City Network Resilience Quantification under Systemic Risks: A Hybrid Machine Learning-Genetic Algorithm Approach

City Network Resilience Quantification under Systemic Risks: A Hybrid Machine Learning-Genetic Algorithm Approach

By

RASHA AHMED MAHMOUD ALI HASSAN

B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Master of Applied Science

McMaster University

October 2020

Master of Applied Science (2020) McMaster University

(Civil Engineering) Hamilton, Ontario

TITLE: City Network Resilience Quantification under Systemic Risks: A Hybrid Machine Learning-Genetic Algorithm Approach

AUTHOR: Rasha Hassan

B.Sc. (Ain Shams University)

SUPERVISORS: Dr. Wael El-Dakhakhni

Dr. Mohamed Ezzeldin

NUMBER OF PAGES: 68

ABSTRACT

Disruptions due to either natural or anthropogenic hazards significantly impact the operation of critical infrastructure networks because they may instigate network-level cascade (i.e., systemic) risks. Therefore, quantifying and enhancing the resilience of such complex dynamically evolving networks ensure minimizing the possibility and consequences of systemic risks. Focusing only on robustness, as one of the key resilience attributes, and on transportation networks, key critical infrastructure, the current study develops a hybrid complex network theoretic-genetic algorithms analysis approach. To demonstrate the developed approach, the robustness of a city transportation network is quantified by integrating complex network theoretic topology measures with a dynamic flow redistribution model. The network robustness is subsequently investigated under different operational measures and the corresponding absorptive capacity thresholds are quantified. Finally, the robustness of the network under different failure scenarios is evaluated using genetic algorithms coupled with k-means clustering to classify the different network components. The hybrid approach developed in the current study is expected to facilitate optimizing potential systemic risk mitigation strategies for critical infrastructure networks under disruptive events.

KEYWORDS: Absorptive Capacity; Cascade Failure; City Networks; Clustering Analysis; Complex Network Theory; Genetic Algorithms; Network Topology; Resilience; Robustness.

To

My amazing husband Mohamed,

and my lovely daughters

Marium and Hagar

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my supervisors Dr. Wael El-Dakhakhni and Dr. Mohamed Ezzeldin for the continuous support, help, guidance, and care throughout my study. It was a great opportunity to work with such amazing supervisors who taught me how research should be conducted. I would also like to thank Dr. Ahmed Yosri for his efforts, guidance, and valuable advices. I want to also express my special gratitude to Dr. Michael Isaacson, and Dr. Tarek Sayed from the University of British Columbia for their support and encouragement.

I am very grateful to my parents, my sister Reham and my brothers Ahmed and Mahmoud for their continuous support. At last, I would like to thank my husband Mohamed, who was always there by my side at good and bad times, and my lovely kids Marium and Hagar for being patient and helpful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOTATION AND ACRONYMS

The following Notations and Acronyms are used in this study:

Notation:		**List of Acronyms**	
k_i	Degree of centrality	BTN	Bus transit network
P_k	Degree centrality distribution	LBTN	Local bus transit network
k_i^{out}	Outward degree	MLBTN	Minneapolis local bus transit network
k_i^{in}	Inwards degree	APC	Automated passenger count
S_i	Node strength	AC	Absorptive capacity
w_{ij}	Weight of the link between nodes *i* and *j*	R_N	Network robustness
x_i	Node state	CNT	Complex network theory
N_f	Number of removed nodes	R	Route
γ	Degree exponent	PV	Passenger volume
f_c	Critical failure threshold	RC	Route capacity
ξ_1	The topological coupling coefficient	R	Pearson's correlation coefficient
ξ_2	The flow coupling coefficient	GA	Genetic algorithm
N_{FR}	Total number of failed nodes	f_c	Critical failure threshold
N_{TR}	Total number of nodes along routes	$f_{c\,(th)}$	Theoretical Critical failure threshold
BC	Betweenness centrality	$f_{c\text{-}L}$	Theoretical Critical failure threshold - link failure
CC	Closeness centrality	$f_{c\text{-}N\,(act)}$	Theoretical Critical failure threshold – Node failure
EC	Eigen centrality	$f_{c(act)}$	Actual Critical failure threshold
A_{kj}	The adjacency matrix	BC	Betweenness centrality
Q	Modularity class	CC	Closeness centrality
S_R^{out}	Weighted outdegree	EC	Eigen centrality
		A_{kj}	The adjacency matrix
		Q	Modularity class
		S_R^{out}	Weighted outdegree

1. INTRODUCTION

1.1. BACKGROUND AND MOTIVATION

The economy, security, public health, and safety of cities are continuously impacted by their critical infrastructure networks (e.g., power, transportation, and water). This impact is mainly because such networks may face various types of disruptions including planned incidents and terroristic attacks, natural disasters, extreme weather events, and other operational incidents that may result in network-level cascade (i.e., systemic) risks (Ezzeldin and El-Dakhakhni 2019; Yassien et al. 2020). For example, in 2012, hurricane Sandy impacted several infrastructure networks in New York and New Jersey cities including their subway and bus networks that serve around 7.5 million passengers per day (Schwartz et al. 2014). The subway tunnels were flooded which subsequently induced excess demands on bus networks and other transportation modes (Kaufman et al. 2012). Such unexpected demands resulted in delays, interrupted services, and significant crowdedness (Schwartz et al. 2014). The damages of the subway network only reached $4.5 billion, whereas the overall economic damages of Hurricane Sandy were estimated to be more than $40 billion (Progressive Railroading 2016). Such systemic risks and the associated impacts on city critical infrastructure networks highlight the importance of network resilience (the ability to withstand, adapt, and recover from disruptions).

City critical infrastructure networks typically comprise multiple subsystems that are both complex (interdependent) and dynamic (i.e., evolve with time) (Salama et al. 2020).

Among these networks, public transit is the backbone of city transportation systems as it provides the basic mobility and accessibility services that ensure equity and inclusion of community members (Litman 2014). For example, public transit networks in the United States and Canada have served around 10 and 3 billion trips during 2019, respectively (APTA 2019). In addition, public transit networks support cities' economy and urban expansion, and they also induce transit-oriented developments. Such developments contribute towards multimodal transit neighbourhoods that subsequently reduce traffic congestions, car emissions, and accident rates (Ceder 2016). Moreover, public transit networks play a vital role in emergency responses and evacuation schemes during natural disasters and all other major incidents (Humphrey 2008).

Among the different multimodal transit networks (e.g., light rail, subway, commuter rail, rapid transit, and ferry boats), an urban local bus transit network of a city is a vital component as it serves denser regions of the downtown core, critical land use and the main attractions in this city (Martin et al. 1998; Metropolitan Council 2009). For instance, in the United States in 2019, the Minneapolis Metro transit agency has served an average of 249,300 trips during the weekdays, where the urban local bus transit network contributes towards around 68.8% of the total served trips (Dickens 2011; Dickens et al. 2012; APTA 2019). Therefore, as discussed earlier, any disruptions to bus transit networks can cause a significant reduction in the overall system performance, thus impacting the quality of service provided as well as the reliability of bus transit networks (Kim et al.

2016). This can, in turn, lead to severe adverse socio-economic impacts (Levinson et al. 2011).

Disruptions to urban bus transit networks include: infrastructure maintenance and corresponding road closures, accidents or traffic congestions, unexpected weather and climatological incidents, terroristic attacks, and strikes (Weilant et al. 2019). Examples of recent disruptions that significantly influenced different urban public transit networks in North America include: *i*) the 2020 demonstrations across the United States (e.g., in Minneapolis, Chicago, Miami, Los Angeles) due to the death of George Floyd, which resulted in rerouting several bus lines, cancelling hundreds of trips, reducing the frequency of the bus transit system, interrupting and suspending the public transit service for several days, and stranding thousands of commuters (Fox 9, 2020); and *ii*) the shutdown of Toronto subway in 2020 due to the derailment of a train, which triggered massive crowdedness at several bus stops causing a chaotic situation to the bus transit network in Toronto. Specifically, massive crowdedness occurred at several bus stops and thousands of commuters waited for the bus service up to one hour in chilly weather conditions as a result of the subway service disruption. This chaotic situation lasted for four hours during the morning peak time until it was finally resolved (CTV News 2020).

The aforementioned disruptions, together with others not mentioned herein, have significantly impacted key components of urban public transit networks (e.g., stations and routes) either directly or indirectly. This situation highlights the need for quantifying and enhancing the resilience of urban local bus transit networks in an effort to withstand and

recover from the sequential disruptions of stations and routes with minimum losses in their performance levels. This will ensure reliable services during both normal conditions and disruptions, improve the riding experience, and increase ridership rates.

1.2. RESILIENCE OF PUBLIC TRANSIT NETWORKS

There is a lack of consensus among the definition of resilience within the transportation engineering field (Tamvakis and Xenidis 2012; Hosseini et al. 2016; Matherly et al. 2017; Haggag et al. 2020). However, according to the U.S. Department of Transportation, FHWA 2014, resilience of transportation networks is commonly defined as *"the ability to anticipate, prepare for, adapt and absorb perturbation and withstand, respond to, and recover rapidly from disruptions"* (FHWA 2014). Therefore, in transportation networks, resilience can be represented by different operational measures, including hours of congestion, travel time index, volume of congestion, and the optimal spare capacity (Hollnagel et al. 2008; Weilant et al. 2019). The spare capacity can be further categorized into three different capacities, namely *absorptive*, *adaptive*, and *restorative* (Hosseini et al. 2016; Weilant et al. 2019) **Absorptive capacity** is the ability of a system to absorb and withstand the impacts and consequences of disruptive events, **adaptive capacity is** the potential of a system to change in response to disruptions in order to maintain normal functions, and **restorative capacity** is the ability of a system to achieve quick recovery and maintain its normal conditions at minimum damage after disruptions (Norris et al. 2008; Turnquist et al. 2013). Resilience can also be represented by the **4R** measures: *robustness*,

rapidity, *redundancy*, and *resourcefulness* (Bruneau et al. 2003). **Robustness** is defined as the ability of a system to withstand sudden disruptions with minimum loss in its performance level; **redundancy** is the ability of a system to meet the required functions during disruptions with the presence of alternatives for the failed components; **resourcefulness** represents the capacity of resources to define problems and recognize priorities during disruptions; and **rapidity** reflects the ability of a system to timely satisfy priorities and accomplish goals. As such, based on the aforementioned definitions, the robustness is related to the absorptive capacity, and both measures are subsequently used in the current study to quantify the resilience of local bus transit networks in an effort to minimize the impacts of disruptions and the corresponding induced systemic risks (Hosseini and Barker 2016; Reggiani 2013).

1.3. ROBUSTNESS OF PUBLIC TRANSIT NETWORKS

Complex network theory (CNT) represents a computationally efficient technique that enables the simulation of complex networks (Barabási 2015). In CNT, networks are simulated as *nodes* (i.e., representing the network components) and *links* (i.e., representing the interdependency between these components). The CNT has been applied in previous studies to investigate the robustness of a wide spectrum of transportation networks, including public transit networks. Such studies can be broadly classified into two main groups: 1) static robustness (e.g., Jiang et al. 2004; Wang et al. 2011; Zou et al. 2013); and

2) dynamic robustness (Marchiori et al. 2000; Barrat et al. 2004; Soh et al. 2010; Sullivan et al. 2010; Cats et al. 2015; Huang et al. 2015).

1.3.1. STATIC ROBUSTNESS

The static robustness of public transit networks relies on investigating the topological characteristics of such networks with an assumption that the failure of a single component (e.g., a station or a route) does not trigger a redistribution of its load demands (e.g., number of passengers) throughout the network. For example, the robustness of 33 metro networks across the world was investigated by Derrible et al. (2010). The authors demonstrated that creating transfer stations to provide alternative routes during disruptive events is essential for improving the network robustness. Ren et al. (2016) analyzed also the urban bus transit network of Shenyang City through three different static networks representing the stops, transfers, and lines of the network. The robustness of this network was quantified based on interactions between stop-line and line-transfer networks. The authors proposed an emergency strategy that can be applied to cope with the traffic congestion that may occur. Recently, Liu et al. (2017) modeled the Beijing-Tianjin-Hebei rail transit network as a network of high exposure to natural and targeted attacks. The study highlighted the importance of both creating rescue stations and increasing redundant routes to enhance network robustness. More recently, the robustness of 40 bus transit networks in Canada was quantified and several robustness indices are evaluated by Abdelaty et al. (2020). The results highlighted a significant level of contradictions between the different robustness

indices. In general, the above studies recommended applying other robustness assessment approaches to achieve realistic and practical results, as will be discussed next.

1.3.2 DYNAMIC ROBUSTNESS

The dynamic robustness of public transit networks considers not only the network topology but also the redistribution of the load demands (i.e., passengers) following the failure of some network components that may trigger cascade failures to other components. Various models have been utilized in previous studies to quantify the dynamic robustness of several infrastructure networks including the double value impact models (Watts 2011), the optimal power flow models (Cupac et al. 2013), the sand pile models (Bak et al. 1987; Turalska et al. 2019), the load capacity models (Huang et al. 2018; Motter et al. 2002), and the coupled map lattice models (García-Morales 2016; Kaneko 1992). However, within the context of public transit networks, the dynamic robustness is quantified through either the load-capacity models (Huang et al. 2018) or the coupled map lattice models (e.g., Huang et al. 2015; Sun et al. 2018). The load-capacity model was first introduced by Motter et al. (2002) to model the cascade failures of complex networks based on the assumption that a single unit of weight is exchanged between each pair of nodes overtime through the shortest path connecting these nodes. In public transit networks, the load of a station is defined as the total number of shortest paths that pass through the station, while the capacity of a station is constrained by the maximum load that this station can accommodate. If a station fails, it is removed from the network and its loads are redistributed to the neighboring

nodes. If the new loads exceed the corresponding capacities, the neighbouring nodes fail, and a series of cascade failures is triggered. On the other side, the coupled map lattice (CML) models follow a different concept in modelling cascade failures of complex networks. The CML model was first introduced by Kaneko (1992) to model the spatiotemporal chaos and pattern formation in fluids. The basic assumption of the CML model is that the spatiotemporal chaotic failure is mainly attributed to both the topology of the network as well as the redistribution of the weights on links. The applications of the CML model has been subsequently extended to other fields such as biology, mathematics, and engineering (Chazottes et al. 2005).

