) Klemm and Eguílez networks, (e) Delaunay networks, and (f) Voronoi networks. These are average times for 1000 random restarts for optimization applied to 1000 random network of each type and size. The times are scaled by the exhaustive search time and log transformed for easier interpretation.

Table 2.A4: The network characteristics of the street networks around the schools used for the study. The average weighted clustering coefficient $\overline{C}$ was not calculated since triplets were uncommon in these networks and neither was the proximity ratio S since it depends on $\overline{C}$.

	Suburban school network						Rural school network			
	Elementary			Middle			Elementary			
	SE(1)	SE(2)	SE(3)	SM(1)	SM(2)	SM(3)	RE(1)	RE(2)	RE(3)	RE(4)
N	3952	4222	5149	7750	4895	8059	1727	2160	1539	1996
N_C	1772	1807	1077	4614	1198	2210	416	330	363	735
N_D	2180	2415	4072	3136	3697	5849	1311	1830	1176	1261
L_F	50	96	194	110	43	103	51	95	50	52
$\overline{C_D}$	2	2	2	2	2	2	2	2	2	2
L	80	133	183	135	72	123	77	88	71	82
γ	1179	1300	1604	2353	1432	2427	761	967	679	891
E_G	0.0060	0.0060	0.0045	0.0052	0.0074	0.0056	0.0078	0.0080	0.0104	0.0092

Table 2.A5: The random network parameters and values used for the analysis. Following Barabási and Albert (1999), the degree of new nodes for the Barabási and Albert graphs were equal to the initial network size ($m = m_0$).

Parameter		Description	Values
	N	Number of nodes	500, 1000, 2000
Erdös-Rényi graphs	p	connection probability	0.01
Watts-Strogatz graphs	p_W	rewiring probability	0.01
	k_L	initial node degree	10
Barabási and Albert graphs	m_0	connected network size	10
	m	degree of new nodes	$m0$
Klemm and Eguílez graphs	m_0	connected network size	10
	p_S	node selection probability	0.1
	p	connection probability	0.01
Delaunay random graphs	p_R	edge removal probability	0.1
Voronoi random graphs	p_R	edge removal probability	0.1

ALGORITHM 2.1
Exhaustive search pseudocode

```
for i in N_D do                                          ▷ Select a distant node.
    for j in N_C do                                      ▷ Select each close node.
        if d(i,j) + d(j,S) < D then                      ▷ If node will be within the distance.
            C(i,j) = d(i,j)                              ▷ Calculate the cost of the connection, i.e. the length.
            for k in N_D do                              ▷ Select each distant node.
                if d(k,i) + d(i,j) + d(j,S) < D then     ▷ Calculate the distance to the focal node.
                    k ∈ N'_C                             ▷ If node is within the distance assign it to the new close set.
                end if
            end for
            B(i,j) = |N'_C|                              ▷ Calculate the number of new close nodes.
        end if
    end for
end for
```

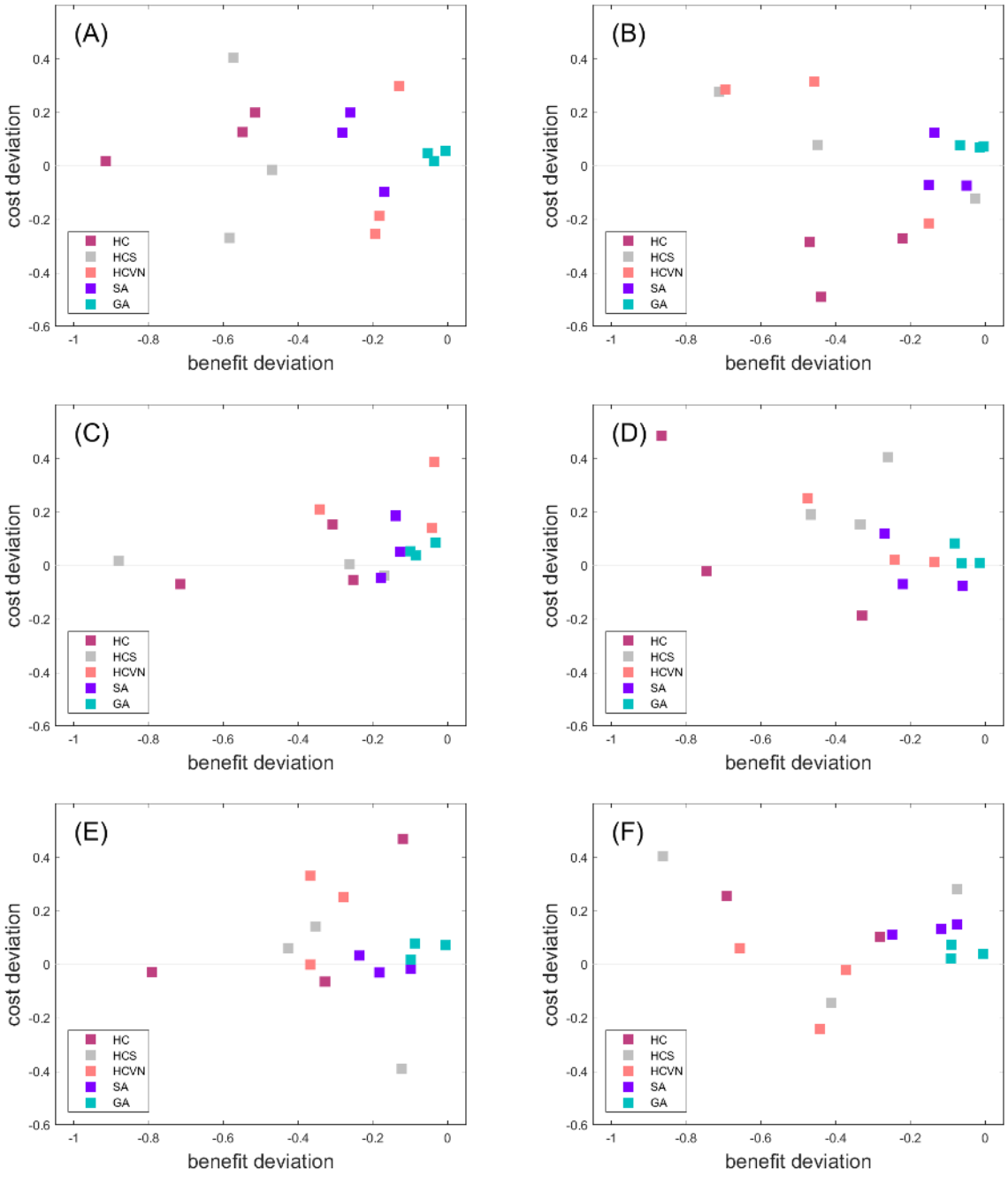


Figure 2.A3: The cost and benefit deviations for the heuristics applied to the random networks: (a) Erdös-Rényi networks, (b) Watts-Strogatz networks, (c) Barabási and Albert networks, and (d) Klemm and Eguílez networks, (e) Delaunay networks, and (f) Voronoi networks. Each point represents the results for a given method and network size ($N = 500, 1000, 2000$). The costs and benefits were scaled by the results from the exhaustive search, where a longer connection length is a positive cost deviation and a shorter connection is a negative cost deviation. These are average times for 1000 random restarts for optimization applied to 1000 random network of each type and size.

ALGORITHM 2.2
Hill climbing

randomly select i from N_D
randomly select j from N_C
$C(i,j) = d(i,j)$
for k **in** N_D **do**
    if $d(k,i) + d(i,j) + d(j,F) < D$ **then**
        $k \in N'_C$
    end if
end for
$B(i,j) = |N'_C|$
$O^0 = \alpha C(i,j) + \beta B(i,j)$
$t = 1$
while $O^t \neq O^{t-1}$ **do**
    N_i^D = neighbors of i in N_D
    N_j^C = neighbors of j in N_C
    for m **in** N_i^D **do**
        for n **in** N_j^C **do**
            $C(m,n) = d(m,n)$
            for k **in** N_D **do**
                if $d(k,m) + d(m,n) + d(n,F) < D$ **then**
                $k \in N'_C$
                end if
            end for
            $B(m,n) = |N'_C|$
        end for
    end for
    $O^t = \max_{(m,n)}(\alpha C(m,n) + \beta B(m,n))$
    $t = t + 1$
end while

ALGORITHM 2.3
Stochastic hill climbing

randomly select i from N_D
randomly select j from N_C
$C(i,j) = d(i,j)$
for k **in** N_D **do**
  if $d(k,i) + d(i,j) + d(j,F) < D$ **then**
    $k \in N'_C$
  end if
end for
$B(i,j) = |N'_C|$
$O^0 = \alpha C(i,j) + \beta B(i,j)$
$t = 1$
while $O^t \neq O^{t-1}$ **do**
  N_i^D = neighbors of i in N_D
  N_j^C = neighbors of j in N_C
  for m **in** N_i^D **do**
    for n **in** N_j^C **do**
      $C(m,n) = d(m,n)$
      for k **in** N_D **do**
        if $d(k,m) + d(m,n) + d(n,F) < D$ **then**
          $k \in N'_C$
        end if
      end for
      $B(m,n) = |N'_C|$
    end for
  end for
  $O^t = \alpha C(i,j) + \beta B(i,j)$ with probability$(i,j) = \dfrac{\alpha C(i,j) + \beta B(i,j)}{\sum_{(m,n)}(\alpha C(m,n) + \beta B(m,n))}$
  $t = t + 1$
end while

