

STRUCTURAL AND DYNAMIC MODELS FOR COMPLEX ROAD
NETWORKS

A Thesis

Submitted to the Faculty

of

Purdue University

by

Jiawei Xue

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science in Engineering

May 2020

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF APPROVAL

Dr. Satish V. Ukkusuri, Chair

Lyles School of Civil Engineering

Dr. Shreyas Sundaram

School of Electrical and Computer Engineering

Dr. Suresh Rao

Lyles School of Civil Engineering & Agronomy Department

Approved by:

Dr. Duley M. Abraham

Head of the School Graduate Program

To my parents and my love.

ACKNOWLEDGMENTS

I would like to show my sincere gratitude to Prof. Satish Ukkusuri for two years of supervision. He provided me the chance to learn optimization, network science, machine learning and mathematics courses at Purdue University and offered good academic environment for me to do the research about electric vehicles, complex networks, and urban congestion.

I also express great thanks to my M.S. advisory committee members: Prof. Shreyas Sundaram and Prof. Suresh Rao. Prof. Sundaram taught me fundamental knowledge about network science and gave me valuable advice on complex road network modeling. And Prof. Rao gave me essential guidance to conduct complex network researches from a broad perspective.

Thanks to UMNI lab members: Dr. Xinwu Qian, Dr. Wenbo Zhang, Dr. Tho Le, Hemant Gehlot, Takahiro Yabe, Zengxiang Lei, Lu Ling. Since September 2018, Dr. Xinwu guided me to investigate the epidemic spreading, electric vehicle charging station location optimization, electric bus route design, and urban congestion analysis. His working style impressed and affected me a lot and I am appreciated to make some contributions to our researches. Moreover, Hemant Gehlot held the weekly meeting with me about the Braess's paradox in scale-free networks research since October 2019 and put great effort to help me to improve academic writing. Additionally, Zengxiang Lei taught me to use Python when I first came to Purdue University in 2018 and was willing to hear my research ideas.

In the end, I would thank my parents for financial support and my girlfriend for mental support. Their devotions constitute the foundation for me to study in the USA.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vii
LIST OF FIGURES	viii
ABSTRACT	xi
1 INTRODUCTION	1
1.1 Road Networks	1
1.2 Network Science	3
1.3 Research Objectives	4
2 LITERATURE REVIEW	5
2.1 Urban Street Networks	5
2.1.1 Primal models	6
2.1.2 Dual models	7
2.2 Traffic Dynamics	9
2.2.1 Basic traffic dynamic variables	9
2.2.2 Network-based traffic dynamic models	13
2.2.3 Dynamic packet routing road network	14
3 COMPLEX ROAD NETWORK STRUCTURAL MODELS	17
3.1 Introduction	17
3.2 Background	19
3.2.1 Complexity of road network	19
3.2.2 Related work	19
3.3 Node Merge Algorithm	20
3.4 Intersection Degree Distribution	23
3.4.1 Spatial degree distribution	23
3.4.2 Spatial weighted degree distribution	24
3.5 Data	25
3.6 Node Merge Results	27
3.7 Degree Distribution Results	31
3.7.1 Degree distribution	31
3.7.2 Weighted degree distribution	35
3.8 Conclusion	37
4 COMPLEX ROAD NETWORK DYNAMIC MODELS	38
4.1 Introduction	38
4.1.1 Previous work	38

	Page
4.1.2	Percolation phase transition 40
4.1.3	Macroscopic fundamental diagram 40
4.1.4	Road network partition 41
4.2	Road Network Dynamic Analysis Framework 43
4.3	Percolation Curve And Critical Value 44
4.4	Road Network Partition Algorithm 46
4.4.1	Modularity maximization and Louvain algorithm 46
4.4.2	Percolation-based road network partition 47
4.5	Data 51
4.6	Percolation Curves And Critical Value Results 51
4.6.1	Existence 51
4.6.2	More spatial and temporal experiments 56
4.7	Network Partition Results 60
4.8	Conclusion 64
5	COMPLEX ROAD NETWORK CONNECTIVITY OPTIMIZATION 66
5.1	Introduction 66
5.2	Background 67
5.2.1	Road management and maintenance policy 67
5.2.2	Equilibrium-based network optimization 68
5.2.3	General network optimization 68
5.2.4	Critical value 69
5.3	Road Network Improvement Optimization 70
5.3.1	Optimization problem definition 70
5.3.2	Method 1: simple allocation 72
5.3.3	Method 2: heuristic optimization 73
5.4	Data 74
5.5	Experiments 77
5.5.1	Parameters preparation 77
5.5.2	Method 1: simple allocation 78
5.5.3	Method 2: heuristic optimization 79
5.5.4	More experiments of method 2 81
5.6	Conclusion 83
6	CONCLUSIONS 86
6.1	Summary 86
6.2	Future Work 87
	REFERENCES 90
	VITA 95