The CML models have been continuously modified to reflect realistic characteristics of transportation networks such that the cascade failure phenomena of these networks can be effectively represented. In such models, cascade failures consider the state, degree and strength of both failed stations and neighbouring stations as well as the weights and capacities of links connected to such stations. Therefore, several CML models have been utilized in previous studies to provide a comprehensive approach and better representation of cascade failures by considering key factors in public transit networks. For instance, Huang et al. (2015) adopted a CML model to assess the robustness of the Beijing bus and rail transit networks considering their passenger flows. Furthermore, Sun et al. (2018) assessed the robustness of an urban rail transit network using an integrated CML-passenger flow redistribution model, where cascade failures were attributed to topological characteristics and passenger flow redistributions.

1.4. RESEARCH OBJECTIVES

Most of the previous CML model-related studies simulated transit networks as undirected networks. However, this can result in misleading conclusions as load demands may vary over directions and centrality measures (e.g., node degree and strength) are also significantly different between directed and undirected networks (Barabási 2015; White et al. 1994). Therefore, limited studies have applied the CML model to quantify the robustness of directed public transit networks. For example, Shen et al. (2019) developed a modified directed version of the CML model to assess the robustness of the Nanjing metro transit network. However, the cascade failure conducted in this study has been limited to node failure scenarios only. On the other hand, Zhang et al. (2019) investigated the impact of route failures on weighted bus transit networks considering the dynamic load redistribution and link prediction method. The authors simulated the underlying networks in an R-space framework, where routes were represented as nodes (i.e., a node failure represented a route failure) and links existed only if there was a connection between these routes. While this representation can facilitate analyzing the dynamic cascade failures, link weights and capacities as well as directed centrality measures were not captured for critical routes. As such, further investigations are still needed in order to: *i*) assess the robustness of directed public transit networks based on different failure scenarios (i.e., nodes, links, and routes); and *ii*) relate the robustness thresholds to operational bus transit measures through a practical mapping approach.

In this respect, the current study aims at developing an approach to quantify the robustness of city networks under systemic risks and subsequently identify the corresponding critical components using a hybrid approach (i.e., machine learning, genetic algorithm and complex network theory). To demonstrate the practical use of the developed modelling approach, the Minneapolis local bus transit network is utilized, where CNT is employed first to obtain the topological measures of the network. A CML model coupled with a direction-based passenger flow redistribution model are subsequently developed based on such CNT measures. The robustness of the network is then quantified based on node, link and route failure scenarios. The robustness of the network is also investigated under different passenger volume to route capacity ratios and absorptive capacity thresholds are accordingly obtained. The robustness of the network under route failures is studied considering 43 different centrality/operation measures and a genetic algorithm is then coupled with a k-means clustering approach to categorize the network routes based on key network measures.

The current study is organized as follows; Section 2 introduces the methods and measures used to assess the robustness of bus transit networks, Section 3 contains a description of the case study and the data processing procedures, Section 4 includes the results of the robustness analysis and the corresponding coupled genetic algorithm-cluster analysis, and Section 5 provides the conclusions of the study and suggestions for future research.

2. METHODS AND MEASURES

This section highlights the approach developed to assess the robustness of bus transit networks. The developed approach is divided into three stages, as shown in Figure 1. In Stage one "*Data Processing and Model Development*", the underlying network is simulated using CNT, where stations and bus routes are represented by nodes and links, respectively. Accordingly, the passenger volumes and route capacities are attributed to each link based on automated passenger count (APC) data, thus constructing a directed weighted bus transit network. Stage two incorporates the "*Network Classification and Topological Characteristics*" and is based on node degree distribution and other centrality measures (e.g., betweenness and closeness). In Stage three "*Robustness Assessment*", the network robustness is quantified under systemic risks using a CML model that considers the redistribution of the passenger flow. The network robustness is evaluated considering the failure of its different components (i.e., nodes, links). The network robustness is also investigated under different passenger volume to route capacity ratios and the corresponding absorptive capacity thresholds are identified. Finally, using an integrated genetic algorithm-clustering approach, the network routes are categorized based on the network robustness under route failures.

2.1. COMPLEX NETWORK THEORETIC MEASURES

Disruptions of stations and bus routes cause significant changes in the network topological characteristics such as the node degree centrality, node strength, and link weights.

2.1.1. THE DEGREE CENTRALITY (k_i) AND DEGREE CENTRALITY DISTRIBUTION (P_k) (CONNECTEDNESS MEASURE)

k_i is defined as the total number of links connected directly to a node. For directed networks, the measure is calculated as the total outward, $k_{i_{out}}$, and inward, $k_{i_{in}}$, links connected directly to a node as follows (Opsahl et al. 2010; Barabási 2016):

$$k_i = k_{i_{out}} + k_{i_{in}} \quad (1)$$

$$k_{i_{out}} = \sum_{j=1}^{N} a_{ij} \quad (2)$$

$$k_{i_{in}} = \sum_{j=1}^{N} a_{ji} \quad (3)$$

where a_{ij} *and* a_{ji} are the adjacency matrix elements of the network given that $a_{ij} \neq a_{ji}$ and $a_{ii} = a_{jj} = 0$, i is the start station (source node), j is the end station (all other nodes connected to the source node), and N is the total number of nodes in the network.

The degree centrality distribution, P_k, represents the percentage of nodes with degree k and is calculated as follows (Albert et al. 2002):

$$P_{k_{out}} = N_{k_{out}} / N \quad (4)$$

$$P_{k_{in}} = N_{k_{in}} / N \quad (5)$$

Where $N_{k_{out}}$ and $N_{k_{in}}$ are the total number of nodes with k_{out} and k_{in}, respectively.

2.1.2. THE NODE STRENGTH (s_i) (PASSENGER FLOW MEASURE)

S_i represents the sum of passenger flow in links connected to a given node (Sun et al. 2018):

$$S_i = (\sum_{j \neq i,}^{N} w_{ij} + \sum_{j \neq i,}^{N} w_{ji}) / n \quad (6)$$

where n is number of nodes connected to node *i*, while w_{ij} and w_{ji} are sectional passenger flows (i.e., demands per hour) on links connected to node *i*. Such links are assigned directions based on the sequence of the stations along the routes.

2.1.3. CML MODEL

A modified CML model is developed in the current study to investigate the cascade failure within a weighted and directed urban local bus transit network. Specifically, the redistribution of the passenger flow is modified to match the characteristics of a directed network, where the state of node *i* is changed over time according to:

$$x_i(t+1) = \left|(1-\xi_1-\xi_2)f\left(x_i(t)\right) + \xi_1 \sum_{j=1\ j\neq i}^{N} a_{ij}\frac{f\left(x_j(t)\right)}{k(i)} + \xi_2 \sum_{j=1\ j\neq i}^{N} w_{ij}\frac{f\left(x_j(t)\right)}{s(i)}\right| + R \qquad (7)$$

where $x_i(t)$ and $x_i(t+1)$ are the node states at times *t* and *t+1*, respectively, $f\left(x_i(t)\right)$ is the local dynamic chaotic behavior of the node, ξ_1 is the topological coupling coefficient, ξ_2 is the flow coupling coefficient, w_{ij} is the passenger flow rate between nodes *i* and *j* and is defined by the passenger volume (*PV*), and *R* is the external perturbation severity. It is worth mentioning that ξ_1 and ξ_2 values are typically between 0 and 1 and their summation should be less than 1. In the current study, $\xi_1$and ξ_2 values are assumed to be 0.25 (A. Huang et al. 2015).

To utilize the CML model, each station *i* is assigned an initial random x_i between 0 and 1. The value of x_i indicates the transportation condition of the node, where higher values reflect worse conditions (Shen et al., 2019). A logistic chaotic map of $[f(x) = 4x(1-x)]$

is used in several related studies and subsequently used herein to assess the corresponding dynamic behavior pattern of the nodes (Sun et al. 2018; Shen et al. 2019). The external perturbation severity, *R*, reflects the effect and magnitude of the failure that may occur, where $R = 1$ is assigned to the failed stations, and a value of zero is assigned to the operating counterparts.

It is worth mentioning that if both x_i and $f(x_i)$ are within the open interval]0,1[and there is no external perturbation (i.e., $R=0$), the network keeps a permanent normal state. However, if a node/link fails at $t = l$, additional passenger volume is produced and propagates downstream. A cascade failure scenario may occur if the redistributed passenger flow results in additional node/link failures due to capacity constraint that is defined by the route capacity (*RC*).

2.2. FAILURE SCENARIOS

2.2.1. NODE AND LINK FAILURES

When a node fails at $t = l$, the node is removed and its load demand is then transferred to the neighbours (i.e., considering the direction) over the downstream outward links, as shown in Figure 2a. Such links fail when their *PV/RC* ratios are larger than one and they are subsequently removed from the network. Another step of redistribution is then applied to trigger the cascade failure process. To assess the robustness of a directed bus transit network under random node failures, *RC* is estimated for each link based on the

frequency of hourly operating busses and the desired bus occupancy. This enables including *RC* as a constraint that varies according to the link attributes.

On the other side, when a link fails at $t = l$, the associated *PV* is transferred to the destination node and subsequently distributed over the connected links, as shown in Figure 2b, using Equation 8. Such links may also fail when the corresponding load to capacity ratio is large than one. The state of the downstream nodes is also updated based on Equation 7. If an additional node fails due to the *PV* redistribution, this node is also removed from the network and another step of redistribution is applied, as discussed earlier.

$$w_{jh}^{new} = w_{jh} + w_{ij}\frac{w_{jh}}{S_j^{out}} \quad (8)$$

2.2.2. ROBUSTNESS INDEX FOR NODE AND LINK FAILURES

The network robustness can be assessed under random node and link failure scenarios, where a node/link is selected randomly to initiate the failure. The percentage of active nodes (R_N) is used to measure the network robustness and is calculated as follows:

$$R_N = \frac{N - N_f}{N} \quad (9)$$

where N_f is the number of failed nodes. A single value of R_N is estimated for each failure scenario (i.e., considering the cascade effect) using Equation 9 and is updated for each subsequent scenario. As the network robustness decreases with the increasing number of removed nodes/links, critical threshold values, $f_{c\text{-}N}$ and $f_{c\text{-}L}$, are defined to indicate the fraction of nodes and links, respectively, that result in a minimum R_N value.

2.2.3. ROUTE FAILURE

The impact of a route failure on the network robustness is represented by the route failure impact ratio (*RFIR*):

$$RFIR = N_{FR} / N_{TR} \quad (10)$$

where N_{FR} is the total number of failed nodes and N_{TR} represents the total number of nodes along the route.

2.2.3.1. CORRELATION ANALYSIS

Correlation analysis is generally used to assess the strength of the interrelationship between variable pairs (Pearson 1896). To date, several correlation measures have been developed (e.g., Pearson's, Spearman's, Kendall's tau correlation coefficient), each of which has its strengths and limitations (Winter et al. 2016). A correlation coefficient with a high absolute value indicates a strong linear relationship (or at least a strong monotonical variation) between a variable pair, and vice versa. In the current study, Pearson's correlation coefficient (*R*) was used to assess the strength of the linear relationship between the *RFIR* and the corresponding topology- and operation-related variables are presented in Table 1. Such an assessment can be used to identify the most influential variables that control the route failure impact and severity on the network robustness. It is noteworthy that low *R* values may reflect the interdependence between multiple variables rather than a weak interrelationship between variable pairs. As such, further investigations are necessary when low *R* values are encountered.

2.2.3.2. GENETIC ALGORITHMS AND CLUSTER ANALYSIS

Cluster analysis is typically used to group observations in an unsupervised manner based on the degree of similarity. Observations allocated to the same group should therefore share similar characteristics, whereas those allocated to different groups should be extremely different. Several similarity measures have been developed to date (e.g., Euclidean distance, Manhattan distance, Cosine Similarity), each of which is suitable for certain applications (Pandit et al., 2011). K-means clustering is a widely used clustering measure; however, it requires a prior definition of the number of clusters (*K*). A range of *K* values is therefore assumed, and the performance of the resulting models is assessed. The percentage of variance explained is the typical measure of the model performance and can be implicitly expressed through several alternatives (e.g., within cluster sum of squares, silhouette score).

A feature selection technique can be coupled with the clustering approach when the number of variables in the dataset is relatively large, the variables are highly interdependent, or a high degree of collinearity presents in the dataset. The feature selection step aims at removing irrelevant/redundant information such that the model estimates are unbiased (Celebi et al. 2013). Feature selection techniques are generally classified into: *i*) filters, where unimportant variables are removed based on prespecified evaluation criteria (e.g., the value of *R*); and *ii*) wrappers, where an optimization algorithm is introduced to select the optimum (or near optimum) feature subset (Saeys et al., 2007; Beniwal et al., 2012). In the current study, the genetic algorithm (GA) was used as a feature selection

technique and was then integrated with K-means clustering to classify the network routes based on their common characteristics.

GA are evolutionary algorithms that rely on the survival of the most adapted solution (Beniwal et al. 2012). Specifically, an initial set of solutions (initial population) is assumed, and a fitness function is used to assess the suitability of each set. A set of genetic operations (i.e., elitism, mutation, and crossover) is used within the GA to prevent the solution from being trapped in local minima or maxima. Such operations aim at improving the population fitness over generations (i.e., iterations) such that an optimal (or a near optimal) solution can be obtained. When GA is used as a feature selection technique within a cluster analysis, each individual (i.e., solution) is represented by a binary vector. A value of one is used to indicate that a variable is included for clustering, while a value of zero is used to indicate that the variable is not included. For example, an individual in the form of [1 0 1 0 0 0 0 … 0 1] indicates that the first, third, and last variable will be included for clustering, while all other variables will not be included. In the current study, the *K* value was between two and thirty and the GA was applied in each case to identify the optimal variable set used within the K-means clustering. The maximum number of clusters employed in the present study (i.e., 30) was selected to ensure the homogeneity of the resulting clusters (i.e., the clusters include a similar number of observations). The percentage of explained variance $\left(S_{exp}^2\right)$ was used as the fitness function, with the aim of maximizing its value for each *K* value. Accordingly, the GA-based clustering can be formulated mathematically as:

$$\max_{x_i} S_{exp}^2 \quad \text{subject to: } x_i \text{ is binary } \forall\, i \tag{11}$$

where $S_{exp}^2 = 1 - \frac{SS_c}{SS_T}$, SS_c is the within cluster sum of squared distance, and SS_T is the sum of squared distance between observation pairs (Jones and Harris 1999). As the optimum variable set differs based on the *K* value, the frequency of including a specific variable in the optimum set is used to reflect its importance for clustering the network routes. The optimum value of *K* and the optimum variable set were chosen based on assessing the performance of the optimal GA solution over the different *K* values.