ALGORITHM 2.4
Hill climbing with a variable neighborhood

set n_{max}
randomly select i from N_D
randomly select j from N_C
$C(i,j) = d(i,j)$
for k **in** N_D **do**
    if $d(k,i) + d(i,j) + d(j,F) < D$ **then**
        $k \in N'_C$
    end if
end for
$B(i,j) = |N'_C|$
$O^0 = \alpha C(i,j) + \beta B(i,j)$
$\eta = 1$
$t = 1$
while $\eta < \eta_{max}$ **do**
    N_i^D = neighbors of i in N_D
    N_j^C = neighbors of j in N_C
    for m **in** N_i^D **do**
        for n **in** N_j^C **do**
            $C(m,n) = d(m,n)$
            for k **in** N_D **do**
                if $d(k,m) + d(m,n) + d(n,F) < D$ **then**
                $k \in N'_C$
                end if
            end for
            $B(m,n) = |N'_C|$
        end for
    end for
    $O^t = \max_{(m,n)}(\alpha C(m,n) + \beta B(m,n))$
    if $O^t > O^{t-1}$ **then**
        $\eta = 1$
    else
        $\eta = \eta + 1$
    end if
    $t = t + 1$
end while

ALGORITHM 2.5
Simulated annealing

randomly select i from N_D
randomly select j from N_C
$C(i,j) = d(i,j)$
for k **in** N_D **do**
  if $d(k,i) + d(i,j) + d(j,F) < D$ **then**
    $k \in N'_C$
  end if
end for
$B(i,j) = |N'_C|$
$O^0 = \alpha C(i,j) + \beta B(i,j)$
$t = 1$
while $O^t \neq O^{t-1}$ **do**
  N_i^D = neighbors of i in N_D
  N_j^C = neighbors of j in N_C
  for m **in** N_i^D **do**
    for n **in** N_j^C **do**
      $C(m,n) = d(m,n)$
      for k **in** N_D **do**
        if $d(k,m) + d(m,n) + d(n,F) < D$ **then**
          $k \in N'_C$
        end if
      end for
      $B(m,n) = |N'_C|$
      $O(m,n) = \alpha C(m,n) + \beta B(m,n)$
    end for
  end for
  if $\max_{(m,n)} O(m,n) > O^{t-1}$ **then**
    $O^t = O(m,n)$
  else
    $O^t = O(m,n)$ with probability$(m,n) = \exp\left(-\dfrac{O^{t-1} - O(m,n)}{t}\right)$
  end if
  $t = t + 1$
end while

ALGORITHM 2.6
Genetic algorithm

$\mu =$ mutation rate
$s =$ selection coefficient
$P =$ randomly selected population
Chromosome$(i) = (1...1)$
Chromosome$(j) = (1...1)$
for (i, j) **in** P **do**
    $C(i, j) = d(i, j)$
    for k **in** N_D **do**
        if $d(k, i) + d(i, j) + d(j, F) < D$ **then**
            $k \in N'_C$
        end if
    end for
    $B(i, j) = |N'_C|$
    $O(i, j)^0 = \alpha C(i, j) + \beta B(i, j)$
end for
$t = 1$
while $\max_{(i,j)} O(i, j)^t \neq \max_{(i,j)} O(i, j)^{t-1}$ **do**
    $f(i, j) = \dfrac{O(i, j)}{\sum_{(i,j)} O(i, j)}$
    populate P with (i, j) with probability$(i, j) = \dfrac{s * f(i, j) + (1 - s)}{\sum_{(m,n)}(s * f(m, n) + (1 - s))}$
    for k **in** P **do**
        for ℓ **in** chromosome(k) **do**
            if random number $\leq \mu$ **then**
                gene$(k)_\ell$ = gene$(k)_\ell$ + 1
            end if
        end for
    end for
    for (i, j) **in** P **do**
        randomly select m from N_i^D with probability gene$(i)_m / \sum_k$ gene$(i)_k$
        randomly select n from N_j^C with probability gene$(j)_n / \sum_k$ gene$(j)_k$
        $C(m, n) = d(m, n)$
        for k **in** N_D **do**
            if $d(k, m) + d(m, n) + d(n, F) < D$ **then**
                $k \in N'_C$
            end if
        end for
        $B(m, n) = |N'_C|$
        $O(m, n)^t = \alpha C(m, n) + \beta B(m, n)$
    end for
    $t = t + 1$
end while

Chapter 3

Small Changes in Street Connectivity can Result in Big Gains for Student Walking

Abstract

Student active commuting to school is an important component to student achievement and student health. Increasing street and trail connectivity between residential developments and schools is a key to fostering student active commuting. This study conducts a cost-benefit analysis of increased connectivity around schools. Benefits, which include potential cost-savings to a school system if they had fewer students to bus to school, increased student walking, and the reduction in health-care costs of fewer obese students, are compared to the financial costs of the new connections. Advanced network optimization techniques were applied to several urban and suburban schools from a U.S. school system to locate the optimal new connections that maximize student walking to a school. Results from this representative case study showed that short connections could lead to a large increase of potential student active commuters. This work can inform city planners, housing developers, and school officials on the impact of greater connectivity for student active commuting and residential development.

Introduction

School Active Commuting: Importance and Recent Decline in the U.S.

Physical activity is a valuable component of childhood development. A meta-analysis of the literature on student physical activity and academic achievement in the U.S. found a positive relationship between the two regardless of socioeconomics, demographics, geography, and school characteristics (Centers for Disease Control and Prevention, 2010). Increased physical activity has also been correlated with higher attendance rates and fewer disciplinary incidents (Welk, 2009). Children who actively commute to school, by walking or biking, tend to be more active outside of this commuting (Sirard and Slater, 2008, Lubans et al., 2011). This is important because the lack of physical activity is one of the primary contributors of childhood obesity (Robert Wood Johnson Foundation, 2009) and active commuting to school has been shown to be inversely associated with body mass index (Mendoza et al., 2011, Turrell et al., 2018).

Yet, there has been a rapid decline in physical activity reported for U.S. children (Nader et al., 2008) and active commuting to school (McDonald, 2007, Pedestrian and Bicycle Information Center, 2010, The National Center for Safe Routes to School, 2016). Parents have reported several that have contributed to this decline: the distance between home and school; the perception of possible violence or crime along the route; the speed and volume of traffic along the route; and poor weather or climate in the area; (Pedestrian and Bicycle Information Center, 2010, McDonald and Aalborg, 2009, Centers for Disease Control and Prevention, 2005, Mendoza et al., 2014). School systems also impose barriers to active commuting, often because of liability concerns (Chriqui et al., 2012).

School site design and location often make active transportation difficult. Since the 1970s, school systems have increasingly constructed larger schools on larger tracts of

land, often in rural areas further from population centers. One reason is an attempt to reduce construction and operating costs. Another is the imposition of national minimum-acreage standards, which are still adhered to in many places, even though they are no longer in place. In Knox County, TN, for example, the average acreage of elementary schools jumped from 8.5 acres (for schools built prior to 1977) to 24.5 acres (for schools built since then). There's a trend toward larger schools nationally as well, with the number of schools dropping from 262,000 in 1930 to 95,000 in 2004, while the student population has increased from 28M to 54.5M (Office of Children's Health Protection, 2011). School systems additionally find zoning ordinances limit their site selection choices.

School location is an important factor affecting student active commuting. Clarence Perry's Neighborhood Unit Plan advocated for a centrally located school to promote student walking (Perry, 1929). More centrally located schools have more student walking or biking, even after controlling for other neighborhood characteristics (Kim and Lee, 2016). Yet, active commuting for students is usually not considered part of the school location planning, as site acreage, and building costs are the top priorities.[1] Residential neighborhood developers are incentivized to minimize street connectivity in order to maximize the number of buildable lots, especially lots on cul-de-sacs, which are often popular with buyers because they are low-traffic streets, while the connectivity of the built environment has been found to be positively correlated with student active commuting as it contributes to a shorter distance between school and home (Babey et al., 2009, Bungum et al., 2009, Larsen et al., 2009).

The design of school sites also contributes to the lack of walkability, especially requirements that schools be single-story, and the provision of ample parking and driving spaces on school sites, but often not sidewalks. These siting and design factors contribute to urban sprawl and to greater distances between student homes and schools, and when coupled with a lack of sidewalks and bike paths greatly reduce student opportunities

[1] From personal communication with Knox County employees.

for active commuting (Kouri, 1999). As expected, neighborhood walkability and distance between home and school has been shown to be negatively correlated with childhood obesity (Spence et al., 2008, Grafova, 2008, Oreskovic et al., 2009). Due to school site selection in Knox County, from the 2013-14 school year to the 2017-18 school year, the number of students who lived outside the walking distance increased by approximately 6,000 (Transportation Consultants, 2014, Knox County Board of Education, 2018).

This decline in active commuting to schools leads to increased traffic from private driving and school busing, which accounts for 10-14% of morning rush-hour traffic (McDonald et al., 2011), and increased pollution. The additional traffic also intensifies parental safety concerns related to traffic and student active commuting. Still, the number of children killed and injured while walking or biking is dwarfed by the increasing rate of vehicle crashes, which are the leading cause of death among school-age children (National Center for Statistics and Analysis, 2016). This decrease in active commuting also imposes a considerable cost on school systems with increased costs related to purchasing and maintaining buses, hiring drivers (an occupation with a high turnover rate), and fuel and insurance expenditures (DeNisco, 2015). A study of school administrators found that addressing the perceived safety concerns and increasing the number of sidewalks can increase active travel to schools (Price et al., 2011), and one the key elements of the National Center for Safe Routes to School is transportation planning approaches to ensure safe active commuting opportunities (The National Center for Safe Routes to School, 2016).

School Active Commuting Policy Case Study

The Knox County School (KCS) public school system serves 60,000 students in and surrounding Knoxville, Tennessee, with 89 schools and 337 buses. KCS has an established policy to determine if a student is eligible for transportation by a bus, based on each student's residence location in relation to their school (Knox County Schools Transportation

Department, 2009). A student is inside the Parental Responsibility Zone (PRZ) – and ineligible for busing – if the distance from the student's residence to the drop-off location of the student's zoned school is less than 1 mile (for elementary students) or less than 1.5 miles (for secondary students), based upon the existing street network.