LIST OF TABLES

Table	Page
1.1 Concise history of network science	4
3.1 Node number under different distance threshold δ	30
3.2 Node number changes for Beijing and Shanghai after node merging	31
4.1 Modularity results	63
5.1 Parameter setting for the genetic algorithm	80
5.2 Budget constraint binding results (unit: thousand dollars)	82

LIST OF FIGURES

Figure	Page
2.1 Four representations of the road networks [4]	5
2.2 Global road length distribution [8]	7
2.3 Network expansion in Groane near Milan [15]	8
2.4 The relationship between Γ and unloading time τ [20]	11
2.5 Observed delays and modeled delays [24]	12
2.6 Congestion links under different ρ at 8:30 AM in Melbourne, Australia [26]	13
2.7 Congestion spreading in San Francisco [31]	15
3.1 A sample road network in Beijing from OpenStreetMap	21
3.2 Intersections in the research area in Beijing	26
3.3 Intersections in the research area in Shanghai	27
3.4 A road network in Beijing before node merging	28
3.5 A road network in Beijing after node merging	28
3.6 A road network in Shanghai before node merging	29
3.7 A road network in Shanghai after node merging	29
3.8 Out and in degree distribution in Beijing	32
3.9 Total degree distribution in Beijing	32
3.10 Degree distribution in Beijing (within 6-th ring road) [36]	33
3.11 Degree distribution comparison	33
3.12 Out and in degree distribution in Shanghai	34
3.13 Total degree distribution in Shanghai	35
3.14 Spatial degree distribution with varying ω	36
4.1 Impact of distance between loop detector and intersection on MFD curve for Toulouse, France, on June 6, 2008 [54]	42
4.2 Traditional MFD traffic management technical route	43

Figure	Page
4.3 A new traffic management technical route	43
4.4 Two iterations of Phase I (Step I) and Phase II (Step II) in Louvain algorithm [62]	47
4.5 Temporal speed variation	51
4.6 Congested roads in Beijing at 11:46, December 8, 2015 (Tuesday)	52
4.7 Traffic state of roads with $\rho = 0.30$	53
4.8 Traffic state of roads with $\rho = 0.58$	53
4.9 Traffic state of roads with $\rho = 0.85$	54
4.10 Percolation curve in Sanyuanqiao at 11:46, December 8, 2015 (Tuesday)	55
4.11 Percolation curve in Sanyuanqiao at 06:02, December 8, 2015 (Tuesday)	57
4.12 Percolation curve in Sanyuanqiao at 18:14, December 8, 2015 (Tuesday)	57
4.13 Percolation curve in Xidan at 11:46, December 8, 2015 (Tuesday)	59
4.14 Percolation curve in Zhongguancun at 11:46, December 8, 2015 (Tuesday)	59
4.15 Selected region for network partition	60
4.16 Weighted degree in the dual network $G^d(V_d, E_d)$ in Chaoyang District in Beijing at 11:46, December 8, 2015 (Tuesday)	60
4.17 Edges in $G^d(V_d, E_d)$ in Chaoyang District in Beijing at 11:46, December 8, 2015 (Tuesday)	62
4.18 Road network partition results	64
5.1 Percolation curve in Sanyuanqiao at 18:14, December 8, 2015 (Tuesday)	70
5.2 Selected region for optimization	75
5.3 The traffic state with threshold 0.54	75
5.4 Percolation curve	76
5.5 Road segment length distribution	77
5.6 Relative speed distribution	77
5.7 Road improvement for Dongzhimen area at 11:46	79
5.8 Link investment using Method 1 ($B = 2.4M$)	81
5.9 Link investment using Method 2 ($B = 2.4M$)	81
5.10 Road improvement for Dongzhimen area at 18:14	83

Figure	Page
5.11 Investment results using method 2 for Dongzhimen area at 18:14	83
5.12 Road improvement for Xidan area at 11:46	84
5.13 Investment results using method 2 for Xidan area at 11:46	84
6.1 A gating strategy for protected road network N [45]	88
6.2 Relationship between speed u and density k for a motorway [71]	88

ABSTRACT

Xue Jiawei MSE, Purdue University, May 2020. Structural and dynamic models for complex road networks. Major Professor: Dr. Satish V. Ukkusuri.