3. DEMONSTRATION APPLICATION: MINNEAPOLIS LOCAL BUS TRANSIT NETWORK

In the current study, The Minneapolis local bus transit network (MLBTN) is considered as a case study for the application of the developed robustness assessment approach. Minneapolis is the largest (i.e., 430,000 capita) and most populous (i.e., 7.6% of the total population in Minnesota) city in the U.S. state of Minnesota (U.S Census Bureau, 2019). The transportation system in the city is characterized by a highly connected urban multimodal transit network that serves numerous attractions, destinations, and other critical facilities such as universities, hospitals, public schools, and recreational and shopping centers (Minneapolis City 2019). The system is operated by the city's Metro Transit Agency and incorporates light rail, commuter rail, and local bus transit networks. The

public transit network serves an average of 249,300 passengers per day, with the local bus transit networks only serve approximately 171,600 passengers per day (Metropolitan Council 2019). The city's comprehensive plan for 2040 aims at improving the reliability of the public transit in order to increase ridership rates as well as to support new housing, job schemes and other transit-oriented developments (Minneapolis 2040 — The City's Comprehensive Plan 2019).

3.1. DATA DESCRIPTION AND PREPROCESSING

The automated passenger count (APC) dataset of the MBLTN is provided by the *Metro Transit Agency under the Metropolitan council agreement*. The dataset is available on the Minnesota Geospatial Commons website under the Minnesota Government Data Practices Act (Minnesota Geospatial commons website 2019). It includes the average daily boarding and alighting at each station during regular working hours from September 2019 to December 2019. The provided APC dataset includes the longitude, latitude, X and Y-coordinates, stops site ID and stops' name, the average daily boarding and alighting at each station during regular working hours (non-standard days are excluded from the average). The open-source data has been filtered and non-representative days are excluded from the dataset, including days of extreme weather conditions, unusual events, and other days with abnormal travel patterns.

Within the dataset, the total boarding and alighting are unequal for different stations, routes and provider levels. This might be attributed to (Kimpel et al. 2002; Lan

2015): 1) APC device faults that may undercount passengers due to high volumes of passengers at blind spots when they are not detected ; 2) behavior of passengers (passenger carry over) that mostly occurs in routes that end in a loop or in a pair of interlined routes, where passengers may stay on board to join the next trip; 3) overcount of passengers due to the bus operator activities that mostly occur at layover points when the bus is empty (i.e., adjusting bus mirrors or using washrooms) ; and 4) errors in data collection, aggregation and filtering.

The above unbalanced trip error has been addressed in the current study through two stages. *First*, the variation between the total boarding (ons) and the total alighting (offs) along the routes has been calculated (i.e., between 0.8 and 1.1) and compared to the allowable thresholds (i.e., between 0.7 and 1.3) according to the TCRP synthesis 77 guidelines (Boyle 2009). *Second*, ons and offs are balanced such that negative loading errors of the cumulative passenger flow are avoided (i.e., correcting negative loads). Three methods are available in the literature and can be applied to balance such errors including factor into the higher count (Lan 2015), pseudo stop (TCRP113 2006), and factoring the average (Lu 2008). In the current study, the latter method was used as Dawei Lu (2008) highlighted that such three methods typically yield similar balanced trips. The factoring to the average method splits the imbalanced trip errors over the boarding and alighting counts such that they are over- or undercounted. A target total number of passengers is computed based on the average total number of boarding and alighting and an adjustment factor is subsequently obtained for each of the boarding and alighting. The boarding and alighting

counts are then multiplied by the corresponding adjustment factor, and total balanced boarding and alighting are obtained. The adjusted passenger flow rates are then checked herein for negative loading errors and are found to be free of such errors.

The transit time-of-day analysis is performed in the current study to calculate the peak hour volumes based on the average daily passenger flow, which are subsequently used as link weights. Since the peak hour volumes oblige the required size of the transit fleet, the transit time-of-day analysis approach is based on a factor of 0.14 [according to the TRB-National Highway research program (Martin et al. 1998)] to convert the total average daily passenger flow (i.e., available within the dataset) to peak hour volumes for all purposes-transit trips.

3.2. MODEL DEVELOPMENT

The topological configuration of a public transit network can be defined through three different mapping methods: a) **Space of changes** *(P-space)*, where nodes represent stations and links exist if there is a direct bus route between any two stations (Kurant et al. 2006); b) **Space of routes** *(R-space)*, where routes are defined by nodes and connection between lines are common stations between these routes (Xu et al. 2007; Zhang et al. 2019); and c) **Space of stations** *(L-space)*, where nodes represent stations, while links represent both the physical and the route connection between these stations (Kurant et al. 2006; Xu et al. 2007). The current study followed the latter method because it can capture the actual

variation in passenger demands over links as well as the operating frequency and capacity of bus transit routes (Zhang et al. 2018).

The Minneapolis local bus transit network (MLBTN) contains 27 routes, resulting in a network of 4,339 nodes and 4,664 links, as shown in Figure 3. The MLBTN is conceptualized as a weighted directed network, where each link is assigned a weight and direction based on the passenger flow and route direction, respectively. The direction of links is represented by either Eastbound, Westbound, Northbound or Southbound based on the bus schedule and stops' sequence, as shown in Figures 3a and 3b. The MLBTN is modelled as a directed network and represented by *G = (S, L)*, where *G* represents the Minneapolis local bus transit network, *S* represents the set of stations and *L* denotes links of *S* components. The adjacency matrix of the network is represented by:

$$a_{ij} = \begin{cases} 1, & (i,j) \in S \\ 0, & (i,j) \notin S \end{cases}, \qquad a_{ij} \neq a_{ij} \ and \ a_{ii} = 0,$$

$where\ a_{ij} = 1$ if there is a direct connection (i.e., a bus route) from i to j.

4. RESULTS

This section demonstrates the topological characteristics and cascade failure analysis results of the MLBTN. First, the topological characteristics of the network are examined based on various centrality measures. In addition, the CML model is used to obtain the dynamic robustness of the MLBTN network based on random node, link and route failure scenarios that represent different disruptions to bus transit components such as traffic

congestion and severe weather conditions (e.g., snowstorms, floods, etc.). Specifically, such disruptions may induce additional demands, thus impacting transit stations or roadway sections or transit corridors (U.S DOT Statistics 2018). The network robustness towards route failures is quantified considering various centrality and operational measures. A sensitivity analysis is conducted on the impact of *PV/RC* ratio variations on the failure thresholds and the corresponding absorptive capacities are obtained.

4.1. TOPOLOGICAL CHARACTERISTICS

The static analysis of the MLBTN's topology demonstrates that approximately 81% of the nodes have K_{out} = 1 and K_{in} = 1, as shown in Figures 4a and 4b, respectively. As can be seen also in the Figures, the MLBTN follows a power-law distribution ($P_{k_{out}}$ = 0.72 $k^{-3.21}$ and $P_{k_{in}}$ = 0.72 $k^{-3.22)}$ and subsequently can be classified as a scale-free network. The power-law exponent values (i.e., 3.21 and 3.22) show that the MLBTN is of significant heterogeneity. Such values indicate also that the MLBTN acts as a random network during random failures (Barabási 2015) and subsequently fragments at a finite critical threshold independent of the network size, as will be discussed next.

4.2. ROBUSTNESS

4.2.1. LINK AND NODE FAILURES

Figure 5 shows a significant difference between the overall robustness of the MBLTN under node and link failure scenarios. For example, at a fraction of removed nodes and links, *f*, of 0.3, the corresponding network robustness values are 0.38 and 0.61, respectively. However, both scenarios yield completely disconnected networks at similar critical threshold values of $f_{c\text{-}N(act)}$ = 0.58 and $f_{c\text{-}L(act)}$ = 0.63 when nodes and links are triggered, respectively, as shown in Figure 5. By comparing such values to the theoretical critical threshold value, $f_{c(th)}$, developed by Molloy et al. (1995) for scale-free networks and presented in Equation 12, there are variations of 20% and 15% for node and link failures, respectively. These variations are mainly because $f_{c(th)}$ is based on the network topology ($k_{\min}$ and γ) only; while $f_{c\text{-}N\ (act)}$ and $f_{c\text{-}L\ (act)}$ values depend on the developed CML model that considers both the topology and the passenger flow when systemic risks within bus transit networks are quantified.

$$f_{c_{(th)}} = 1 - \frac{1}{\frac{(\gamma-2)}{(\gamma-3)}k_{\min}-1}, \quad for \ \ \gamma > 3 \tag{12}$$

4.2.2. CAPACITY DISRUPTIONS

The passenger volume to route capacity (*PV/RC)* ratios can be utilized to represent any disruptions that may occur as a result of either variation in the frequency of buses or excess in demands with respect to the route capacities. In this respect, disruptions to the network

capacities have been investigated in the current study and the corresponding network robustness is quantified under various *PV/RC* ratios. Figures 6a and 6b show the sensitivity of f_c to different *PV/RC* ratios under node and link failure scenarios, respectively. The MLBTN exhibits an initial *PV/RC* value of 0.18 and corresponding critical threshold value of $f_{c\text{-}N(act)} = 0.58$ and $f_{c\text{-}L(act)} = 0.63$ for node and link failures, respectively. However, in the current study, the absorptive capacity is represented by the *PV/RC* ratio that the network can withstand with minor changes in the corresponding critical thresholds. Therefore, the absorptive capacity of the MLBTN is 0.35 (i.e., 200% increase) and 0.45 (i.e., 250% increase) under node and link failures, with corresponding minor variations in $f_{c\text{-}N(act)}$ and $f_{c\text{-}L(act)}$, respectively.

As can be seen also in Figures 6a and 6b, for all *PV/RC* ratios, the $f_{c\text{-}N(act)}$ and $f_{c\text{-}N(act)}$ are lower than $f_{c(th)}$. For instance, at *PV/RC* = 0.09 (i.e., 50% less than the original value), *the* $f_{c(act)}$ values are 0.59 and 0.64 for node and link failures, respectively, which are smaller than $f_{c(th)}$. This finding emphasises the importance of considering the *PV/RC* ratio as well as the topological characteristics of the network in obtaining the critical thresholds of local bus transit networks.

4.2.3. ROUTE FAILURES

As every route encompasses a set of nodes and links, the robustness of the MBLTN under route failures highly depends on the characteristics of such components. Therefore, for each route failure, 43 different variables are calculated to represent a wide range of operational

(e.g., N_L, N_N, $(PV/RC)_R$ (Aver)) and topological (e.g., BC_R^{aver}, D_R^{aver}, EC_R^{max}, CC_R^{aver}) measures, as presented in Table 1. A preliminary analysis is conducted to investigate the correlation between the route failure and such measures. As can be seen in Figure 7, although the *RFIR* has shown a correlation that is higher than 0.7 with the PR_R^{avg} and D_R^{avg}, the correlation matrix indicates the insignificant correlation between the robustness indices (i.e., *RFIR*, $N_{f\text{-}L}$ % and $N_{f\text{-}N}$ %) and the majority of the considered measures. These results highlight the interdependence, rather than the direct interrelation, between such measures, which renders the robustness of the MLBTN to be a function of the interplay between all topology and operational measures.

GA was therefore integrated with K-means clustering in order to categorize the MLBTN routes considering the interplay between their topology and operational measures. The GA was applied using a population size of 200, crossover ratio of 0.2, maximum generations of 100, and a tournament selection with size 2. The integrated GA-clustering analysis was applied 29 times, each of which includes: 1) predefining a *K* value (between 2 and 30); and 2) applying a binary GA to identify the optimum variable set that can represent the whole dataset through such number of clusters (i.e., *K*). An optimum variable set is thus identified for each number of clusters and is referred to as the GA solution. Figure 8 shows the frequency of including each of the 43 measures in the GA solution. A high frequency indicates that the corresponding measure was frequently used for clustering, which reflects its importance for representing the whole dataset. For example, the top five measures of the highest importance are the maximum weighted degree, S_R^{max}, the *RFIR*, the

average degree, D_R^{aver}, the average Eigen-centrality, EC_R^{aver}, and the maximum weighted outdegree, $S_R^{out\text{-}aver}$. These measures appeared 21, 18, 17, 14, and 14 times in the GA solutions, respectively. Measures of the highest importance include topology and centrality measures. This supports the limitation of characterizing bus transit routes based on operational measures only and the need for including topology measures to assess the overall bus transit network robustness under route failures.

The percentage of the variance explained by each GA solution and the number of variables included at each of the predefined *K* values are shown in Figure 9. As can be seen in the figure, the percentage of variance explained increases with the number of clusters (i.e., *K*), where the best cluster model is that with six variables and 27 clusters (describing approximately 100% of the variance). However, this is considered an infeasible model due to the large number of clusters generated. As such, a model with a smaller number of clusters can be more useful. The use of six clusters with eight variables can represent approximately 96% of the variance, which can be increased by only 3.7% using a model with more clusters. Accordingly, a model with six clusters and eight variables is considered herein as the near optimal cluster model. Figure 10 shows the statistical behavior of each variable over the six clusters. The variables included for clustering represent three different measures; robustness (*RFIR*), connectedness (BC_R^{aver}, CC_R^{aver}, EC_R^{aver}, D_R^{in-max}, Q_R^{max}), and passenger flow ($S_R^{out\text{-}max}$ and S_R^{max}). These variables along with their definitions, physical representations, and mathematical formulations are presented in Table 2, whereas a detailed description of each cluster is provided in Table 3. Based on the *RFIR* values, route impacts

can be classified into *low*, *moderate*, and *high*. Specifically, clusters 1 to 4 include routes with low *RFIR* values that span between 0.18 and 0.48. Cluster 5 represents routes with high *RFIR* values that are between 0.62 and 0.99, while cluster 6 contains moderate impact routes with *RFIR* values that vary between 0.34 and 0.63. The results show that BC_R^{aver}, EC_R^{aver} were found to have a significant impact on the severity of the failure compared to other connectedness indicators, highlighting the importance of the location, accessibility levels, and characteristics of the neighboring nodes in controlling the severity of failure propagation through the network. Moreover, routes with relatively high connectedness indicators and moderate to high passenger flow indicators can significantly impact the overall network vulnerability. In contrast, routes with relatively low connectedness indicators and high passenger flow indicators are characterized by low or moderate route failure impact ratio (i.e., their failures have low to moderate effects on the overall network vulnerability).

Figure 11 shows the spatial distribution of the six clusters over the MLBTN. Routes of relatively higher failure impacts (i.e., those with moderate to high *RFIR* values) are highly connected and located in the core of the network. Most of the stations along such routes exhibit high connectedness (expressed through higher BC_R^{avg}, CC_R^{avg} *and* EC_R^{avg}) compared to other routes, as presented in Table 3. On the other hand, routes of lower failure impacts (i.e., those with low *RFIR* values) are located near the network boundaries. However, a small fraction of the stations along these routes is highly connected to other routes in the network.