A report by the Knox County Department of Engineering and Public Works identified 22,322 students who live within a PRZ, a figure equal to 37.6% of the total number of students in the school district (Transportation Consultants, 2014). The report calculated the likelihood of each student walking to school based on the distance from home to school, and predicted that there were 7,434 potential daily walk trips, if there were good conditions for walking. More than 30% of those predicted walk trips originated outside the PRZ. Therefore, KCS is responsible for busing large numbers of students to school who could potentially be within active transportation distance, which places a significant financial burden on the school district.

In addition to distance, poorly connected street networks may pose a barrier to students walking or bicycling to school. An analysis conducted by the authors of this study found many Knox County schools sited near neighborhoods with street networks that have few intersections or frequent dead-ends (cul-de-sacs), i.e. neighborhoods with low street connectivity. As such, the network PRZs fail to capture many residences that are close to the school "as the crow flies," that is, the straight-line or "Euclidean distance" (see Figure 3.1).

Scope of Research

The purpose of this study is to estimate the potential benefits of policies that would lead to larger PRZs around new and existing schools as new neighborhoods are built with greater street connectivity. In order to do that, it models how existing PRZs could be expanded via greater street connectivity in a way that would capture more households and students. Several potential benefits may result from this research. First, the study will examine

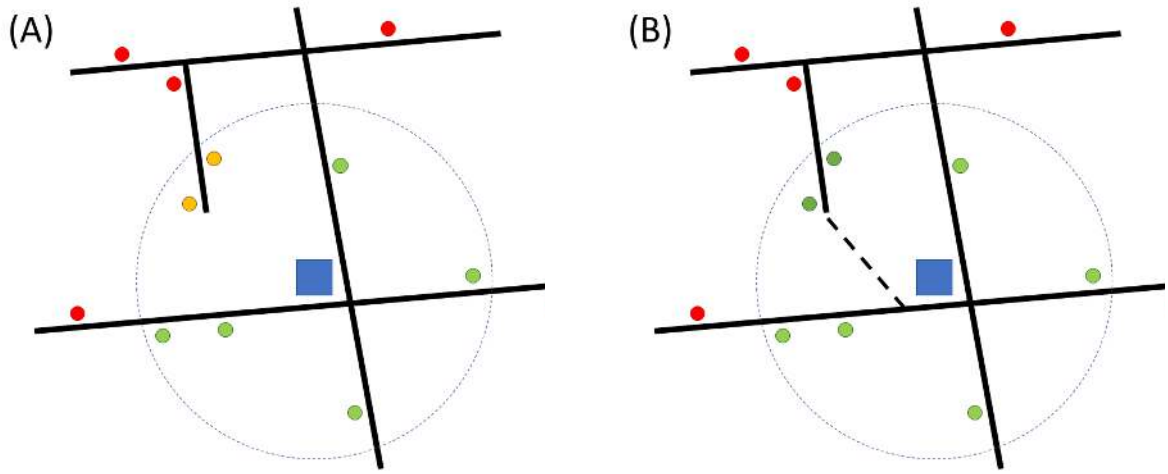


Figure 3.1: The left figure provides an example of residential parcels (orange circles) within the Euclidean PRZ distance (blue dashed line) of the school (blue square) but not within the network PRZ distance. The residential parcels within the network PRZ distance are represented by green circles and the residences outside of the Euclidean PRZ distance are the red circles. In the right figure the dashed line depicts the optimal new connection that maximizes the number of residences now within the network PRZ distance and minimizes the length of the connection.

the potential cost-savings to the Knox County Schools if they had to provide busing for a smaller proportion of students. Second, it will look at the expected health impact to students (e.g., increased physical activity) by increasing the chance that they will walk or bike to school. Third, the reduction of health costs from reduced obesity prevalence. Fourth, an analysis of subdivision trails (greenways) to explore how modifying the PRZ school policy include these types of connections increases the above benefits. Currently, KCS does not consider greenways or trails as walkable thoroughfares.

Combinatorial optimization techniques were employed to identify and evaluate new street connections, expanding on optimization approaches used for greenway planning and bus stop locations (Linehan et al., 1995, Ibeas et al., 2010, Delmelle et al., 2012). These techniques were utilized with data provided by the Knoxville-Knox County Metropolitan Planning Commission (MPC) to identify the lengths and economic costs of the new streets and trails, and the benefits of these additional connections, specifically the number

of additional residences within the walking distance to a school, the increased physical activity of the additional students who can actively commute, and the possible reduction of health care costs related to childhood obesity. Results from this work will be shared with Knox County planners and other decision-makers with the intent of making them aware of the costs imposed by lack of street connectivity in neighborhoods (Thaler and Sunstein, 2008). Results from this work could also influence planners to consider street connectivity and student active commuting when siting schools. Furthermore, these results could also start the dialogue of allowing greenways to be included as student commuting paths for school systems that currently do not consider them as permitted thoroughfares for student active commuting. The presence of greenways near residences have shown to increase property values (Nichols and Crompton, 2005) as prospective home buyers are willing to pay more for a home in a walkable neighborhood (Knoxville Area Association of Realtors, 2017). Finally, after optimal connections are computationally identified, residents should be involved with the design of their future communities (Forester, 1999).

Methods

Data

Since student residences for a given school can change over the course of the year and from year to year, as students enroll, graduate, or move, we used residential parcels as the proxy for student residence locations. Data about residential parcel locations and types, such as single-family residence or multi-family residence, were provided by MPC, which is the GIS administrator for Knoxville, Knox County, and the Knoxville Utilities Board. MPC has also provided the average number of students for residential parcel types in the study area, which was used to estimate the number of students that would be affected by changes in street connectivity. MPC has also supplied the number of students for each school in Knox County, the number of students within each school's PRZ, and the number

of students that are within the PRZ and network distance to their respective school. The data was collected in 2017 and engineered according to the methods provided in the Supplementary Material.

Figure 3.2: Map of the study area with the Euclidean PRZ distance shown for each school selected for analysis. Elementary schools have a 1-mile Euclidean PRZ distance and middle schools have a 1.5-mile Euclidean PRZ distance.

School Selection and Optimal Location of Connections

To identify schools that would benefit most from additional street connectivity, we developed a metric for a school's PRZ distance disparity. PRZ distance disparity is the difference between the number of students in the Euclidean PRZ distance and the number of students in the network PRZ. A large proportion of students within both the Euclidean PRZ

distance and the network PRZ is associated with an environment that has high street connectivity, typically urban. A large proportion of students within the Euclidean PRZ distance with and a low proportion in the network PRZ is indicative of a built-up yet low-connectivity environment, typically suburban. Low proportions in both the Euclidean PRZ distance and the actual PRZ is associated with a low-density, typically rural, environment. The ten schools with the largest PRZ distance disparity were selected for the analysis in this study (see Figure 3.3). These ten schools were from suburban and rural environments, as urban schools already had large proportions of students within the network PRZs and therefore low PRZ distance disparity. An exhaustive search of all possible connections between all residences and street intersections was conducted for each of the ten schools (see Algorithm 3.1 in the Appendix for the optimization algorithm pseudocode and Chapter 2 for optimization heuristics that can reduce the computational costs for large networks).

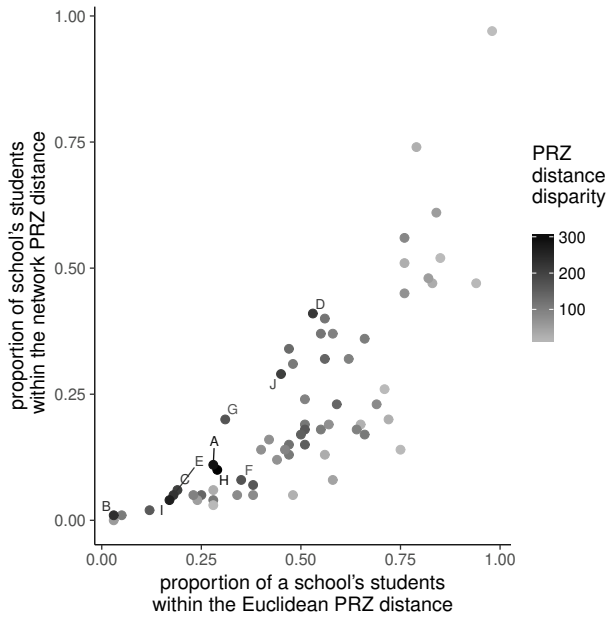


Figure 3.3: PRZ distance disparity for Knox County schools. The ten schools with highest PRZ distance disparity, i.e. the number of students who could walk if there was additional street and trail connections, are labeled and selected for analysis.

Energy Expenditure

To estimate the health benefit of potential new street and trail connections, the number of additional residences within the PRZ were converted to the number of potential students who could walk or bike to school. The total benefit was calculated by taking this number of students and multiplying it by the estimated duration of the physical activity. The average U.S. student who walks to school acquires 16 minutes of moderate-intensity activity and they walk at a rate of 71.7 meters per minute Crouter et al. (2013). Assuming a linear relationship between distance walked and walking time, it would take an average student 22.5 minutes to walk 1 mile, and this rate was used to calculate the walking benefit for the students along the optimal connection for each school evaluated. These times of physical activity can be converted to the metabolic equivalent of task to compare active commuting to other forms of exercise (Bassett et al., 2013, Basset, 2013).

Busing Costs

Knox County Schools contracts for 337 passenger buses with each bus having the capacity to carry either 65 or 90 passengers. Assuming that each bus completes approximately four runs per day, KCS pays its bus contractors daily rates of $219.03 and $259.85, respectively. [2] However, if a bus travels more than 52 miles a day, there is an overmile charge of $0.01 for each additional mile with the school system subsidizing fuel costs if gasoline costs more than $2.80 per gallon at that point in time. Afternoon ridership is typically greater than morning, when parents are more likely to drive students. There is also joint ridership for several different schools, for example students at a nearby elementary school and middle school may share a bus trip.