The interplay between network topology and traffic dynamics in road networks impacts various performance measures. There are extensive existing researches focusing on link-level fundamental diagrams, traffic assignments under route choice assumptions. However, the underlying coupling of structure and dynamic makes network-level traffic not fully investigated. In this thesis, we build structural and dynamic models to deal with three challenges: 1) describing road network topology and understanding the difference between cities; 2) quantifying network congestion considering both road network topology and traffic flow information; 3) allocating transportation management resources to optimize the road network connectivity.

The first part of the thesis focuses on structural models for complex road networks. Online road map data platforms, like OpenStreetMap, provide us with reliable road network data of the world. To solve the duplicate node problem, an $O(n)$ time complexity node merging algorithm is designed to pre-process the raw road network with n nodes. Hereafter, we define unweighted and weighted node degree distribution for road networks. Numerical experiments present the heterogeneity in node degree distribution for Beijing and Shanghai road network. Additionally, we find that the power law distribution fits the weighted road network under certain parameter settings, extending the current knowledge that degree distribution for the primal road network is not power law.

In the second part, we develop a road network congestion analysis and management framework. Different from previous methods, our framework incorporates both network structure and dynamics. Moreover, it relies on link speed data only, which is

more accessible than previously used link density data. Specifically, we start from the existing traffic percolation theory and *critical relative speed* to describe network-level traffic congestion level. Based on traffic component curves, we construct A_{ij} for two road segments i and j to quantify the necessity of considering the two road segments in the same traffic zone. Finally, we apply the Louvain algorithm on defined road segment networks to generate road network partition candidates. These candidate partitions will help transportation engineers to control regional traffic.

The last part formulates and solves a road network management resource allocation optimization. The objective is to maximize critical relative speed, which is defined from traffic component curves and is closely related to personal driving comfort. Budget upper bound serves as one of the constraints. To solve the simulation-based nonlinear optimization problem, we propose a simple allocation and a meta-heuristic method based on the genetic algorithm. Three applications demonstrate that the meta-heuristic method finds better solutions than simple allocation. The results will inform the optimal allocation of resources at each road segment in metropolitan cities to enhance the connectivity of road networks.

1. INTRODUCTION

1.1 Road Networks

The last half-century has witnessed rapid growth in both urban road networks and travel demand in metropolitan cities worldwide. Traffic congestion is caused by the imbalance between the network supply and travel demand, bringing about negative effects like economical loss, decline in road capacity, air pollution, and road rage. For instance, road congestion is severer even in developed countries where the road infrastructure is most developed. The average annual individual congestion hours for 494 urban areas in the United States were 20, 38, 47, 54 hours in 1982, 2000, 2012, 2017, respectively [1]. In 2018, Boston was the most congested city in the USA with 164 congestion hours, causing a loss of \$ 2,291 per driver [2]. In developing countries, rapid urbanization gives birth to congested traffic on streets because the road network planning is not done properly. Therefore, it is important to understand the traffic congestion phenomenon on road networks.

To analyze traffic congestion, transportation researchers in the last several decades have deeply focused on traffic assignment problems. Note that the route choice assumptions in traffic assignment problems may not always apply because people's choices are sometimes unpredictable and irrational. In addition, there has been less work on considering the road congestion from the perspective of network topology and flow dynamics. The merit of considering network topology and traffic analysis is that we do not make any assumption on routing behavior of travelers. Moreover, the theoretical analysis that would be developed can also be applied to regional traffic management.

Consequently, formulating traffic dynamic models at the network level is a promising research direction. Specifically speaking, there are three major existing problems to investigate:

(1) How should the topology of road networks be modeled?