4.2.4. MANAGERIAL INSIGHTS

Enhancing the robustness of transit networks is key towards improving the overall reliability of these networks which is considered one of the main goals of transit agencies. Therefore, evaluating the network robustness during disruptions provides a deeper understanding of the required emergency/mitigation strategies and managing the available resources that aid at minimizing the impact and recovery time of disruptions.

The current study develops an approach that considers the nature of critical infrastructure networks with a demonstration application on public transit networks. The approach provided an in-depth understanding of the failure behavior of the network components (bus stops, accessibility between stations, and bus routes) during disruptions and the impact of these disruptions on the transit service.

Based on the results of this study, transit service providers and decision-makers can evaluate the transit network under various operational conditions (e.g., passenger volume-to-route capacity ratios) and obtain the corresponding robustness thresholds. In addition, the results enable assessing the ability of the spare capacity to absorb the excess demands during disruptions. Moreover, based on the identified critical routes and their characteristics, transit agencies can manage the transit fleet and ensure the availability of alternative routes as a response strategy during disruptive events that ensure the fulfillment of the service requirements during and following such events.

5. CONCLUSIONS AND FUTURE WORK

Urban local bus transit networks provide fundamental accessibility and mobility services for users and any disruptions to such complex and dynamic networks are associated with severe socio-economic impacts. Therefore, more efforts and studies should be directed towards quantifying and enhancing their resilience under cascade failure scenarios to ensure a trustworthy service for the users.

In the current study, a hybrid approach is developed to quantify the robustness of urban local bus transit networks, where complex network theory, machine learning and genetic algorithms are integrated within the transportation engineering field. The developed approach aims at utilizing several topology and operational measures to evaluate the robustness of urban local bus transit networks following the failure of key network components (i.e., stops, links, and routes). As such, the CML model is applied to assess the robustness of the MLBTN considering different *PV/RC* ratios. The MLBTN is simulated as a directed weighted network through an integrated CML-passenger flow redistribution model. In addition, GA is coupled with K-means clustering to identify the primary controllers of the network robustness as well as to categorize routes based on their topological and operational measures.

The results of the current study showed that the MLBTN exhibits a higher vulnerability and a lower absorptive capacity under random node failures compared to those under random link failures. In addition, the degree of accessibility, location, and characteristics of the failed stops along with their neighboring stops are key cascade failure

factors. The actual obtained critical thresholds of the MLBTN are also lower than their corresponding theoretical values, which indicates the importance of considering the redistribution of passenger flows when cascade failures are quantified. Moreover, both connectedness and passenger volumes on outwards links impact significantly the *Route failure impact ratio* of the routes. The results of the clustering analysis, in particular, show that both directions of the routes may not necessarily fall within the same cluster, which emphasizes the importance of simulating local bus transit networks as directed networks. Overall, the result of this study indicates the importance of including both the topology and operation-related measures for the characterization of the local bus transit networks. This can, in turn, facilitate the development of effective planning, management, and risk mitigation strategies.

The approach developed in the current study can be applied during the post-design stage to evaluate the network robustness considering its topology and operational measures. Specifically, the approach provides an attainable post design approach for evaluating the cascade failure-based reliability of urban local bus transit networks, which is essential to comprehensively understand the interplay between key network robustness factors, develop mitigation strategies and emergency schemes and enhance the network robustness. This can ultimately aid the decision-makers and transit service providers to assess the consequences of planning decisions and operational changes.

The results of the current study can further aid the decision-makers in: *i*) evaluating the potential planning- and operation-related decisions and their impacts on the network

robustness under normal and abnormal interventions; and *ii*) preparing effective management and risk mitigation strategies that minimize the impact and consequences of disruptive events. This will, in turn, contribute to improving the network resilience and cascade failure-based reliability, enhancing the attractiveness of the network, and increasing the ridership rates.

Future research studies should incorporate other modes of transportation (e.g., subway, light rail, and taxi) and simulate transit networks as a directed multimodal network. Furthermore, the behavioral response of passengers during disruptive events should be considered when the passenger flow redistribution models are developed.

6. REFERENCES

Abdelaty, Hatem, Moataz Mohamed, Mohamed Ezzeldin, and Wael El-Dakhakhni. 2020. "Quantifying and Classifying the Robustness of Bus Transit Networks." *Transportmetrica A: Transport Science* 16(3): 1176–1216. https://doi.org/10.1080/23249935.2020.1720042.

Albert, Réka, and Albert László Barabási. 2002. "Statistical Mechanics of Complex Networks." *Reviews of Modern Physics* 74(1): 47–97.

American Public Transportation Association. 2019. *PUBLIC TRANSPORTATION RIDERSHIP REPORT*.

Bak, P., Tang, C., & Wiesenfeld, K. 1987. "Self-Organized Criticality: An Explanation of the 1/f Noise. Physical Review Letters." 59(4): 381.

Barabási, A. L. 2016. . ". Network Science." *Cambridge university press.*

Barabási, Albert-László. 2015. "NETWORK SCIENCE. 5. The Barab{á}si-Albert Model." *Network Science*: 1–45. http://barabasi.com/networksciencebook/content/book_chapter_5.pdf.

Barrat, A., M. Barthélemy, R. Pastor-Satorras, and A. Vespignani. 2004. "The Architecture of Complex Weighted Networks." *Proceedings of the National Academy of Sciences of the United States of America* 101(11): 3747–52.

Beniwal, Sunita, and Jitender Arora. 2012. "Classification and Feature Selection Techniques in Data Mining." *International Journal of Engineering* 1(6).

Bonacich, Phillip. 2007. "Some Unique Properties of Eigenvector Centrality."

Social Networks 29(4): 555–64.

Boyle, Daniel K. 2009. Passenger Counting Systems *Passenger Counting Systems*.

Bruneau, Michel et al. 2003. "A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities." *Earthquake Spectra* 19(4): 733–52.

Cats, Oded, and Erik Jenelius. 2015. "Planning for the Unexpected: The Value of Reserve Capacity for Public Transport Network Robustness." *Transportation Research Part A: Policy and Practice* 81: 47–61. http://dx.doi.org/10.1016/j.tra.2015.02.013.

Ceder, A. 2016. *Public Transit Planning and Operation: Modeling, Practice and Behavior*. CRC press.

Celebi, M. Emre, Hassan A. Kingravi, and Patricio A. Vela. 2013. "A Comparative Study of Efficient Initialization Methods for the K-Means Clustering Algorithm." *Expert Systems with Applications* 40(1): 200–210. http://dx.doi.org/10.1016/j.eswa.2012.07.021.

Chazottes, J. R., & Fernandez, B. (Eds.). 2005. "Dynamics of Coupled Map Lattices and of Related Spatially Extended Systems." *Springer Science & Business Media* 671.

ctvnews. 2020. "Subway Service from Jane to Ossington Resumes after Partial Train Derailment." https://toronto.ctvnews.ca/subway-service-from-jane-to-ossington-resumes-after-partial-train-derailment-1.4778157.

Cupac, Valentina, Joseph T. Lizier, and Mikhail Prokopenko. 2013. "Comparing Dynamics of Cascading Failures between Network-Centric and Power Flow Models." *International Journal of Electrical Power and Energy Systems* 49(1): 369–79. http://dx.doi.org/10.1016/j.ijepes.2013.01.017.

Dawei Lu, M.S. 2008. "ROUTE LEVEL BUS TRANSIT PASSENGER ORIGIN-DESTINATION FLOW ESTIMATION USING APC DATA: NUMERICAL AND EMPIRICAL INVESTIGATIONS."

Derrible, Sybil, and Christopher Kennedy. 2010. "The Complexity and Robustness of Metro Networks." *Physica A: Statistical Mechanics and its Applications* 389(17): 3678–91. http://dx.doi.org/10.1016/j.physa.2010.04.008.

Dickens, Matthew. 2011. "PUBLIC TRANSPORTATION RIDERSHIP REPORT Fourth Quarter 2011 HEAVY RAIL PUBLIC TRANSPORTATION RIDERSHIP REPORT Fourth Quarter 2011." (c).

Dickens, Matthew, John Neff, and Darnell Grisby. 2012. "1/28 Reading: APTA 2012 Public Transportation Fact Book." http://trid.trb.org/view.aspx?id=1225089.

Dugué, Nicolas, and Anthony Perez. 2015. "Directed Louvain : Maximizing Modularity in Directed Networks." (November): 0–14. https://hal.archives-ouvertes.fr/hal-01231784.

Ezzeldin, Mohamed, and Wael E. El-Dakhakhni. 2019. "Robustness of Ontario

Power Network under Systemic Risks." *Sustainable and Resilient Infrastructure* 00(00): 1–20. https://doi.org/10.1080/23789689.2019.1666340.

Fox9. 2020. "Metro Transit Suspending All Bus, Light Rail Service through the Weekend Due to Protests." https://www.fox9.com/news/metro-transit-suspending-all-bus-light-rail-service-through-the-weekend-due-to-protests.

García-Morales, Vladimir. 2016. "From Deterministic Cellular Automata to Coupled Map Lattices." *Journal of Physics A: Mathematical and Theoretical* 49(29): 0–17.

Haggag, May, Ezzeldin, Mohamed, El-Dakhakhni, Wael, and Hassini, Elkafi (2020). "Resilient cities critical infrastructure interdependence: a meta-research." *Sustainable and Resilient Infrastructure*, 1-22.

Hosseini, Seyedmohsen, and Kash Barker. 2016. "Modeling Infrastructure Resilience Using Bayesian Networks: A Case Study of Inland Waterway Ports." *Computers and Industrial Engineering* 93: 252–66. http://dx.doi.org/10.1016/j.cie.2016.01.007.

Hosseini, Seyedmohsen, Kash Barker, and Jose E. Ramirez-Marquez. 2016. "A Review of Definitions and Measures of System Resilience." *Reliability Engineering and System Safety* 145: 47–61.

Huang, Ailing et al. 2015. "Cascading Failures in Weighted Complex Networks of Transit Systems Based on Coupled Map Lattices." *Mathematical Problems in*

Engineering 2015.

Huang, Jiajun, Feng Zhou, and Mengru Xi. 2018. "Calculation Method for Load Capacity of Urban Rail Transit Station Considering Cascading Failure." *Journal of Advanced Transportation* 2018.

Humphrey, Nancy. 2008. TR News *The Role of Transit in Emergency Evacuation*.

Jayasinghe, Amila, and Talat Munshi. 2014. "'Centrality Measures' as a Tool to Identify the Transit Demand at Public Transit Stops; A Case of Ahmedabad City, India." *International Journal of Advanced Research* 2(7): 1063–74.

Jiang, Bin, and Christophe Claramunt. 2004. "Topological Analysis of Urban Street Networks." *Environment and Planning B: Planning and Design* 31(1): 151–62.

Jones, Peter R., and Philip W. Harris. 1999. "Developing an Empirically Based Typology of Delinquent Youths." *Journal of Quantitative Criminology* 15(3): 251–76.

Kaneko, Kunihiko. 1992. "Overview of Coupled Map Lattices." *Chaos* 2(3): 279–82.

Kaufman, Sarah, Carson Qing, Nolan Levenson, and Melinda Hanson. 2012. "Transportation During and After Hurricane Sandy." *Rudin Center for Transportation NYU Wagner Graduate School of Public Service* (November): 1–36.

Kim, Hyun, Changjoo Kim, and Yongwan Chun. 2016. "Network Reliability and

Resilience of Rapid Transit Systems." *Professional Geographer* 68(1): 53–65.

Kimpel, Thomas J et al. 2002. "Implications for National Transit Database Reporting." *Transportation Research Record: Journal of the Transportation Research Board* 1835(03): 93–100.

Kurant, Maciej, and Patrick Thiran. 2006. "Extraction and Analysis of Traffic and Topologies of Transportation Networks." *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 74(3).

Lan, Cheng. 2015. "Route-Level Transit Passenger Origin-Destination Trip Estimation from Automatic Passenger Counting Data: A Case Study in Edmonton." https://era.library.ualberta.ca/files/xd07gw76v/Lan_Cheng_201501_MSc.pdf.

Leicht, E. A., and M. E.J. Newman. 2008. "Community Structure in Directed Networks." *Physical Review Letters* 100(11): 1–4.

Levinson, D., Liu, H. X., & Bell, M. (Eds.). 2011. "Network Reliability in Practice: Selected Papers from the Fourth International Symposium on Transportation Network Reliability." *Springer Science & Business Media*.

Litman, Todd. 2014. "Evaluating Public Transit Benefits and Costs - Best Practices Guidebook." *Victotia Transport Policy Institute*: 1–138. www.vtpi.orgInfo@vtpi.org.

Liu, Jing, Huapu Lu, He Ma, and Wenzhi Liu. 2017. "Network Vulnerability

Analysis of Rail Transit Planin Beijng-Tianjin-Hebei Region Considering Connectivity Reliability." *Sustainability (Switzerland)* 9(8).

Marchiori, Massimo, and Vito Latora. 2000. "Harmony in the Small-World." *Physica A: Statistical Mechanics and its Applications* 285(3): 539–46.

Martin, William A, and Nancy A McGuckin. 1998. NCHRP Report *NCHRP Synthesis 365: Travel Estimation Techniques for Urban Planning*. http://isbnplus.org/9780309053655.

Matherly, Deborah, Jon A. Carnegie, and Jane Mobley. 2017. Improving the Resilience of Transit Systems Threatened by Natural Disasters, Volume 1: A Guide *Improving the Resilience of Transit Systems Threatened by Natural Disasters, Volume 1: A Guide*.

Metropolitan Council. 2009. *Twin Cities Transit System Performance Evaluation*.

———. 2019. *TRANSIT SYSTEM PERFORMANCE EVALUATION-Twin Cities Metropolitan Region*.

Minneapolis 2040 — The City's Comprehensive Plan. 2019. Minneapolis City Council.

Minneapolis City, Council. 2019. *Minneapolis 2040 — The City's Comprehensive Plan*.

Minnesota Geospatial commons website. 2019. "Minnesota Geospatial Commons Website." https://gisdata.mn.gov/.

Mishra, Sabyasachee, Timothy F. Welch, and Manoj K. Jha. 2012. "Performance Indicators for Public Transit Connectivity in Multi-Modal Transportation Networks." *Transportation Research Part A: Policy and Practice* 46(7): 1066–85.

Motter, Adilson E., and Ying Cheng Lai. 2002. "Cascade-Based Attacks on Complex Networks." *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 66(6): 4.

Newman, M. E.J. 2001. "Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks, and Centrality." *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 64(1): 7.

Norris, Fran H. et al. 2008. "Community Resilience as a Metaphor, Theory, Set of Capacities, and Strategy for Disaster Readiness." *American Journal of Community Psychology* 41(1–2): 127–50.