Busing costs for Knox County Schools are increasing. The proposed budget for fiscal year 2019 includes an increase of $1.9 million for busing costs (Knox County Board of

[2] Busing rates and over-mile charges were collected during an interview with a Knox County School official.

Education, 2018). $1 million of that proposed increase is to pay to contractors. The remaining increase is to serve additional routes and new schools set to open in the 2018-19 school year. KCS has also raised its standards for bus contractors following a crash of two school buses in 2014 that resulted in the deaths of two students and a teacher's aide. Officials describe finding enough bus operators as an "everyday struggle," because of the more stringent requirements on bus operators at both the state and local level (Deese, 2018).

Health Economics

According to 2016 data, in the U.S. 1 in 5 school-age children are obese (Hales et al., 2017). Obese children ages 2-19 in the United States were found to stay on average 0.85 days longer for hospital treatment, incur $2000 additional charges, and $925 more in hospital costs than non-obese children (in 2017 dollars) (Trasande et al., 2009). Furthermore, obese children incurred $210 in higher outpatient visit expenditures, $124 higher prescription drug expenditures, and $13 higher emergency room expenditures (Trasande and Chatterjee, 2012). There were approximately 73.8 million children under the age of 18 in the United States in 2017 (United States Census Bureau, 2016) and children were hospitalized in the U.S. at a rate of 0.0014 in 2012 (Witt et al., 2014). Assuming obese children were hospitalized a this rate, it is estimated that the hospital treatments of obese children in 2017 resulted in $40 million dollars in additional hospital costs and $20 million in additional charges.

Connection Costs

The financial costs of streets and greenways per mile for Knox County are $1M and $500K, respectively. This is a very general cost and does not include site-specific engineering requirements (such as bridges) and land costs, which vary widely. The cost for residential developers is approximately half a million per mile, excluding ROW, crossing

water/wetlands, and major topography (Bushell et al., 2013).

Results

The location of the optimal connections are given in Figure 3.4 and Supplementary Figure 3.A1. Results for the costs and benefits were calculated by taking the average of 1,000 random samples, using the proportion of student per residence, of the new residences from the optimal connection solution for each school. The length of the optimal connection, the number of residences and students included in the network PRZ with the new connection, and the total time spent walking to and from the school for these students are provided in Table 3.1.

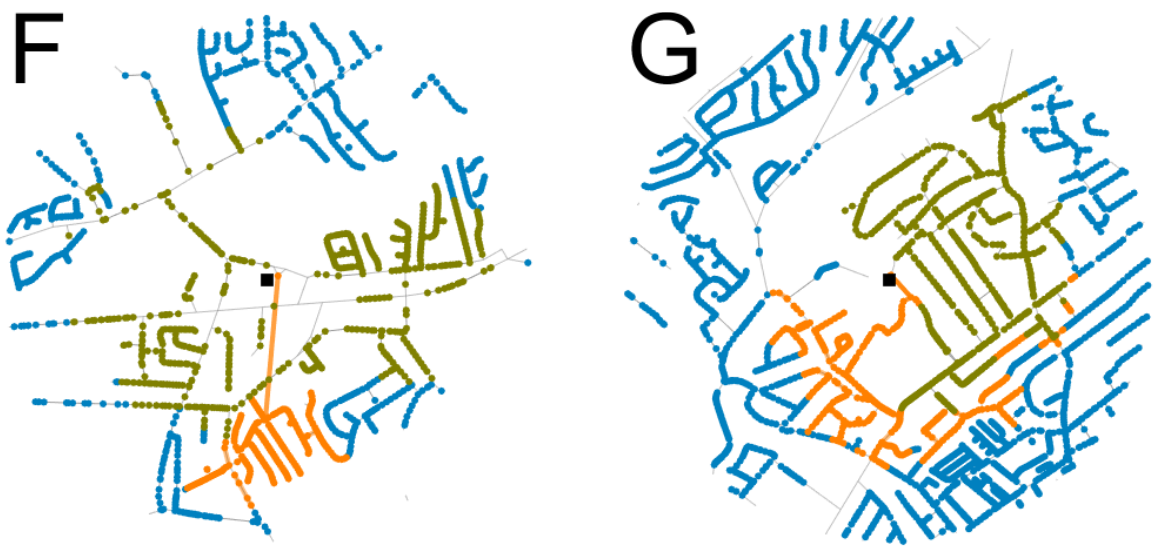


Figure 3.4: Optimal connection results for a rural elementary school (left) and a suburban elementary school (right). The black square represents the school, and grey lines indicate streets. Dark green points are residences within the network PRZ distance, blue points signify residences not within the network PRZ distance and orange points are the residences that are within the Euclidean PRZ distance after the new connection is made, denoted by the orange line.

Table 3.1: Active commuting characteristics and optimal results for the ten schools with the greatest walking potential. Schools denoted with an * are located in a suburban area, otherwise the school is located in a rural area. Standard deviations for the optimal walking times are given in the parentheses.

	School	Proportion students within Network PRZ	Proportion students within Euclidean PRZ	walking potential	# distant residences	students/ residence	Optimal # residences	Optimal length (ft)	Optimal # students	Optimal walking (m)
Elementary	A*	0.11	0.28	292	1680	0.17	453	3704	77	2936 (30)
	B	0.01	0.03	239	1035	0.23	137	3891	32	1308 (16)
	C	0.05	0.18	228	1173	0.19	203	3613	39	1536 (24)
	D*	0.41	0.53	220	1911	0.12	255	3133	31	1168 (26)
	E	0.06	0.19	210	1154	0.18	287	3307	52	2008 (30)
	F	0.08	0.35	173	1120	0.15	287	2907	43	1502 (32)
	G*	0.20	0.31	171	1027	0.17	703	602	120	4220 (76)
Middle	H*	0.10	0.29	316	2033	0.13	940	3260	122	6452 (112)
	I	0.04	0.17	269	1058	0.25	464	3554	116	5994 (100)
	J*	0.29	0.45	209	4246	0.05	381	5027	19	1104 (28)

Discussion

The results show that even the addition of short street or trail segments can greatly increase the number of residences within the PRZ of a school. These additional residences now within the PRZ increase the potential number of students who could actively commute to school, and commuting is positively correlated with improved health and education achievement. Planners and other decision-makers can use this result to increase their awareness about the costs, to families and to school systems, imposed by the lack of street connectivity in new neighborhoods. Another way to reduce busing costs is to stagger bell times for different school, which Knox County Schools does not do, which would reduce the number of vehicles and drivers needed.

Limitations

There are several issues with the current analysis. First, the study area is limited to one U.S. county, and to make these results more robust, the networks for schools in additional study areas should be evaluated. Second, residences may be placed on the nearest street but that may not be the street they are physically located on. The error rate for this occurring is small for this study area but could vary for different study areas. Third, restrictions on the locations of new streets and trails need to be included. For example, streets and trails cannot be built on slopes beyond a specific threshold, and it may be cost-prohibitive or impossible to route them through specific parcel types (such as commercial areas, mining sites and landfills) or highways. Fourth, solutions are dependent on the location of nodes at every parcel and intersection. This can lead to large sparse areas of possible connectivity and should be corrected with the placement of artificial nodes at regular intervals. Fifth, the issue that newly constructed connections can intersect existing streets (planarity) was not resolved in this study. Calculating the determinants for each pair of lines is computationally expensive. To counter this, a coloring scheme where each

region of the space segmented by edges is classified as a unique color has been proposed. Every node is labeled with the colors its edges touch. A new connection can only be established between nodes that share a color. This would be used to reduce the set of lines needed to calculate the determinants for. Lastly, due to the limitations of shapefiles, it is recommended that government agencies move away from their dependence on them for spatial data. There are several emerging data formats, such as OGC GeoPackage and GeoJSON, that are robust and free.

Conclusions

Knox County, TN, like many other places in the U.S., has one group that make decisions about school policies, such as siting and busing (the KCS School Board), and another that makes decisions about the design of new neighborhoods and other developments (the MPC). Those groups may not always be aware of how decisions they make affect each other. This paper attempts to describe and quantify the ways in which decisions made by the MPC – specifically about street connectivity in developments near schools – can impose additional costs on Knox County Schools, in the form of the need to provide more busing. Reduced street connectivity around schools also imposes costs on families and on children themselves, in the form of less opportunity for physical activity, potentially lower academic achievement, and increased risk of obesity and other health problems.

It would be in the best interest of KCS to request that the MPC take into account the impact that low street connectivity around schools has on busing costs. It is probably too late to remedy the low street connectivity around the schools examined in this paper, because of cost and likely neighborhood opposition. But the key finding of this paper – that short connections can vastly increase connectivity and walkability around schools, while decreasing future busing and health costs – should inform future decision-making about neighborhood design around schools with the goal of reducing costs and improving

students' health and academic achievement. An expanded working relationship between MPC and KCS may eventually lead to additional opportunities to collaborate toward cost savings, such as better coordination of school siting with other land use goals for Knox County.

Future directions for this work include the optimization of school bus drop-off points and allowing for multiple connections to be evaluated simultaneously, a significant nontrivial extension. Once these additions have been added, including the solution to the planarity issue, a tool should be developed for planners in Python for ArcGIS. MPC provided socioeconomic data at the parcel level that could be used to identify inequalities in active commuting, such as car ownership, rental identification, household income, and free lunch eligibility. An evaluation of student walking in the context of their socioeconomic status should also be conducted to determine if there are inequalities in active commuting and how they should be addressed (Althoff et al., 2017).