Opsahl, Tore, Filip Agneessens, and John Skvoretz. 2010. "Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths." *Social Networks* 32(3): 245–51. http://dx.doi.org/10.1016/j.socnet.2010.03.006.

Pandit, Shraddha, and Suchita Gupta. 2011. "A Comparative Study on Distance Measuring Approaches for Clustering." *International Journal of Research in Computer Science* 2(1): 29–31.

Pearson, K. 1896. "VII. Mathematical Contributions to the Theory of Evolution.—

III. Regression, Heredity, and Panmixia. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character." 187: 253–318.

Progressive railroading. 2016. "Hurricane Sandy: Four Years Later, New York City Transit Is Still Fixing, Fortifying the Rail System." https://www.progressiverailroading.com/.

Reggiani, Aura. 2013. "Network Resilience for Transport Security: Some Methodological Considerations." *Transport Policy* 28: 63–68. http://dx.doi.org/10.1016/j.tranpol.2012.09.007.

Ren, Tao, Yi Fan Wang, Miao Miao Liu, and Yan Jie Xu. 2016. "Analysis of Robustness of Urban Bus Network." *Chinese Physics B* 25(2): 0–12.

Rodrigues, Francisco Aparecido. 2019. "Network Centrality: An Introduction." : 177–96.

Rubulotta, Elena, Matteo Ignaccolo, Giuseppe Inturri, and Yodan Rofé. 2013. "Accessibility and Centrality for Sustainable Mobility: Regional Planning Case Study." *Journal of Urban Planning and Development* 139(2): 115–32.

Saeys, Yvan, Iñaki Inza, and Pedro Larrañaga. 2007. "A Review of Feature Selection Techniques in Bioinformatics." *Bioinformatics* 23(19): 2507–17.

Salama, Mohamed, Ezzeldin, Mohamed, El-Dakhakhni, Wael, and Tait, Michael. (2020). "Temporal networks: a review and opportunities for infrastructure

simulation." *Sustainable and Resilient Infrastructure*, 1-16.

Schwartz, H. G., M. Meyer, C. J. Burbank, M. Kuby, C. Oster, J. Posey, E. J. Russo, and A. Rypinski. 2014. *Ch. 5: Transportation. Climate Change Impacts in the United States: The Third National Climate Assessment, J. M. Melillo, Terese (T.C.) Richmond, and G. W. Yohe, Eds., U.S. Global*.

Shen, Yi, Gang Ren, and Bin Ran. 2019. "Cascading Failure Analysis and Robustness Optimization of Metro Networks Based on Coupled Map Lattices: A Case Study of Nanjing, China." *Transportation* (0123456789). https://doi.org/10.1007/s11116-019-10066-y.

Soh, Harold et al. 2010. "Weighted Complex Network Analysis of Travel Routes on the Singapore Public Transportation System." *Physica A: Statistical Mechanics and its Applications* 389(24): 5852–63. http://dx.doi.org/10.1016/j.physa.2010.08.015.

States Census Bureau, US. 2019. "United States Census Bureau." https://www.census.gov/quickfacts/fact/table/MN,minneapoliscityminnesota/PST045219.

Sullivan, J. L., D. C. Novak, L. Aultman-Hall, and D. M. Scott. 2010. "Identifying Critical Road Segments and Measuring System-Wide Robustness in Transportation Networks with Isolating Links: A Link-Based Capacity-Reduction Approach." *Transportation Research Part A: Policy and Practice*

44(5): 323–36. http://dx.doi.org/10.1016/j.tra.2010.02.003.

Sun, Lishan, Yuchen Huang, Yanyan Chen, and Liya Yao. 2018. "Vulnerability Assessment of Urban Rail Transit Based on Multi-Static Weighted Method in Beijing, China." *Transportation Research Part A: Policy and Practice* 108(October 2017): 12–24. https://doi.org/10.1016/j.tra.2017.12.008.

Tamvakis, Pavlos, and Yiannis Xenidis. 2012. "Resilience in Transportation Systems." *Procedia - Social and Behavioral Sciences* 48: 3441–50.

TCRP113, REPORT. 2006. *Using Archived AVL-APC Data to Improve Transit Performance and Management*.

Turalska, Malgorzata et al. 2019. "Cascading Failures in Scale-Free Interdependent Networks." *Physical Review E* 99(3): 1–9.

Turnquist, Mark, and Eric Vugrin. 2013. "Design for Resilience in Infrastructure Distribution Networks." *Environmentalist* 33(1): 104–20.

U.S DOT Statistics. 2018. *Transportation Statistics Annual Report*.

Wang, Jiaoe, Huihui Mo, Fahui Wang, and Fengjun Jin. 2011. "Exploring the Network Structure and Nodal Centrality of China's Air Transport Network: A Complex Network Approach." *Journal of Transport Geography* 19(4): 712–21. http://dx.doi.org/10.1016/j.jtrangeo.2010.08.012.

Watts, Duncan J. 2011. "A Simple Model of Global Cascades on Random Networks." *The Structure and Dynamics of Networks* 9781400841(9): 497–

502.

Weilant, Sarah, Aaron Strong, and Benjamin Miller. 2019. “Incorporating Resilience into Transportation Planning and Assessment.” *Incorporating Resilience into Transportation Planning and Assessment*.

White, Douglas R., and Stephen P. Borgatti. 1994. “Betweenness Centrality Measures for Directed Graphs.” *Social Networks* 16(4): 335–46.

de Winter, Joost C.F., Samuel D. Gosling, and Jeff Potter. 2016. “Comparing the Pearson and Spearman Correlation Coefficients across Distributions and Sample Sizes: A Tutorial Using Simulations and Empirical Data.” *Psychological Methods* 21(3): 273–90.

Xu, Xinping, Junhui Hu, Feng Liu, and Lianshou Liu. 2007. “Scaling and Correlations in Three Bus-Transport Networks of China.” *Physica A: Statistical Mechanics and its Applications* 374(1): 441–48.

Yassien, Yassien, Ezzeldin, Mohamed, Mohamed, Moataz, and El-Dakhakhni, Wael. (2020). "Air Transportation Infrastructure Robustness Assessment for Proactive Systemic Risk Management." Journal of Management in Engineering, 36(4), 04020037.

Zhang, Lin, Jian Lu, Bai bai Fu, and Shu bin Li. 2019. “A Cascading Failures Model of Weighted Bus Transit Route Network under Route Failure Perspective Considering Link Prediction Effect.” *Physica A: Statistical Mechanics and its*

Applications 523: 1315–30. https://doi.org/10.1016/j.physa.2019.04.122.

Zhang, Lin, Jian Lu, Bai Bai Fu, and Shu Bin Li. 2018. "A Review and Prospect for the Complexity and Resilience of Urban Public Transit Network Based on Complex Network Theory." *Complexity* 2018.

Zou, Zhiyun, Yao Xiao, and Jianzhi Gao. 2013. "Robustness Analysis of Urban Transit Network Based on Complex Networks Theory." *Kybernetes* 42(3): 383–99.

Table 1: Variables included in the feature selection stage

Variable	Acronym
Maximum weighted degree	S_R^{max}
Route failure impact ratio	$RFIR$
Average degree	D_R^{avg}
Average eigen centrality	EC_R^{avg}
Maximum weighted outdegree	$S_R^{out\text{-}max}$
Average harmonic closeness centrality	CCH_R^{avg}
Maximum modularity class	Q_R^{max}
Maximum eigen centrality	EC_R^{max}
Average weighted outdegree	$S_R^{out\text{-}avg}$
Average strong component number	SC_R^{avg}
Maximum indegree	$D_R^{in\text{-}max}$
Average indegree centrality	$D_R^{in\text{-}avg}$
Maximum weighted indegree	$S_R^{in\text{-}max}$
Average closeness centrality	CC_R^{avg}
Maximum harmonic closeness centrality	CCH_R^{max}
Average weighted degree	S_R^{avg}
Maximum betweenness centrality	BC_R^{max}
Maximum degree	D_R^{max}
Number of links along the route	N_L
Average betweenness centrality	BC_R^{avg}
Percentage of nodes with outdegree =1	$N(D_{out}=1)\,\%$
Average outdegree centrality	$D_R^{out\text{-}avg}$
Average modularity class	Q_R^{aver}
Maximum closeness centrality	CC_R^{max}
Number of stations along the route	N_N
Average authority score	$A_R(v)^{avg}$
Maximum page ranks	PR^{max}
Percentage of failed nodes based on link failure	$N_{f\text{-}L}\,\%$
Percentage of nodes with indegree = 1	$N(D_{in}=1)\,\%$
Average clustering	C_R^{avg}
Maximum clustering	C_R^{max}
Average *PV/RC*	$(PV/RC)_R^{avg}$
Average weighted indegree	$S_R^{in\text{-}avg}$
Average page ranks	PR_R^{avg}
Maximum outdegree	$D_R^{out\text{-}max}$
Maximum eccentricity	$E_R(v)^{max}$
Maximum authority	$A_R(v)^{max}$
Maximum hub	$H_R(v)^{max}$
Maximum strong component number	SC_R^{max}
Percentage of failed nodes based on node failure	$N_{f\text{-}N}\,\%$
Maximum PV/RC	$(PV/RC)_R^{max}$
Average eccentricity	$E_R(v)^{avg}$
Average hub	$H_R(v)^{avg}$

Table 2: Measures and definitions

Variables included in the clusters	Definition and Representation	Formula
Vulnerability or robustness indicator:		
Route failure impact ratio (*RFIR*)	• The ratio between the total number of failed nodes (N_{FR}) when a specific route failed and the number of nodes along that route (N_{TR}). • *RFIR* is used to identify the critical routes that can facilitate cascade failures in the network.	$RFIR = N_{FR} / N_{TR}$
Connectedness and accessibility indicators (represent the location and importance of a station):		
Average betweenness centrality (BC_R)	• The average betweenness centrality (BC) of the nodes along the route *R*, where BC is the fraction of shortest paths passing through each node *k* along that route (Barabási 2015; Newman 2001). • BC_R is used to identify the route containing the relatively important stops in terms of location compared to other stops included in the network (Jayasinghe et al. 2014).	$BC_R = \frac{1}{N_{TR}} \sum_k \sum_{l \neq k} \sum_{m \neq l \neq m} \frac{\rho(l,k,m)}{\rho(l,m)}$, k
	The average closeness centrality (CC) of the nodes along the route	

AVER-closeness-centrality (CC_R)	*R* (Rubulotta et al. 2013), where CC represents the average distance of each node *k* on that route to all other nodes in the network (Barabási 2015; Opsahl et al. 2010). • CC_R represents the degree of accessibility to the stations belonging to the route R compared to other stations in the network (Jayasinghe et al. 2014).	$$CC_R = \frac{1}{N_{TR}} \sum_k \frac{N-1}{\sum_j d_{kj}}$$ where, d_{kj} is the length of the shortest path between nodes *k* and *j*.
AVER-eigen centrality (EC_R)	• The average eigen centrality (EC) of the nodes along the route *R*, where EC measures the influence of each node *k* along that route based on the importance of the neighboring nodes (Bonacich 2007; Rubulotta et al. 2013). • EC_R represents the importance of the stations of a certain route based on and the importance of their neighboring stations (Mishra et al. 2012; Rodrigues 2019).	$$EC_R = \frac{1}{N_{TR}} \sum_k \frac{1}{\lambda} \sum_j A_{kj} D_{EC_j}$$ where D_{EC} is the right leading eigenvector, A_{kj} is the adjacency matrix, and λ is the eigenvalue of A_{kj} for which D_{EC} exists.
MAX-indegree (D_R^{in})	• The maximum number of links directed towards any node *k* over the route R (Barabási 2015). • It corresponds to the stops with the largest size (major stop like terminals, or main stations) that exit along a route based on the number of routes directed towards the station.	$$D_R^{in} = \max_k \sum_j a_{jk}$$ where a_{jk} is the adjacency matrix entry corresponding to the connection between nodes *j* and *k*.

MAX-modularity class (Q_R)	The maximum modularity class (Q) amongst the nodes located along the route R, where Q is used to measure the strength of divisions of a network into modules (Dugué et al. 2015; Leicht et al. 2008) . • It represents the likelihood of being within the same community, where higher modularity denotes that a group of edges are most likely within the same community.	$Q = \frac{1}{m}\sum_{i,j}\left[A_{ij} - \frac{D_i^{in} D_j^{out}}{m}\right]\delta(c_i, c_j)$ where A_{ij} is the adjacency matrix, m is the number of edges in the network, D_i^{in} is the indegree of the node i, $D_j^{out_t}$ is the outdegree of the node j, c_i (c_j) is the community to which the node i (j) belongs, and δ_{ij} is the Kronecker delta function (Leicht et al. 2008).
Passenger flow indicators:		
MAX-weighted outdegree (S_R^{out})	• The maximum outward node strength along the route R. • It represents the maximum passenger volume (*PV*) originated from a specific route (Barabási 2015).	$S_R^{out} = \max_k \sum_j w_{kj}$ where k is the node index along the route R and w_{kj} is the entry of the weighted adjacency matrix corresponding to the connection between nodes k and j.
MAX-Weighted Degree (S_R)	• The maximum node strength along the route R (Barabási 2015). • It corresponds to the station with the highest passenger circulation along a specific route (Mishra et al. 2012).	$S_R = \max_k \left(\sum_j w_{kj} + \sum_j w_{jk}\right)$

Table 3: Clustering analysis results

Cluster #	Routes Included	Features of each cluster	Comments
Clusters of low route failure impact ratio (*RFIR*):			
1	5N, 5S, 9E 14N,14S, 19N, 19S 22N, 22S	Robustness indicator: • Relatively low *RFIR* values that range between 0.23 and 0.3, indicating that the failure of such routes does not significantly impact the overall network performance Connectedness and accessibility indicators: • Relatively low BC_R values (between 0.21 and 0.41), indicating that the stations (nodes) along such routes are of a secondary importance in terms of the number of shortest paths passing through them. • Low CC_R values (between 0.13 and 0.17), indicating that the stations along these routes are less accessible and far from other stations across the network. • Relatively low EC_R values of (between 0.12 and 0.39), indicating that the bus stops along such routes are of a secondary importance as they are connected to a relatively less important stop. • Low D_R^{in} values (between 0.35 and 0.45), indicating the presence of more local bus stops (with a limited number of inward connections), compared to major stations or terminals, along such routes. • Low values of Max-modularity class (between 0-0.01) indicates a low strength of divisions of the networks into modules.	**The overall network vulnerability is not significantly impacted by the failure of these routes**

		Passenger flow indicators: • Relatively high S_R^{out} values (between 0.6 and 0.78), indicating the high passenger volume originated from the stops along these routes. • Relatively high S_R values (between 0.68 and 0.72), indicating the high passenger circulation in the stops along the included routes.	
2	6N,6S 4N, 7N, 10N, 11N 12E, 17E, 18N 25N, 61E,141N	Robustness indicator: • Relatively low *RFIR* values that range between 0.18 and 0.32, indicating that the failure of such routes does not significantly impact the overall network performance. Connectedness and accessibility indicators: • Moderate values of BC_R (between 0.38 and 0.7), indicating that the stations along such routes are of relatively high importance in terms of the number of shortest paths passing through them. • Low CC_R values (between 0.8 and 0.14), indicating that the stations along these routes are less accessible and far from other stations across the network. • Relatively low EC_R values of relative (between 0.22 and 0.4, indicating that the bus stops along the included routes are of secondary importance as they are connected to a relatively less important stop. • Low D_R^{in} values (between 0.22 and 0.44), indicating the low number of major stations or terminals along the included routes. • Low values of Max-modularity class (between 0.02 and 0.38) indicates a low strength of divisions of the networks into modules.	**The overall network vulnerability is not significantly impacted by the failure of these routes**

		Passenger flow indicators: • Very high S_R^{out} values (0.9), indicating the very high passenger volume originated from the stops along such routes. • Very high S_R values (1.0), indicating the significant passenger circulation at the stops along the included routes.	
3	4S, 7S, 10S, 11S, 12W, 17W 18S, 25S, 59S 61W, 141S	Robustness indicator: • Relatively low *RFIR* values that range between 0.18 and 0.38, indicating that the failure of such routes does not significantly impact the overall network performance. Connectedness and accessibility indicators: • Moderate BC_R values (between 0.3 and 0.5), indicating that the stops (nodes) along such routes are of secondary importance in terms of the number of shortest paths passing through it. • Low CC_R values (between 0.08 and 0.14), indicating that stations along these routes are less accessible and far from other stops across the network. • Low values of relative EC_R (between 0.3 and 0.5), indicating that the stations along such routes are connected to a relatively less important neighboring station. • Very high D_R^{in} values (1.0), indicating the presence of major stations or terminals along these routes. • Very high values of Max-modularity class (equals to 1) indicates a very high strength between divisions of the networks into modules.	**The overall network vulnerability is not significantly impacted by the failure of these routes**

		Passenger flow indicators: • Very high S_R^{out} values (1.0), indicating the very high passenger volume originated from the stations along the routes in this cluster. • Very high S_R values (0.9), indicating the high passenger circulation at the stations along these routes.	
4	2E, 2W, 3E 3W, 9W, 21E,21W, 67E, 67W 74E,74W, 824N 825N,825S, A-LINE-N A-LINE-S	Robustness indicator: • Low *RFIR* values that range (between 0.24 and 0.48), indicating that the failure of such routes does not significantly impact the overall network performance. Connectedness and accessibility indicators: • Moderate BC_R values (between 0.2 and 0.57), indicating that stops (nodes) along such routes are of secondary importance in terms of the number of shortest paths passing through it. • Relatively low CC_R values (between 0.08 and 0.24), indicating that the stations along these routes are less accessible and far from other stations across the network. • Relatively low EC_R values (between 0.2 and 0.44), indicating that the stations along such routes are connected to a relatively less important neighboring station. • Relatively low D_R^{in} values (between 0.2 and 0.42), indicating the low number of major stations or terminals along the routes included in this cluster. • Very low values of Max-modularity class (equals to 0) indicates a very high strength between divisions of the networks into modules. Passenger flow indicators: • Relatively low S_R^{out} values (between 0.12 and 0.38), indicating the very high	**The overall network vulnerability is not significantly impacted by the failure of these routes**

		passenger volume originated from the stations along the routes in this cluster. • Relatively low S_R values (between 0.18 and 0.42), indicating the low passenger circulation at the stations along these routes.	
Cluster of high route failure impact ratio (*RFIR*):			
5	53W, C-LINE-N, C-LINE-S	Robustness indicator: • Relatively high *RFIR* values (between 0.62 and 0.99), indicating that the failure of such routes can significantly impact the overall network performance. Connectedness and accessibility indicators: • High BC_R values (between 0.75 and 0.95), indicating that the stops (nodes) along such routes are of relatively high importance in terms of the number of shortest paths passing through it. • Low CC_R values (between 0.14 and 0.16), indicating that stations along these routes are less accessible and far from other stations across the network. • High EC_R values (between 0.62 and 1), indicating that the stations along such routes are connected to a relatively less important neighboring station. • Moderate D_R^{in} values (between 0.35 and 0.48), indicating the low number of major stations or terminals along the routes included in this cluster. • Very low values of Max-modularity class (equals to 0) indicates a very high strength between divisions of the networks into modules. Passenger flow indicators: • Moderate to high values of S_R^{out} values (between 0.32 and 0.7), indicating the relatively high passenger volume	**The overall network vulnerability is significantly impacted by the failure of these routes**

		originated from the stations along the routes included in this cluster. • Moderate to high S_R values (between 0.36 and 0.68), indicating the relatively high passenger circulation at the stations along these routes.	
Cluster of moderate route failure impact ratio (*RFIR*):			
6	53E, 59N, 824S	Robustness indicator: • Moderate *RFIR* values that range between 0.34 and 0.63, indicating that the failure of such routes can impact the overall network performance. Connectedness and accessibility indicators: • Low - Moderate BC_R values (between 0.3 and 0.63), indicating that the stops (nodes) along such routes are of relatively high importance in terms of the number of shortest paths passing through it. • High CC_R values (between 0.68 and 1.0), indicating that the stations along these routes are of a high accessibility level and are closer to other stations in the network. • Moderate EC_R values (between 0.35 and 0.63), indicating that the stations along such routes are connected to some important stations. • Relatively Low D_R^{in} values (between 0.35 and 0.48), indicating the low number of major stations or terminals along the routes included in this cluster • Very low values of Max-modularity class (ranges between 0 and 0.06) indicate a very low strength between divisions of the networks into modules.	**The overall network vulnerability is moderately impacted by the failure of these routes**

		Passenger flow indicators: • Moderate to high values of S_R^{out} values (between 0.32 and 0.9), indicating the relatively high passenger volume originated from the stations along the routes included in this cluster. • Moderate to high S_R values (between 0.32 and 1.0), indicating the relatively high passenger circulation at the stations along these routes.	

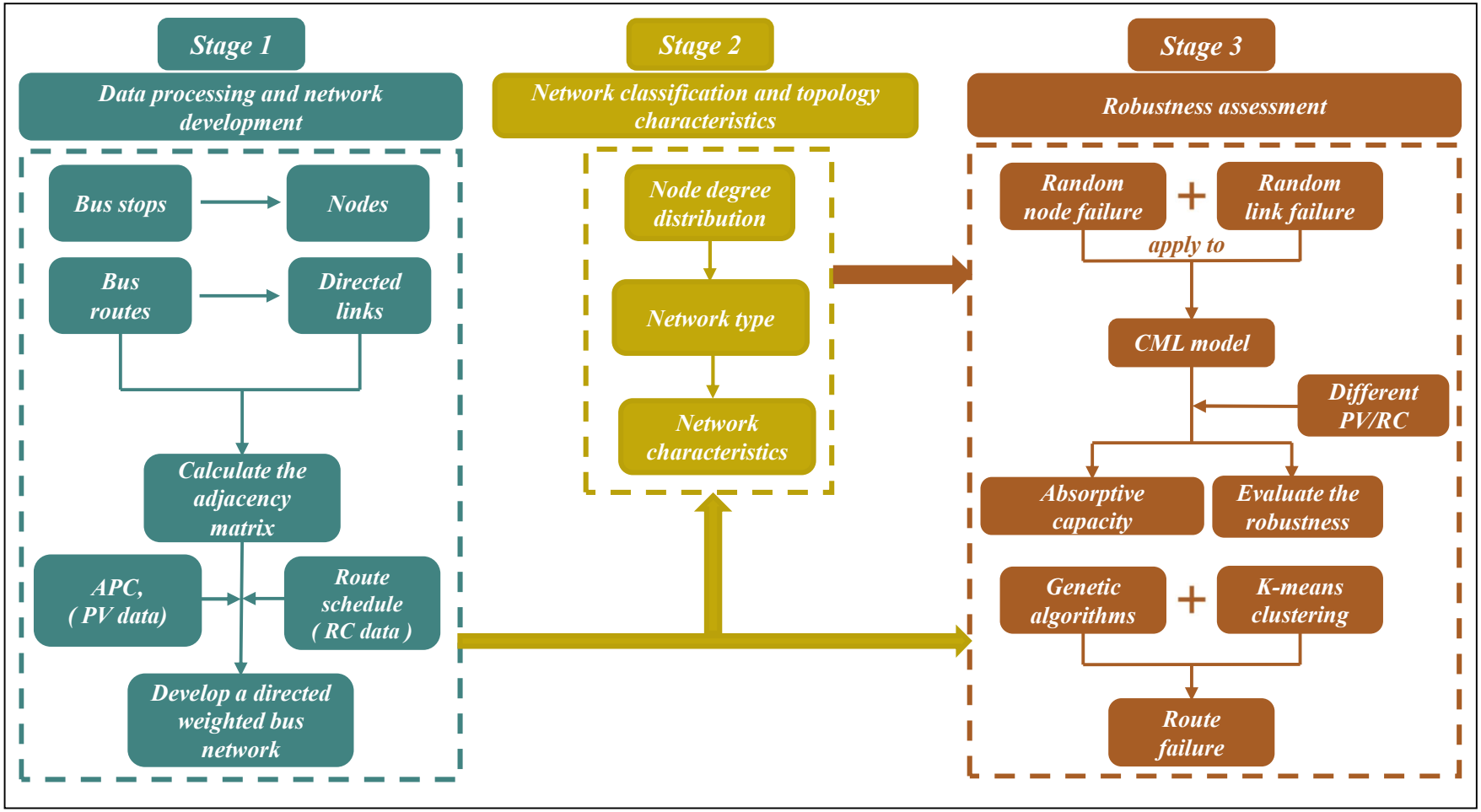

Figure 1: Methodology for robustness assessment of Minneapolis Local Bus Transit Network

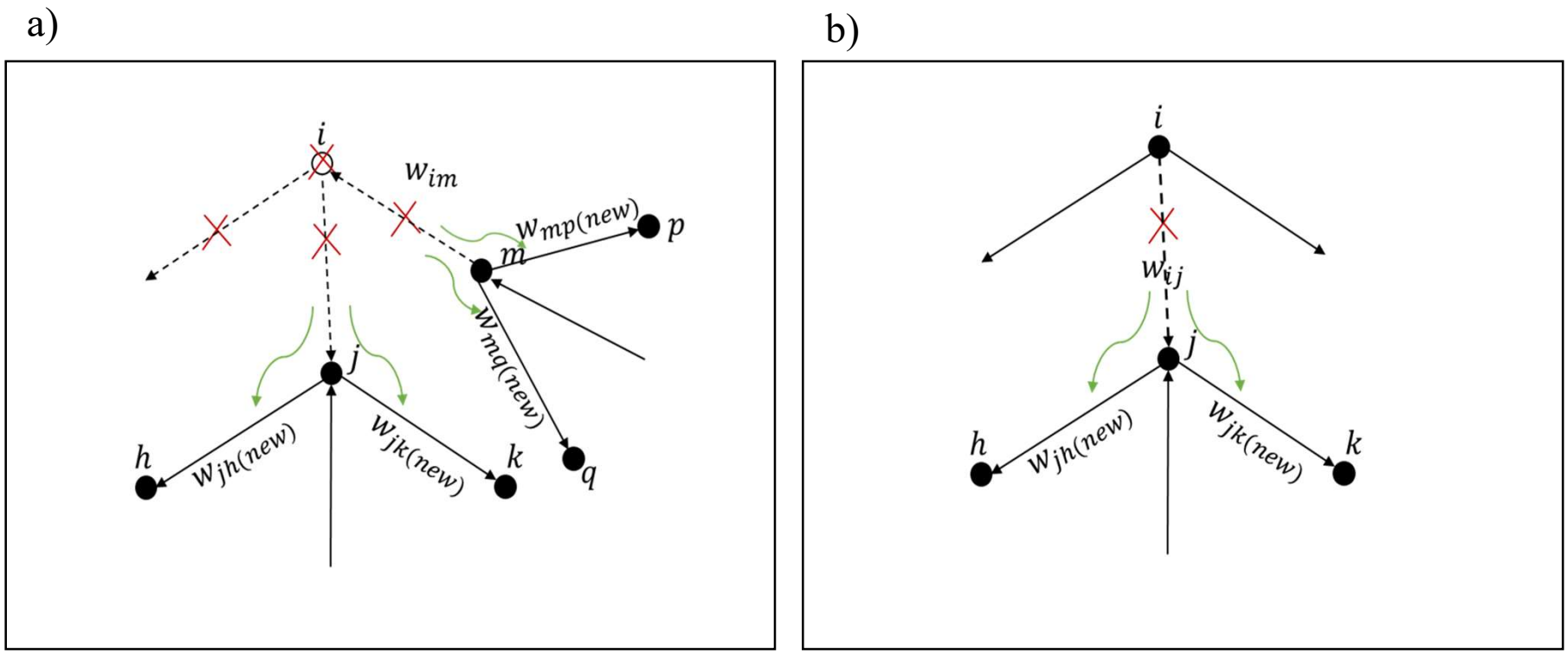

Figure 2: Passenger flow redistribution based on: a) Node failure; b) Link failure