Appendix

Additional Results

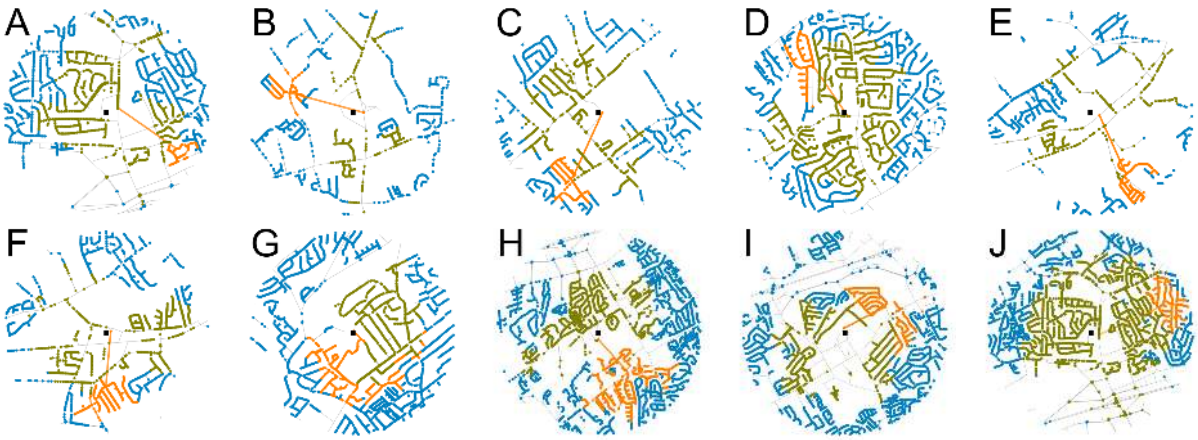


Figure 3.A1: Optimal connection results for the ten schools. The black square represents the school, grey lines indicate streets, dark green points are residences within the network PRZ distance, blue points signify residences not within the network PRZ distance, and orange points are the residences now within the network PRZ distance after the new connection is made, denoted by the orange line.

Data Engineering

The spatial data provided by KGIS was in shapefiles, which are commonly used by government agencies but have significant limitations for analysis. The data cleaning and processing methods described below are provided for planners and researchers to use for their own systems (see Karduni et al. (2016) and Oliver et al. (2007) for additional useful methods to process shapefile data for network analysis).

To identify optimal connections a network approach is used and the data was converted to a network with nodes at residences and street intersections and the streets were converted to edges. To accomplish this data transformation, in ArcMap (version 10.6):

1. buffers were created around each school (1 mile for elementary schools and 1.5 miles for middle schools).

2. street edge and street intersection node development:

   (a) street data outside of the school buffer were removed (using the clip tool),

   (b) highways were removed, as students will not use them for active commuting,

   (c) streets part of the school property were also removed (as schools do not want direct connections),

   (d) nodes at the street intersections were created and labeled as intersections,

      i. points at the street intersections were created (intersect tool with output type point),

      ii. two numeric fields were added to this layer using a field type DOUBLE to ensure maximum precision and the spatial coordinates were identified for the street intersection points (calculate geometry),

      iii. ArcGIS creates these intersections as multipoint geometries (coinciding with the number of streets that meet at the intersection), therefore these intersections were converted to single points (feature to point tool)

      iv. the identical intersection points were removed by their coordinates (delete identical tool),

      v. a numeric field was added to this layer and all street intersection points were given a value of 0,

3. residential node development[3]:

   (a) residential parcels outside of the school buffer were removed (using the clip tool),

   (b) vacant residential lots were removed (select by attributes),

   (c) nodes at the street intersections were created and labeled as intersections,

[3]Apartments were already divided into separate single family residence points. If this is not the case then apartments should be labeled with the number of units for the optimization method to consider when calculating the benefit of the new connections.

i. the nearest street location was identified for each residence (the near tool),

ii. the attributes of these points were converted to a database table (output attribute as .dbf),

iii. these residential coordinates plotted and saved as a layer (display XY data with the Near_X and Near_Y coordinates),

iv. a numeric field was added to the layer and all residential points were given a value of 1,

4. a school node was placed at the nearest street intersection and labeled,

5. the street intersection point data, the school point, and the residential point data were merged (merge tool) with the following important fields kept: (1) X-coordinate, (ii) Y-coordinate, and (iii) node type and exported as a text file,

6. the street polylines were cut at each residential and intersection point (split line at point tool). A search radius of 100 ft was used to offset any residential nodes that were not exactly on a street and this also provides additional solutions for the optimization method.

Due to the limitations of ArcGIS, for example the inability of the software to produce a network connectivity matrix (which is an extremely useful mathematical representation), the cut street data files were imported into Python and the NetworkX package (version 2.1) was used to produced adjacency (connectivity) matrices (Hagberg, 2017). These adjacency matrices and the node data (school, residence, intersection) were imported into MATLAB (version 9.3 R2017b) for the optimization routines. Note, the street cutting process mentioned above (step 6) also creates nodes at the end of streets and these are considered as intersections for the purpose of the analysis. Network distances between all nodes were calculated and nodes were labeled as close or distant distant based on the walking network distance.

Optimal Connection Search Algorithm

To find the optimal connection that maximizes the number of residences within the network PRZ, exhaustive search optimization algorithms were performed in MATLAB (version 9.3 R2017b, see Algorithm 3.1). The exhaustive search optimization routine creates an edge for every combination of distant and close nodes and evaluates the cost of the new connection, i.e. the length of the new connection, and the benefit of the new connection, i.e. the number of residences that are now within the network PRZ. Note we parallelize the optimization routine as finding all of the solutions for large networks is computationally expensive (thousands of nodes results in millions of possible connections to evaluate). Heuristics have been developed to find near-optimal solutions when the exhaustive search is not feasible (See Chapter 2).

ALGORITHM 3.1
Exhaustive search pseudocode

```
for i in N_D do                                          ▷ Select a distant node.
    for j in N_C do                                      ▷ Select a close node.
        if d(i,j) + d(j,S) < D then                      ▷ If node is within the PRZ.
            C(i,j) = d(i,j)                              ▷ Calculate the distance of the new connection.
            for k in N_D do                              ▷ Select each distant residence.
                if d(k,i) + d(i,j) + d(j,S) < D then     ▷ Calculate the distance to the school.
                    k ∈ N'_C                             ▷ If residence is within PRZ, then assign it to the new close set.
                end if
            end for
            B(i,j) = |N'_C|                              ▷ Calculate the number of new close residences.
        end if
    end for
end for
```

For the optimization algorithm, the network distance between two nodes is given by $d(i, j)$, and if the distance between a node and the school node (S) is less than the PRZ distance, $d(i, S) < D$ (1 mile for elementary schools and 1.5 miles for middle schools) then the node i is assigned to the set of close nodes (N_D), otherwise it is assigned to the distant set (N_D). After a new connection is established, the nodes that are now within the network PRZ distance are assigned to the set N'_C. The cost of the new connection is the

connection length, $C(i,j) = d(i,j)$, and the benefit of this new connection $B(i,j)$ is the number of additional residences now within the PRZ, i.e. $B(i,j) = |N'_C|$ minus the number of nonresidential nodes. The optimal solution is the solution with the greatest benefit, or number of new residences now within the distance to the school, which can be expressed as the bi-objective function

$$O^* = \max_{(i,j)}(B(i,j) + \alpha C(i,j)),$$

where $\alpha = \infty$ and implies the objective is only to minimize costs for the same benefit.

Conclusion

This dissertation provided several methods to address issues in transportation planning using network theory approaches. Each chapter explored a different question where understanding the geography of the system was paramount and viewing these spatial systems as networks shed new light on how to understand them.

Traditional measures appear to not be enough to evaluate the robustness of transit networks and other multi-line systems. By redefining network components and beginning the development of new indexes in chapter 1 we can begin to capture some of the information previously lost in these multi-line systems and identify differences in network criticality. These new measures appear to be consistent in their response and highly correlated to the existence of additional lines in a network and provide additional information about the system's tradeoff between a resilient structure and one that attempts to cover the geography of demands. These new indexes can be applied to categorizing networks and have the same range of values, both important characteristics for measuring the robustness of transportation systems.

Rail systems are recognized as the major public transit networks of US major metro areas as they provide high-speed mobility for passengers with separate rights-of-way from which all other vehicular and foot traffic are excluded. Measuring robustness and assessing the system resilience are critical concerns as the portion of service for ridership is considerably increasing in highly dense urban areas. By using our measures, we can help decision makers and transportation planners prioritize the protection or maintenance of stations, not based on a simple but monotonic traditional approach, but our comprehensive approach.

There are several avenues for future work in this area including disruption scenarios and their impact on the indexes. Extending these global measures for systems with directed graphs and circuits should also be considered. Once the robustness measures

mentioned in Ellens and Kooij (2013) and established Laplacian techniques are applied to multi-line matrices and their F-matrices, results can be compared with the findings presented in this work. Another extension is developing additional local measures for robustness, centrality, and connectivity in multi-line networks using network spatial partitioning (Ji and Geroliminis, 2012). Finally, these new indexes can also provide novel ways to look at distance based matrices, resilience simulations, and longitudinal analysis.

The network connectivity problem introduced in chapter 2 is relevant to a wide range of applications and is nontrivial as the number of solutions can become large even for small systems. This type of combinatorial optimization problem highlights the difficulty in determining local search routines a priori. The nodal characteristics were nonlinearly related to the solutions while different characteristics varied in their correlation with solution quality for different networks making it difficult to exclude specific characteristics for network connectivity optimization. Distance to the focal node was consistently related to the quality of the solution as this lowers connectivity length costs, while centrality was intermittently correlated with solution quality it provides greater benefit through more connections. Clustering nodal characteristics did not provide additional useful information from the nodal characteristics for the random networks. This could arise from the following issues: the curse of dimensionality, i.e. large sparse subspaces in the solution space; the nodal characteristics are highly correlated with each other; outliers; finding the appropriate influential nodal characteristics is not possible a priori; and the influence of specific characteristics is dynamic as the heuristics converge. For the school networks: the clustering coefficient was a poor measure due to the lack of triplets in the networks; the network modularity also had poor results possibly due to the measure's inability to account for the spatial component of the nodes; and the network-constrained clusters were also poor in explaining the solutions, due to the complexity of the network topology.