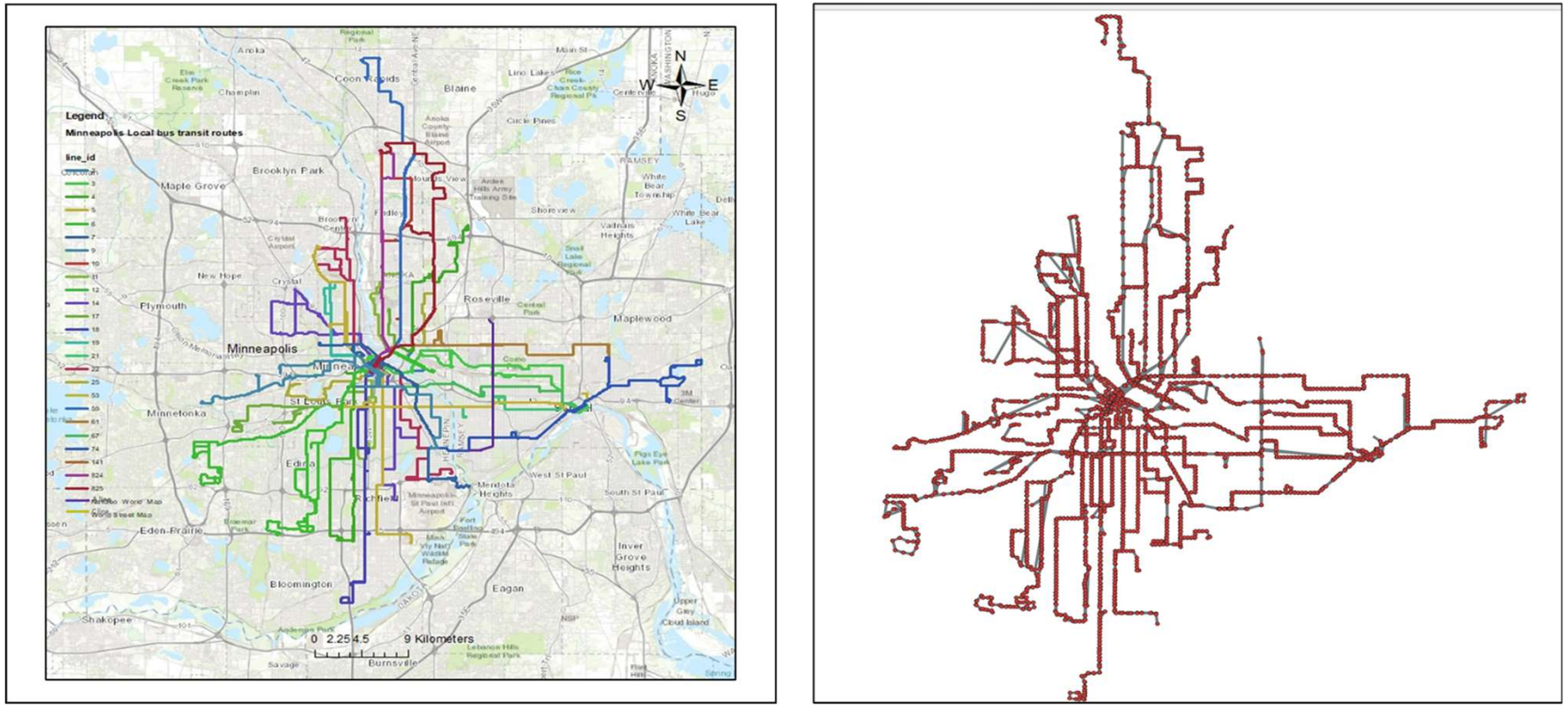

Figure 3: Minneapolis local bus transit network a) GIS Map; b) Directed Network using CNT

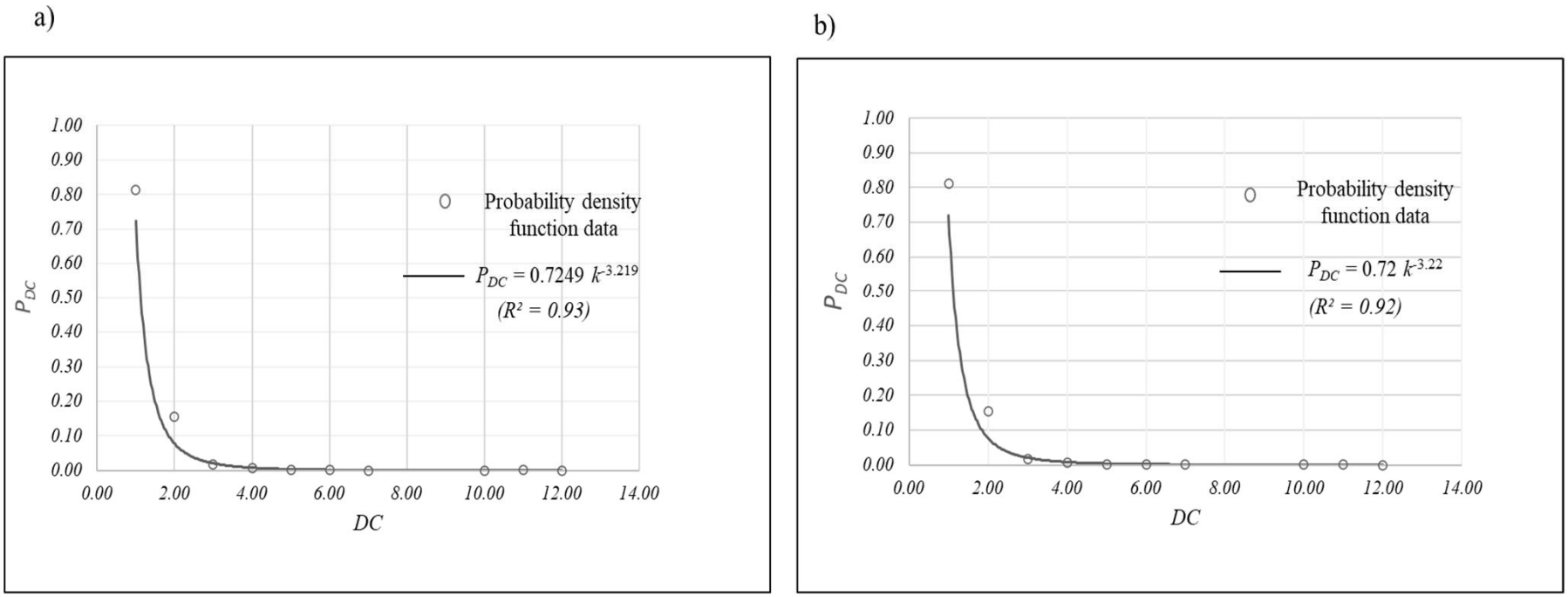

Figure 4: The probability density function data of the network: a) Outgoing degree (k_{out}); b) Incoming degree (k_{in})

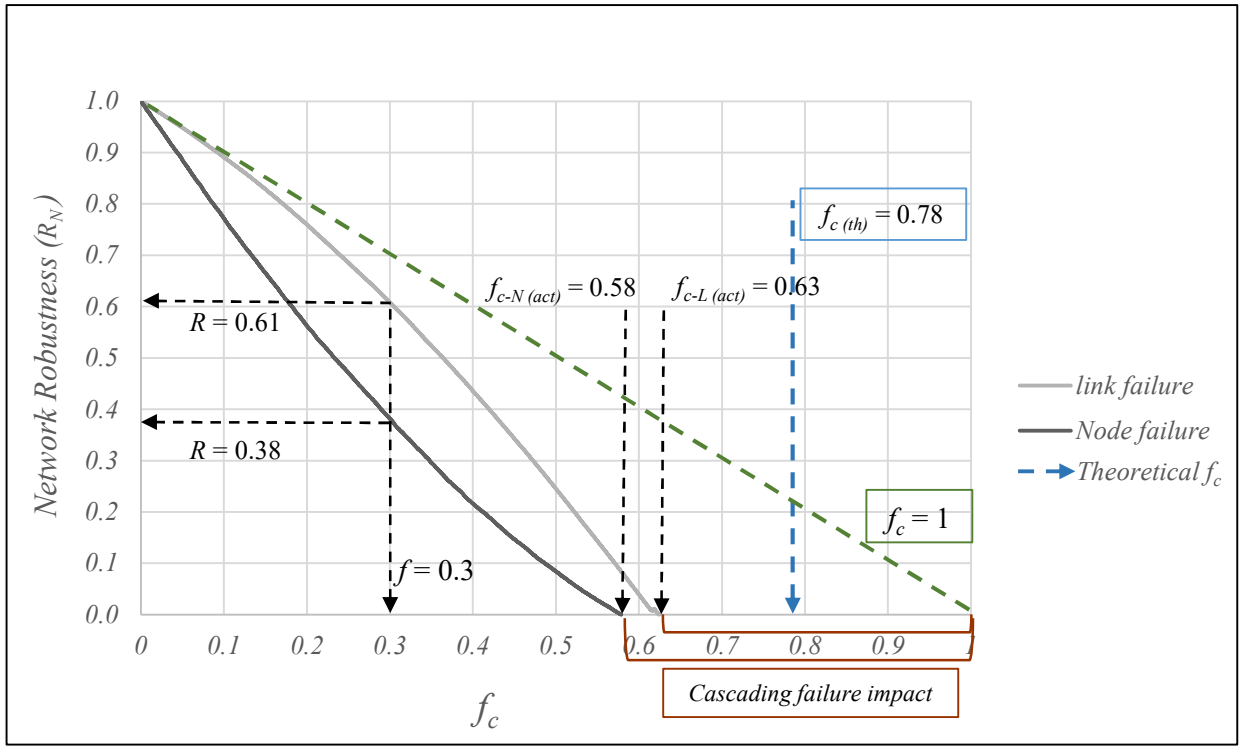

Figure 5: Robustness of Minneapolis local bus transit network (MLBTN) based on random node and link failures.

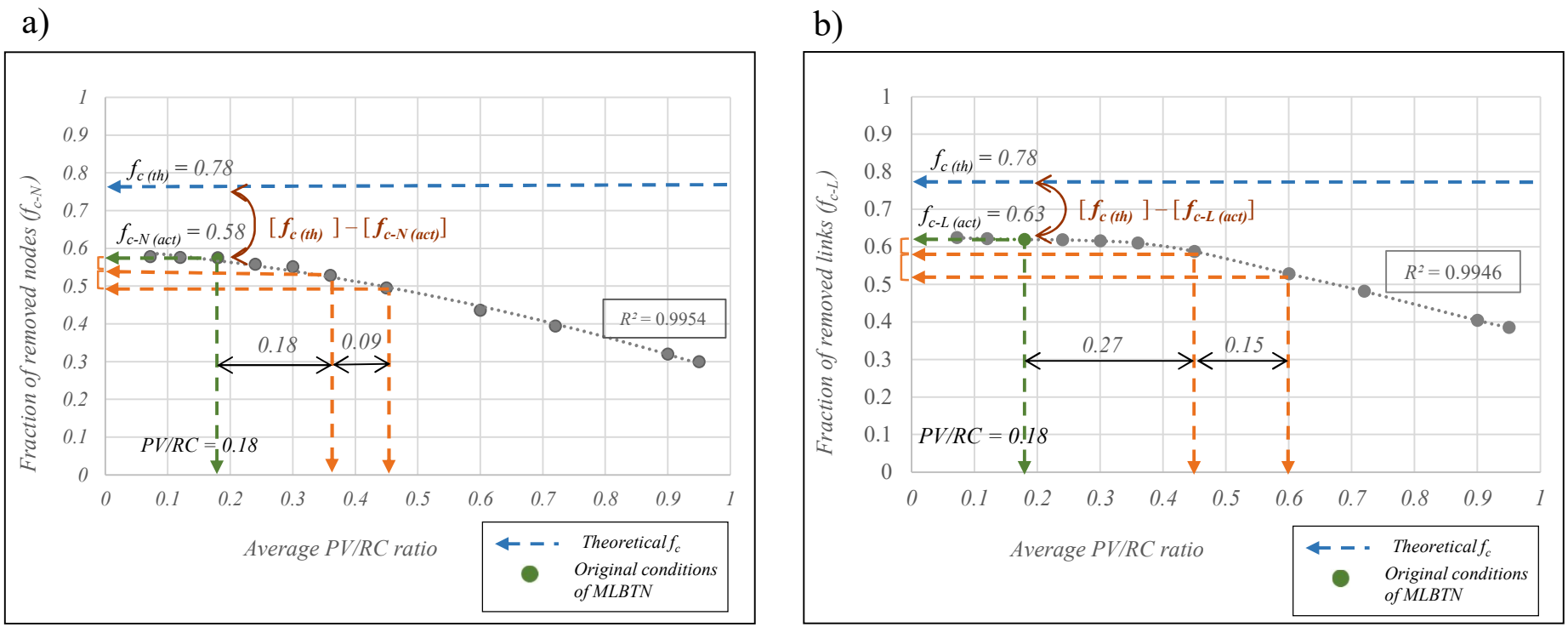

Figure 6: Robustness of the MLBTN under different PV/RC ratios: a) Random node failure; b) Random link failure.

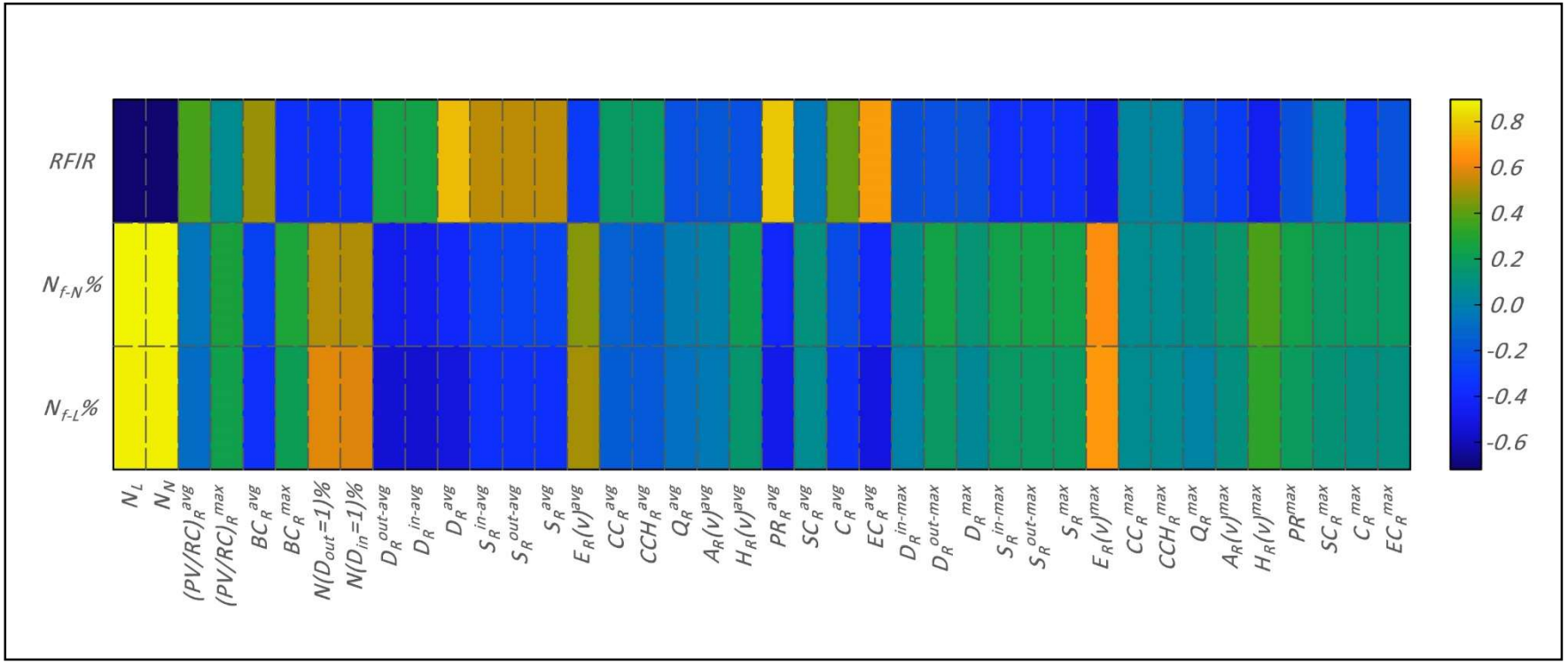

Figure 7: Correlation matrix.

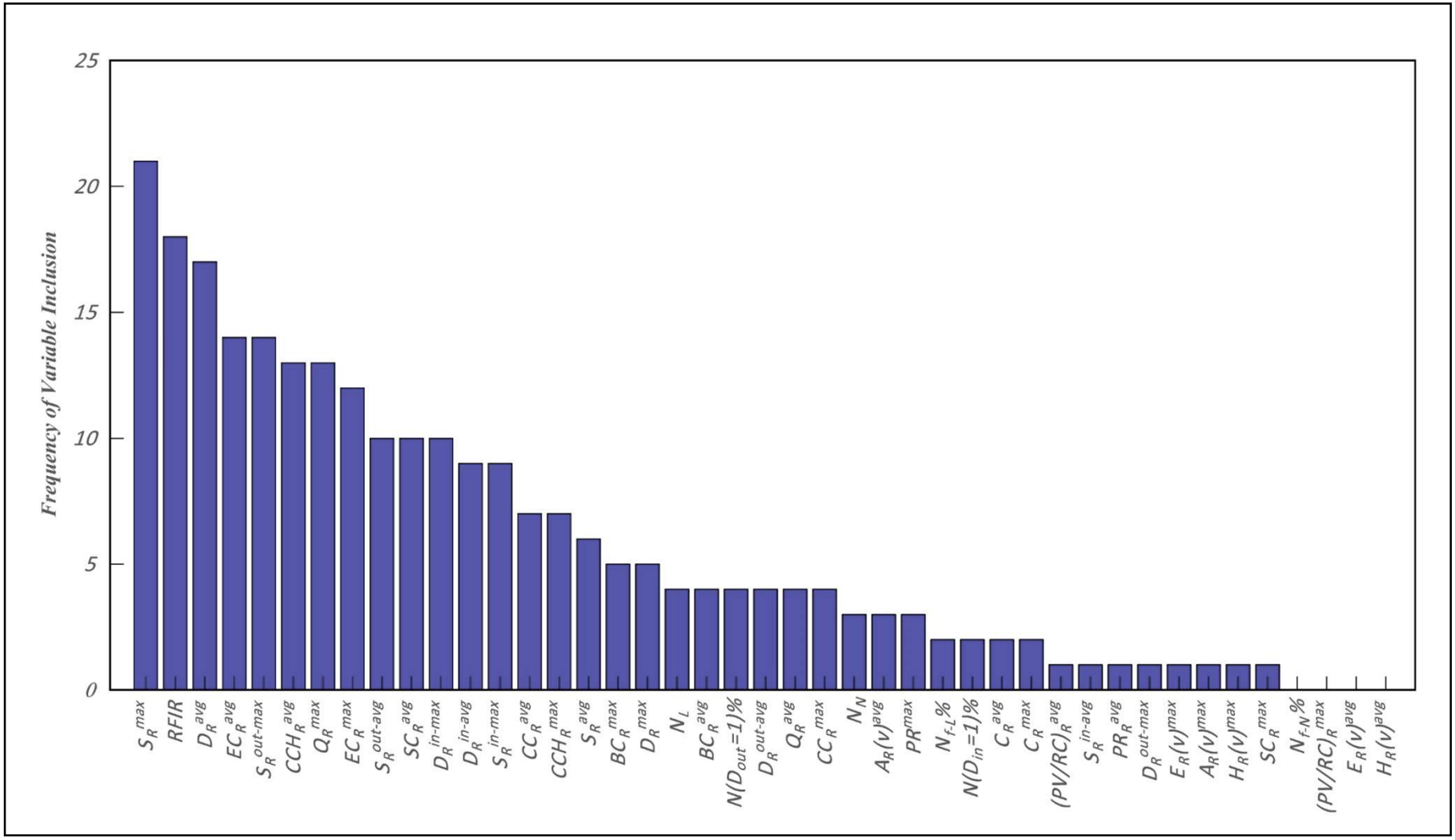

Figure 8: Frequency of variable inclusions in the GA solution

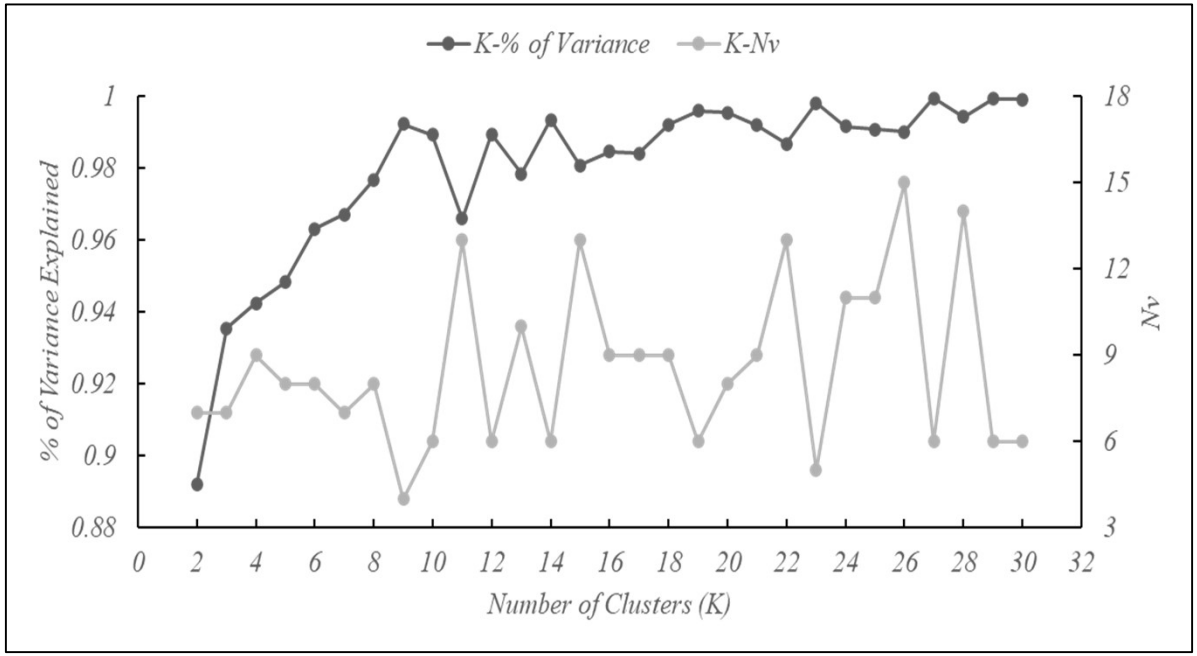

Figure 9: Results of the GA-based feature selection

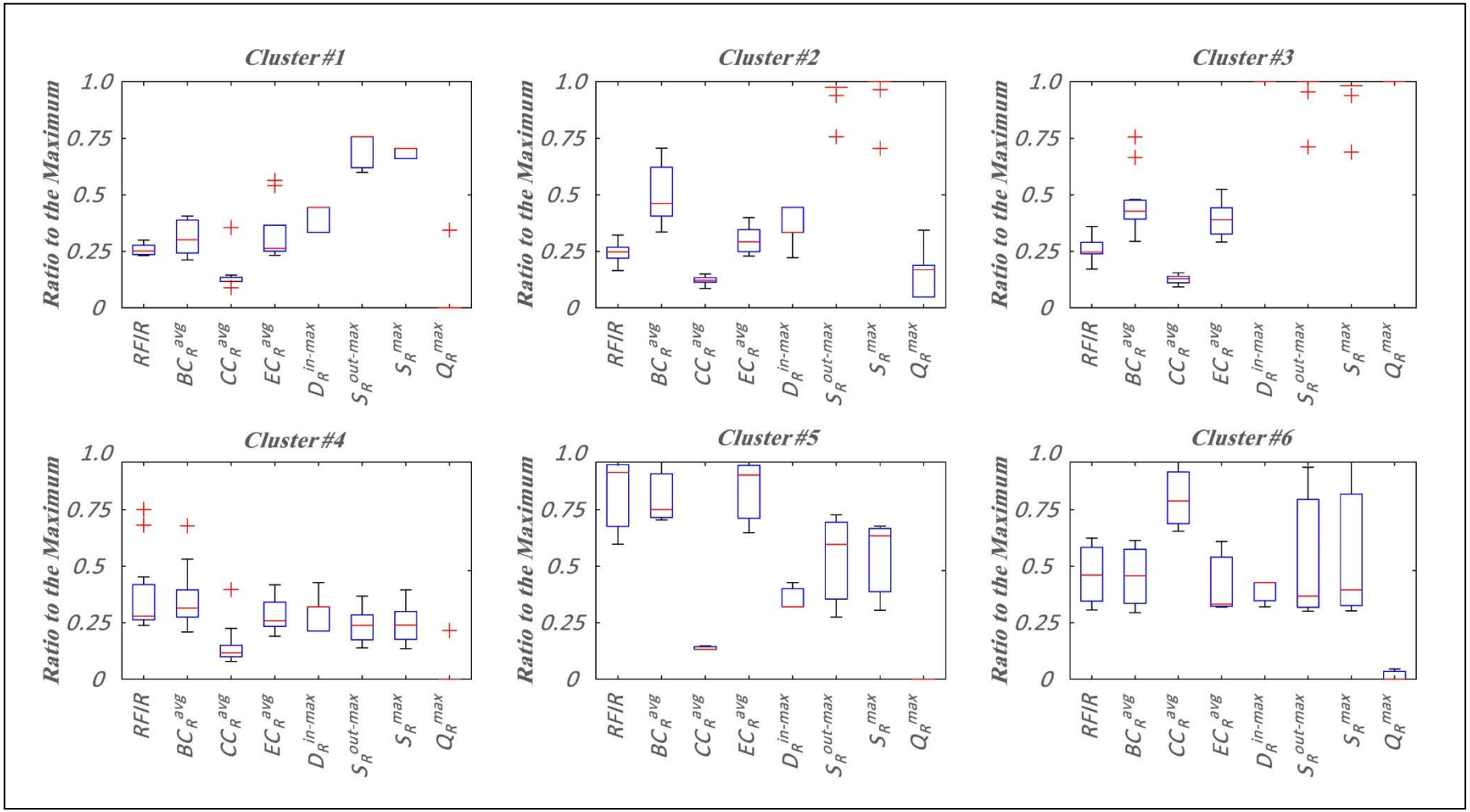

Figure 10: Clusters properties

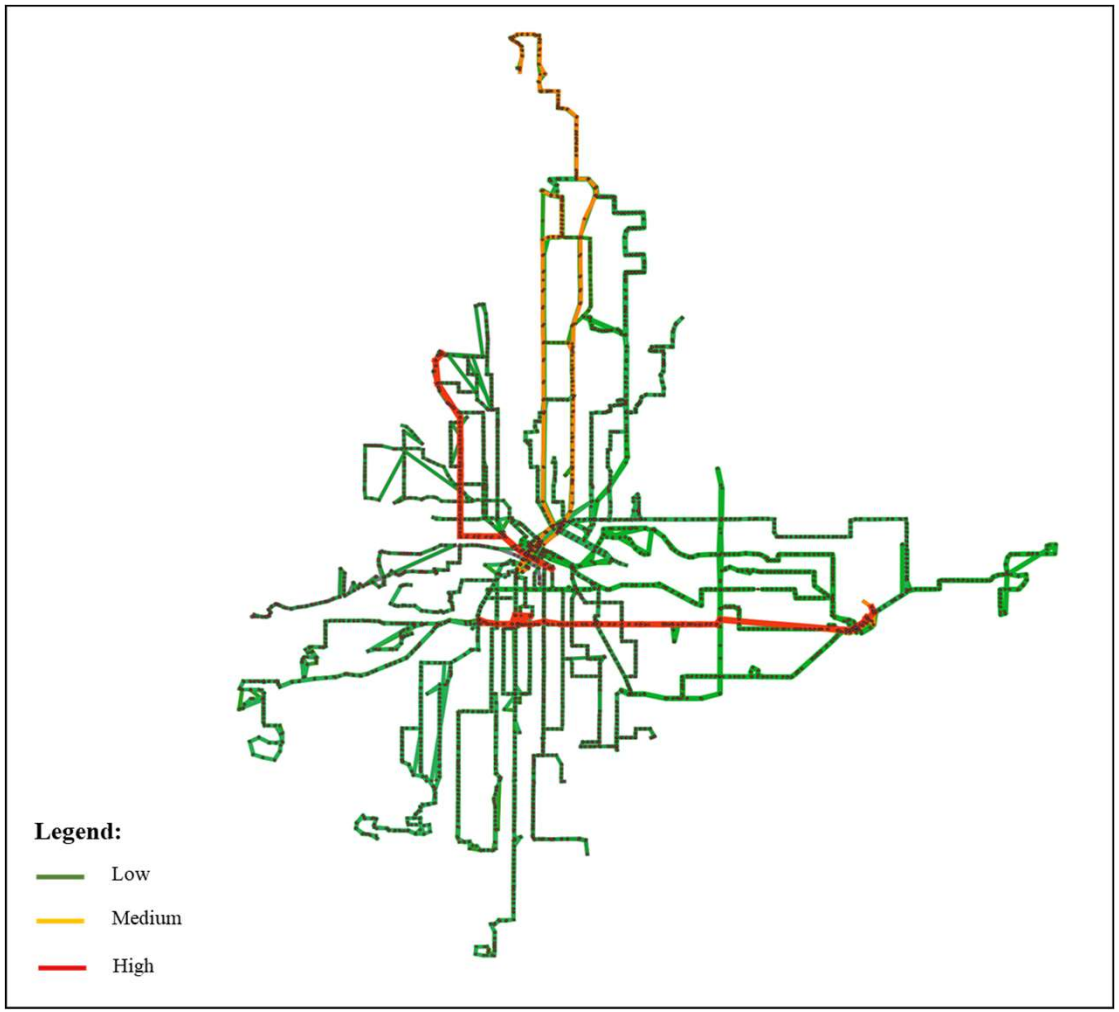

Figure 11: Classification of route impacts based on the clustering analysis