The optimization heuristics save computational time but vary considerably in their ability to find a solution near the optimal. The stochastic hill climbing search was not effective

due to the large neighborhood search space explored. In our experiment, the number of solutions checked at each iteration is > 300 and resulted in a skewed probability distribution of objective values favoring the selection of low values. This degraded the efficiency of the method resulting in the selection of poor solutions. The variable neighborhood search method was similarly not reliable because of the significantly large neighborhood search space (the number of possible solutions explored at a given iteration could be $> 5,000$), and had intermediate results with cost and benefit deviations. The simulated annealing heuristic consistently took longer to converge than the other optimization methods from the exploration of suboptimal solutions prior to moving towards better solutions, yet it was able to converge to values close to the optimal solution.

The computational costs and the variance in the importance of nodal characteristics for the random networks and real-world systems highlights the need for a heuristic that is able to quickly and effectively explore the solution space. The genetic algorithm provided in this work offers a solution to this issue and outperformed the other algorithms in terms of the consistently higher solution precision and accuracy. The genetic algorithm is able to dynamically reduce the size of the neighborhood search space and what variables to analyze. This reduction in the local solution search space allows the genetic algorithm presented here to converge on solutions near the optimal in a timely fashion. This shows the power of biologically inspired algorithms to effectively explore multidimensional spaces (commonly found in natural systems) and their potential use in a wide variety of disciplines, including specific applications for planning and health care.

Application of these methods and heuristics to multi-level networks, such as telecommunication systems, higher dimensional real-world networks (transportation networks with elevation), directed networks, and additional planar random networks (e.g. Gabriel graphs) should be conducted. Different distance measures to the focal node, such as the Hamming distance, could also be evaluated for different applications, and other real world examples should be used for analysis. The methods presented here do not evaluate

whether the new connections intersect existing edges and attempts to incorporate such a feature resulted in unrealistic computational times. Optimizing this feature is currently being developed as is a tool for ArcGIS and Python for planners and researchers to utilize.

Knox County, TN, like many other places in the U.S., has one group that make decisions about school policies, such as siting and busing (the KCS School Board), and another that makes decisions about the design of new neighborhoods and other developments (the MPC). Those groups may not always be aware of how decisions they make affect each other. Chapter 3 describes and quantifies the ways in which decisions made by the MPC – specifically about street connectivity in developments near schools – can impose additional costs on Knox County Schools, in the form of the need to provide more busing. Reduced street connectivity around schools also imposes costs on families and on children themselves, in the form of less opportunity for physical activity, potentially lower academic achievement, and increased risk of obesity and other health problems.

It would be in the best interest of KCS to request that the MPC take into account the impact that low street connectivity around schools has on busing costs. It is probably too late to remedy the low street connectivity around the schools examined in this paper, because of cost and likely neighborhood opposition. But the key finding of this paper – that short connections can vastly increase connectivity and walkability around schools, while decreasing future busing and health costs – should inform future decision-making about neighborhood design around schools with the goal of reducing costs and improving students' health and academic achievement. An expanded working relationship between MPC and KCS may eventually lead to additional opportunities to collaborate toward cost savings, such as better coordination of school siting with other land use goals for Knox County.

Future directions for this work include the optimization of school bus drop-off points and allowing for multiple connections to be evaluated simultaneously, a significant nontrivial extension. Once these additions have been added, including the solution to the pla-

narity issue, a tool should be developed for planners in Python for ArcGIS. MPC provided socioeconomic data at the parcel level that could be used to identify inequalities in active commuting, such as car ownership, rental identification, household income, and free lunch eligibility. An evaluation of student walking in the context of their socioeconomic status should also be conducted to determine if there are inequalities in active commuting and how they should be addressed (Althoff et al., 2017).

The results from this dissertation can contribute to a better understanding of transportation issues and help planners and researchers working in these domains. Extensions of this work are also not limited to transportation and geography, but provide avenues of research for interdisciplinary scientists working at the intersection of planning, computational methodology, and the social sciences.

References

Albert, R. and Barabási, A. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97.

Althoff, T., Sosič, R., Hicks, J. L., King, A. C., Delp, S. L., and Leskovec, J. (2017). Large-scale physical activity data reveal worldwide activity inequality. *Nature*, 547:336–339.

Alvarez-Socorro, A. J., Herrera-Almarza, G. C., and González-Díaz, L. A. (2015). Eigencentrality based on dissimilarity measures reveals central nodes in complex networks. *Scientific Reports*, 5.

An, K. and Lo, H. (2014). Service reliability-based transit network design with stochastic demand. *Transportation Research Record: Journal of the Transportation Research Board*, 2467:101–109.

Anderson, E. J. and Ferris, M. C. (1994). Genetic algorithms for combinatorial optimization: The assembly line balancing problem. *ORSA Journal on Computing*, 6:161–173.

Anderson, S. J., Karumanchi, S. B., and Iagnemma, K. (2012). Constraint-based planning and control. In *Proceedings of the Intelligent Vehicles Symposium (IV)*. IEEE.

Babey, S., Hastert, T., Huang, W., and Brown, E. (2009). Sociodemographic, family, and environmental factors associated with active commuting to school among US adolescents. *Journal of Public Health Policy*, 30(Suppl 1):S203–S220.

Barabási, A. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286:509–512.

Barrat, A., Barthélemy, M., Pastor-Satorras, R., and Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences USA*, 101(11):3747–3752.

Basset, D. R. (2013). Active living research: Using evidence to prevent childhood obesity and create active communities. Technical report, Robert Wood Johnson Foundation.

Bassett, D. R., Fitzhugh, E. C., Heath, G. W., Erwin, P., M., F. G., Wolff, D. L., Welch, W. A., and Stout, A. B. (2013). Estimated energy expenditures for school-based policies and active living. *American Journal of Preventative Medicine*, 44(2):108–113.

Bavelas, A. (1950). Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, 22:725.

Benezech, V. and Coulombel, N. (2013). The value of service reliability. *Transportation Research Part B: Methodological*, 58:1–15.

Berche, B., von Ferber, C., Holovatch, T., and Holovatch, Y. (2009). Resilience of public transport networks against attacks. *The European Physical Journal B*, 71(1):125–137.

Branas, C. C., MacKenzie, E. J., Williams, J. C., Schwab, C. W., Teter, H. M., Flanigan, M. C., Blatt, A. J., and ReVelle, C. S. (2005). Access to trauma centers in the united states. *Journal of the American Medical Association*, 293(21):2626–2633.

Brimberg, J. and Hodgson, J. M. (2011). Heuristics for location models. In Eiselt, H. A. and Marianov, V., editors, *Foundations of Location Analysis*, chapter 15, pages 335–355. Springer.

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117.

Bryan, K. and Leise, T. (2006). The $25,000,000,000 eigenvector: The linear algebra behind Google. *SIAM Review*, 48(3):596–581.

Bungum, T., Lounsbery, M., Moonie, S., and Gast, J. (2009). Prevalence and correlates of walking and biking to school among adolescents. *Journal of Community Health*, 34(2):129–134.

Bushell, M., Poole, B., Rodriguez, D., and Zegeer, C. (2013). Costs for pedestrian and

bicyclist infrastructure improvements: A resource for researchers, engineers, planners and the general public. Technical report, UNC Highway Safety Research Center.

Cats, O. and Jenelius, E. (2014). Dynamic vulnerability analysis of public transport networks: Mitigation effects of real-time information. *Networks and Spatial Economics*, 14(3-4):435–463.

Centers for Disease Control and Prevention (2005). Barriers to children walking to or from school–United States, 2004. Technical report, U.S. Department of Health and Human Services.

Centers for Disease Control and Prevention (2010). The association between school based physical activity, including physical education, and academic performance. Technical report, U.S. Department of Health and Human Services.

Chriqui, J. F., Taber, D. R., Slater, S. J., L., T., Lowrey, K. M., and Chaloupk, F. J. (2012). The impact of state safe routes to school-related laws on active travel to school policies and practices in U.S. elementary schools. *Health & Place*, 18(1):8–15.

Crouter, S. E., Horton, M., and Bassett, D. R. (2013). Validity of ActiGraph child-specific equations during various physical activities. *Medicine & Science in Sports & Exercise*, 45(7):1403–1409.

Crucitti, P., Latora, V., and Marchiori, M. (2004). Model for cascading failures in complex networks. *Physical Review E*, 69.

Deese, H. (2018). What's driving the school bus driver shortage. *The Ledger*, 42(14). `http://www.tnledger.com/editorial/Article.aspx?id=105555`.

Delaunay, B. (1934). Sur la sphère vide. *Bulletin de l'Académie des Sciences de l'URSS, Classe des sciences mathématiques et naturelles*, 6:793–800.

Delmelle, E., Shuping, L., and Murray, A. (2012). Identifying bus stop redundancy: A gis-based spatial optimization approach. *Computers, Environment and Urban Systems*, 36:445–455.

Demaine, E. D. and Zadimoghaddam, M. (2010). Minimizing the diameter of a network using shortcut edges. In *Proceedings of the 12th Scandinavian conference on Algorithm Theory*. Springer.

DeNisco, A. (2015). School bus driver shortage drives new incentives. `https://www.districtadministration.com/article/school-bus-driver-shortage-drives-new-incentives`.

Derrible, S. (2012). Network centrality of metro systems. *PLOS One*, 7(7).

Derrible, S. and Kennedy, C. (2010a). Applications of graph theory and network science to transit network design. *Transport Reviews*, 4:498–519.

Derrible, S. and Kennedy, C. (2010b). The complexity and robustness of metro networks. *Physica A: Statistical Mechanics and its Applications*, 389(17):3678–3691.

Dickens, M. (2016). Public transportation ridership report. Technical report, American Public Transportation Association.

Donoso, Y. and Fabregat, R. (2007). *Multi-Objective Optimization in Computer Networks Using Metaheuristics*. Auerbach Publications.

Ellens, W. and Kooij, R. E. (2013). Graph measures and network robustness. *arXiv*.

Erdös, P. and Rényi, A. (1959). On random graphs. *Publicationes Mathematica*, 6:290–297.

Eubank, S., Guclu, H., Kumar, V. S. A., Marathe, M. V., Srinivasan, A., Toroczkai, Z., and Wang, N. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, 429:180–184.

Floriani, L. D., Falcidieno, B., and Pienovi, C. (1985). Delaunay-based representation of surfaces defined over arbitrarily shaped domains. *Computer Vision, Graphics, and Image Processing*, 32(1):127–140.

Forester, J. F. (1999). *The Deliberative Practitioner: Encouraging Participatory Planning Processes*. MIT Press.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41.

Garrison, W. L. and Marble, D. F. (1958). Analysis of highway networks: A linear programming formulation. *Proceedings of the Highway Research Board*, 37:1–17.

Gavrilets, S., Auerbach, J., and van Vugt, M. (2016). Convergence to consensus in heterogeneous groups and the emergence of informal leadership. *Scientific Reports*, 6.

Golden, B. L. and Skiscim, C. C. (1986). Using simulated annealing to solve routing and location problems. *Naval Research Logistics Quarterly*, 33(2):261–279.

Grafova, I. (2008). Overweight children: assessing the contribution of the built environment. *Preventative Medicine*, 47(3):304–308.

Greiner, R. (1992). Probabilistic hill-climbing: Theory and applications. In *Proceedings of the Ninth Canadian Conference on Artificial Intelligence*. AAAI.

Gu, W., Wang, X., and McGregor, S. E. (2010). Optimization of preventive health care facility locations. *International Journal of Health Geographies*, 9(17).

Hagberg, A. (2017). NetworkX. `https://networkx.github.io/`.

Hales, C. M., Carroll, M. D., Fryar, C. D., and Ogden, C. L. (2017). Prevalence of obesity among adults and youth: United States, 2015–2016. Technical report, National Center for Health Statistics.

Horn, R. A. and Zhang, F. (2005). Basic properties of the Schur complement. In Zhang, F., editor, *The Schur Complement and Its Applications*, pages 17–46. Springer.

Ibeas, A., dell'Olio, L., Alonso, B., and Sainz, O. (2010). Optimizing bus stop spacing in urban areas. *Transportation Research Part E: Logistics and Transportation Review*, 46:446–458.

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37:547–579.

Ji, Y. and Geroliminis, N. (2012). On the spatial partitioning of urban transportation networks. *Transportation Research Part B: Methodological*, 46(10):1639–1656.

Jiang, Z., Liang, M., and Guo, D. (2011). Enhancing network performance by edge addition. *International Journal of Modern Physics C*, 22(11):1211–1226.

Johnson, D. S., Aragon, C. R., McGeoch, L. A., and Schevon, C. (1989). Optimization by simulated annealing: An experimental evaluation; part 1, graph partitioning. *Operations Research*, 37(6):865–892.

Johnson, S. (2007). *The Ghost Map: The Story of London's Most Terrifying Epidemic-and How It Changed Science, Cities, and the Modern World*. Riverhead Books.

Jun-qiang, L., Long-hai, Y., Liu, W., and Zhao, L. (2017). Measuring road network vulnerability with sensitivity analysis. *PLOS One*, 12.

Kansky, K. and Danscoine, P. (1989). Measures of network structure. *Flux*, pages 89–121.

Karduni, A., Kermanshah, A., and Derrible, S. (2016). A protocol to convert spatial polyline data to network formats and applications to world urban road networks. *Scientific Data*, 3.

Kieu, L. M., Bhaskar, A., and Chung, E. (2015). A modified density-based scanning algorithm with noise for spatial travel pattern analysis from smart card AFC data. *Transportation Research Part C: Emerging Technologies*, 58:193–207.

Kim, H. (2012). p-Hub protection models for survivable hub network design. *Journal of Geographic Systems*, 14(4):437–461.

Kim, H., Kim, C., and Chun, Y. (2016a). Network reliability and resilience of rapid transit systems. *The Professional Geographer*, 68(1):53–65.

Kim, H. and O'Kelly, M. E. (2009). Reliable p-hub location problems in telecommunication networks. *Geographical Analysis*, 41(3):283–306.

Kim, H. and Ryerson, M. S. (2017). The q-Ad hoc hub location problem for multi-modal networks. *Networks and Spatial Economics*, 17(3):1015–1041.

Kim, H. J. and Lee, C. (2016). Does a more centrally located school promote walking to school? Spatial centrality in school-neighborhood settings. *Journal of Physical Activity and Health*, 13:481–487.

Kim, K., Dean, D., Kim, H., and Chun, Y. (2016b). Spatial optimization for regionalization problems with spatial interaction: a heuristic approach. *International Journal of Geographic Information Science*, 30(3):451–473.

Kinney, R., Crucitti, P., Albert, R., and Latora, V. (2005). Modeling cascading failures in the north american power grid. *The European Physical Journal B*, 46(1):101–107.

Kirkpatrick, S. (1984). Optimization by simulated annealing: Quantitative studies. *Journal of Statistical Physics*, 34(5/6):975–986.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220:671–680.

Klemm, K. and Eguílez, V. M. (2002). Growing scale-free networks with small-world behavior. *Physical Review E*, 65.

Knox County Board of Education (2018). Knox County Schools general purpose school fund fy 2019 recommended budget. Technical report, Knox County Schools.

Knox County Schools Transportation Department (2009). Bus rider eligibility: Parent Responsibility Zone (PRZ). Technical report, Knox County Schools.

Knoxville Area Association of Realtors (2017). Walkability survey. Technical report, Knoxville Area Association of Realtors.

Kooij, R. E., Schumm, P., and Scoglio, C. (2008). A new metric for robustness with respect to virus spread. *arXiv*.

Kouri, C. (1999). Wait for the bus: How low country school site selection and design deter walking to school and contribute to urban sprawl. Technical report, South Carolina Coastal Conservation League.

Kovacs, I. A. and Barabási, A. (2015). Network science: Destruction perfected. *Nature*, 524:38–39.

Kulkarni, D., Schmidt, D., and Tsui, S. (1999). Eigenvalues of tridiagonal pseudo-Toeplitz matrices. *Linear Algebra and its Applications*, 297(1-3):63–80.

Larsen, K., Gilliland, J., Hess, P., Tucker, P., Irwin, J., and He, M. (2009). The influence of the physical environment and sociodemographic characteristics on children's mode of travel to and from school. *American Journal of Public Health*, 99(3):520–526.

Larson, R. C. and Odoni, A. R. (1981). *URBAN OPERATIONS RESEARCH*. Prentice-Hall.

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

Levinson, D. (2005). The rational locator reexamined: Are travel times still stable? *Transportation*, 32(2):187–202.

Li, Y. and Kim, H. (2014). Assessing survivability of the Beijing subway system. *International Journal of Geospatial and Environmental Research*, 1(1).

Liebeherr, J. and Nahas, M. (2001). Application-layer multicast with Delaunay triangulations. In *Proceedings of the Global Telecommunications Conference, GLOBECOM*. IEEE.

Linehan, J., Gross, M., and Finn, J. (1995). Greenway planning: developing a landscape ecological network approach. *Landscape and Urban Planning*, 33(1-3):179–193.

Lubans, D. R., Boreham, C. A., Kelly, P., and Foster, C. E. (2011). The relationship between active travel to school and health-related fitness in children and adolescents: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 8(5).

Luce, R. and Perry, A. (1949). A method of matrix analysis of group structure. *Psychometrika*, 14(1):95–116.

Lyer, S., Killingback, T., Sundaram, B., and Wang, Z. (2013). Attack robustness and centrality of complex networks. *PLOS One*, 8.

Magnanti, T. L. and Wong, R. T. (1984). Network design and transportation planning: Models andalgorithms. *Transportation Science*, 18(1):1–55.

Mattsson, L. and Jenelius, E. (2015). Vulnerability and resilience of transport systems – A discussion of recent research. *Transportation Research Part A: Policy and Practice*, 81:16–34.

Mazloumi, E., Rose, G., Currie, G., and Moridpour, S. (2011). Prediction intervals to account for uncertainties in neural network predictions: Methodology and application in

bus travel time prediction. *Engineering Applications of Artificial Intelligence*, 24(3):534–542.

McDonald, N. C. (2007). Active transportation to school: Trends among U.S. schoolchildren, 1969-2001. *American Journal of Preventive Medicine*, 32:509–516.

McDonald, N. C. and Aalborg, A. E. (2009). Why parents drive children to school: Implications for Safe Routes to School programs. *Journal of the American Planning Association*, 75(3):331–342.

McDonald, N. C., Brown, A. L., Marchetti, L. M., and Pedroso, M. S. (2011). U.S. school travel, 2009 an assessment of trends. *American Journal of Preventive Medicine*, 41:146–151.

Meguerdichian, S., Koushanfar, F., Qu, G., and Potkonjak, M. (2001). Exposure in wireless Ad-Hoc sensor networks. In *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking*, MobiCom '01, pages 139–150. ACM.

Mendoza, J. A., Cowan, D., and Liu, Y. (2014). Predictors of children's active commuting to school: an observational evaluation in 5 U.S. communities. *Journal of Physical Activity & Health*, 11(4):729–733.

Mendoza, J. A., Watson, K., Nguyen, N., Cerin, E., Baranowski, T., and Nicklas, T. A. (2011). Active commuting to school and association with physical activity and adiposity among US youth. *Journal of Physical Activity & Health*, 8(4):488–495.

Meyerson, A. and Tagiku, B. (2009). Minimizing average shortest path distances via shortcut edge addition. In Dinur, I., Jansen, K., Naor, J., and Rolim, J., editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 272–285. Springer.

Mladenović, N., Brimberg, J., Hansen, P., and Moreno-Pérez, J. A. (2007). The p-median problem: A survey of metaheuristic approaches. *European Journal of Operational Research*, 179:927–939.

Mladenović, N. and Hansen, E. (1997). Variable neighborhood search. *Computers and Operations Research*, 24(11):1097–1100.

Morone, F. and Makse, H. A. (2015). Influence maximization in complex networks through optimal percolation. *Nature*, 524:65–68.

Nader, P. R., Bradley, R. H., and Houts, R. M. (2008). Moderate-to-vigorous physical activity from ages 9 to 15 years. *Journal of the American Medical Association*, 300:295–305.

National Center for Statistics and Analysis (2016). 2015 motor vehicle crashes. Technical report, National Highway Traffic Safety Administration.

Nelson, D. O. and Streit, A. E. (2011). Rail transit safety: An empirical evaluation. In *American Public Transportation Association: Annual Rail Conference*.

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences USA*, 103(23):8577–8582.

Newman, M. E. J. (2008). Mathematics of networks. In Blume, L. and Burlauf, S., editors, *The New Palgrave Encyclopedia of Economics*. Palgrave Macmillan, Basingstoke, 2nd edition.

Newman, M. E. J. (2010). *Networks: An introduction*. Oxford University Press.

Nichols, S. and Crompton, J. L. (2005). The impact of greenways on property values: Evidence from Austin, Texas. *Journal of Leisure Research*, 37(3):321–341.

Office of Children's Health Protection (2011). School siting guidlines. Technical report, United States Environmental Protection Agency.

O'Kelly, M. E. (2015). Network hub structure and resilience. *Networks and spatial economics*, 15(2):235–251.

O'Kelly, M. E. and Bryan, D. L. (1998). Hub location with flow economies of scale. *Transportation Research Part B: Methodological*, 32:605–616.

Oliver, I. M., Smith, D. J. D., and Holland, R. C. J. (1987). Study of permutation crossover operators on the traveling salesman problem. In *Proceedings of the Second International Conference on Genetic Algorithms on Genetic algorithms and their application*, pages 224–230. MIT.

Oliver, L. N., Schuurman, N., and Hall, A. W. (2007). Comparing circular and network buffers to examine the influence of land use on walking for leisure and errands. *International Journal of Health Geographics*, 6(41).

Oreskovic, N., Kuhlthau, K., Romm, D., and Perrin, J. (2009). Built environment and weight disparities among children in high- and low-income towns. *Academic Pediatrics*, 9(5):315–321.

Pablo-Martì, F. and Sánchez, A. (2017). Improving transportation networks: Effects of population structure and decision making policies. *Scientific Reports*, 7.

Pedestrian and Bicycle Information Center (2010). Safe routes to school guide. Technical report, University of North Carolina Highway Safety Research Center.

Perry, C. A. (1929). The neighborhood unit, a scheme of arrangement for the family-life community. In *Monograph 1 of Regional Plan of N.Y. Regional survey of N.Y. and its environs*, volume 7, pages 2–140. Reprinted Routledge/Thoemmes, London, 1998, 25-44.

Prettejohn, B., Berryman, M., and McDonnell, M. (2011). Methods for generating complex

networks with selected properties for simulations: a review and tutorial for neuroscientists. *Frontiers in Computational Neuroscience*, 5.

Price, A. E., Pluto, D. M., Ogoussan, O., and Banda, J. A. (2011). School administrators' perceptions of factors that influence children's active travel to school. *Journal of School Health*, 81(12):741–748.

Resende, M. G. C. and Pardalos, P. M., editors (2006). *Handbook of Optimization in Telecommunications*. Springer.

Robert Wood Johnson Foundation (2009). Active education: Physical education, physical activity and academic performance. Technical report, Robert Wood Johnson Foundation.

Rodriguez-Nunez, E. and Garcia-Palomares, J. (2014). Measuring the vulnerability of public transport networks. *Journal of Transport Geography*, 35:50–63.

Russell, S. J. and Norvig, P. (2004). *Artificial Intelligence: A Modern Approach*. Prentice Hall.

Schrijver, A. (2002). On the history of the transportation and maximum flow problems. *Mathematical Programming*, 91(3):437–445.

Scott, D. M., Novak, D. C., Aultman-Hall, L., and Guo, F. (2006). Network robustness indicator: A new method for identifying critical links and evaluation the performance of transport network. *Journal of Transport Geography*, 14(3):215–227.

Sirard, J. R. and Slater, M. E. (2008). Walking and bicycling to school: A review. *American Journal of Lifestyle Medicine*, 32:372–396.

Sohn, J. (2006). Evaluating the significance of highway network links under the flood damage: An accessibility approach. *Transportation Research Part A: Policy and Practice*, 40:491–506.

Spence, J., Cutumisu, N., Edwards, J., and Evans, J. (2008). Influence of neighbourhood design and access to facilities on overweight among preschool children. *International Journal of Pediatric Obesity*, 3(2):109–116.

Steffen, B. and Seyfried, A. (2010). Methods for measuring pedestrian density, flow, speed and direction with minimal scatter. *Physica A: Statistical Mechanics and its Applications*, 389(9):1902–1910.

Taeffe, E. J., Gauthuer, H. L., and O'Kelly, M. E. (1996). *Geography of Transportation*. Prentice Hall, 2nd edition.

Taylor, M. A. P. and Susilawati (2012). Remoteness and accessibility in the vulnerability analysis of regional road networks,. *Transportation Research Part A: Policy and Practice*, 46:761–771.

Thaler, R. H. and Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.

The National Center for Safe Routes to School (2016). Trends in walking and bicycling to school from 2007 to 2014. Technical report, U.S. Department of Transportation Federal Highway Administration.

Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B*, 63(2):411–423.

Transportation Consultants (2014). Walk-to-school prioritization analysis for the schools of Knox County, Tennessee. Technical report, Knox County Department of Engineering & Public Works.

Trasande, L. and Chatterjee, S. (2012). The impact of obesity on health service utilization and costs in childhood. *Obesity*, 17(9):1749–1754.

Trasande, L., Liu, Y., Fryer, G., and Weitzman, M. (2009). Effects of childhood obesity on hospital care and costs, 1999-2005. *Health Affairs*, 28(4):751–760.

Turrell, G., Hewitt, B., Rachele, J., Giles-Corti, B., Busija, L., and Brown, W. (2018). Do active modes of transport cause lower body mass index? Findings from the HABITAT longitudinal study. *Journal of Epidemiology and Community Health*, 72:294–301.

United States Census Bureau (2016). Current population survey.

Varga, R. (2002). *Matrix Iterative Analysis*. Springer-Verlag, 2nd edition.

von Ferber, C., Berche, B., Holovatch, T., and Holovatch, Y. (2012). A tale of two cities: Vulnerabilities of the London and Paris transit networks. *Journal of Transportation Security*, 5(3):199–216.

Walsh, T. (1999). Search in a small world. In Dean, T., editor, *Proceedings of the 16th International Joint Conference on Artificial Intelligence*. Morgan Kaufmann.

Wang, Y., Chakrabarti, D., Wang, C., and Faloutsos, C. (2003). Epidemic spreading in real networks: An eigenvalue viewpoint. In *22nd International Symposium on Reliable Distributed Systems*.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58:236–244.

Watts, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences USA*, 99(9):5766–5771.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393:440–442.

Welk, G. (2009). Cardiovascular fitness and body mass index are associated with academic achievement in schools. Technical report, Cooper Institute.

Whitley, D., Starkweather, T., and Bogart, C. (1990). Genetic algorithms and neural networks: optimizing connections and connectivity. *Parallel Computing*, 14(3):347–361.

Wilson, R. J. (1986). An eulerian trail through Königsberg. *Journal of Graph Theory*, 10(3):265–275.

Witt, W. P., Weiss, A. J., and Elixhauser, A. (2014). Overview of hospital stays for children in the United States, 2012. Technical report, Agency for Healthcare Research and Quality.

Wu, F., Huberman, B. A., Adamic, L. A., and Tyler, J. R. (2004). Information flow in social groups. *Physica A: Statistical Mechanics and its Applications*, 337(1-2):327–335.

Wu, J., Tan, Y., Deng, H., Li, Y., Liu, B., and Lv, X. (2011). Spectral measure of robustness in complex networks. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 41(6):1244–1252.

Yamada, I. and Thill, J. (2006). Local indicators of network-constrained clusters in spatial point patterns. *Geographical Analysis*, 39:268–292.

Yao, B., Ping, H., Lu, X., Gaoa, J., and Zhang, M. (2014). Transit network design based on travel time reliability. *Transportation Research Part C: Emerging Technologies*, 43:233–248.

Zhao, D., Wang, L., Zhi, Y., and Wang, J. Z. (2016). The robustness of multiplex networks under layer node-based attack. *Scientific Reports*, 6.

Vita

Jeremy Auerbach was born in New Jersey and spent his childhood in the Philadelphia area. He is a veteran of the United States Air Force, where he served two tours in Iraq, a tour in Afghanistan, and was part of the first military response to New Orleans after Hurricane Katrina. He has undergraduate degrees in Humanities and Applied Mathematics and a Master's degree in Applied Mathematics, with a research focus in mathematical and evolutionary biology. His interest in interdisciplinary research led him to graduate studies in the Department of Geography at the University of Tennessee where he focused his work on the urban planning and transportation. Along with his studies he held graduate fellowships with the National Institute of Mathematical and Biological Synthesis and the United States Census Bureau. He has also interned with an Advanced Research and Methods team at the United States Government Accountability Office, the Geographic Information Science and Technology group at the Oak Ridge National Laboratory, and the Computer Science Department at the Georgia Institute of Technology. After completion of his PhD in Geography, Jeremy pursued a postdoctoral research program in the Department of Epidemiology at Colorado State University where he evaluated the impact of environmental pollutants on child health and behavior.