In metropolitan cities, the existence of structures like roundabouts, side roads, ramps, and overpass brings ambiguity in categorizing elements of a road network into intersections and segments. Note that these structures are crucial to regional traffic because traffic gridlocks and spillovers can happen in these structures, declining the total system performance. Thus, we need to classify nodes and edges clearly in network modeling.

(2) How to model traffic congestion considering both network topology and traffic dynamics?

Note that local traffic state variables (for e.g., flow, speed, and density) and global traffic measurement variables (for e.g., average speed), are both useful in depicting the traffic states in the city. But these variables ignore the coupling between flow and network structure. In addition, density data cannot be easily obtained for all road segments due to less penetration of detectors and equipment failure problems of detectors. Hence, it is better to quantify traffic congestion that is a function of both the topology and dynamics of the network. Additionally, the traffic data used in the function should be accessible in most cities, else, the results of theoretical work would be limited to a small number of cities. Therefore, we would use the data that is easily accessible.

(3) How to optimally allocate transportation management resources to improve network performance?

When urban congestion occurs, two straightforward ideas that arise in tracking congestion are to either increase infrastructure supply or to decrease the travel demand. However, Braess's paradox tells us that the addition of a new road might make the system performance worse if all travelers follow the user equilibrium (UE) routing strategy [3]. Also, commonly used demand adjustment methods including congestion

pricing, parking pricing, license plate rationing, transit oriented development (TOD), are strongly dependent on assumptions on human behavior. For these reasons, it is more appropriate to allocate transportation management resources from network optimization point of view. The optimization results that will be presented would provide the optimal spatial distribution of investment on each road segment.

1.2 Network Science

Network science is the knowledge about the representation, structure, and dynamics of the networks. A network is a data structure that consists of the entities that are referred to as nodes and the relationships between them that are modeled as edges. The node-edge system of a network can capture diverse internal and external incidents happening on the nodes and edges. In fact, a lot of systems in the real world can be modeled as networks. For example, researchers and their academic citation records, Internet websites and their link relationship, individuals and their friendships, all can be modeled as networks. Many theoretical frameworks in network science were established in the 1950s and since then network science researches has witnessed considerable development. Table 1.1 exhibits a brief history of research progress in this area. Various measures like degree, neighborhoods, centrality metrics (like degree centrality, eigenvector centrality, betweenness centrality, PageRank), clustering coefficients have been developed to describe the importance of different nodes. Researchers have proposed the concepts of the degree distribution, communities, largest components, diameter to characterize the overall structural properties of the networks, and have investigated the process like the epidemic spreading and flow percolation to understand various dynamics on networks. Furthermore, there have been significant advancements in other problems such as network alignment, community detection, motif recognition, graph summary. These advancements have equipped us with powerful tools to conduct the analysis of distinctive network-based systems.

Table 1.1.: Concise history of network science

Year	Researchers	Contributions
1959	P. Erdos, A. Renyi	Random network
1960s	S. Milgram	Six degree of separation
1998	D. J. Watts, S. Strogatz	Random graph generation model
1999	A. Barabasi	Scale-free network
2001	A. Barabasi	Complex network
2001	R. Pastor-Satorras	Epidemic on scale-free network
2010	M. E. J. Newman	Network: an introduction
2011	M. Barthelemy	Spatial network

1.3 Research Objectives

As mentioned above, conventional transportation analysis techniques are not sufficient for us to study traffic congestion considering both the network topology and traffic dynamics. Therefore, we would develop both structural and dynamic models for road networks. Our models would use easily accessible traffic data, would not make any assumption on behavior, and concentrate on the natural phenomenon of urban congestion. These models would help us to understand road dynamics at the network-level and distinguish the difference between road networks and general complex networks. In the context of realistic transportation management, we will present the transportation management resource allocation methodologies to support engineers for better operating the traffic.

2. LITERATURE REVIEW

2.1 Urban Street Networks

Urban street network modeling is one fundamental topic in transportation research. It serves as an initial step to network-level congestion analysis and road improvement. Without proper adoption of different road network modeling techniques, we may come across some inefficiencies or even biases during transportation system modeling and analysis. In the year 2018, paper [4] conducted a comprehensive literature review about the previous street network modeling approaches. These models are mainly composed of linear elements connecting locations. Below show four different representations for the road networks.

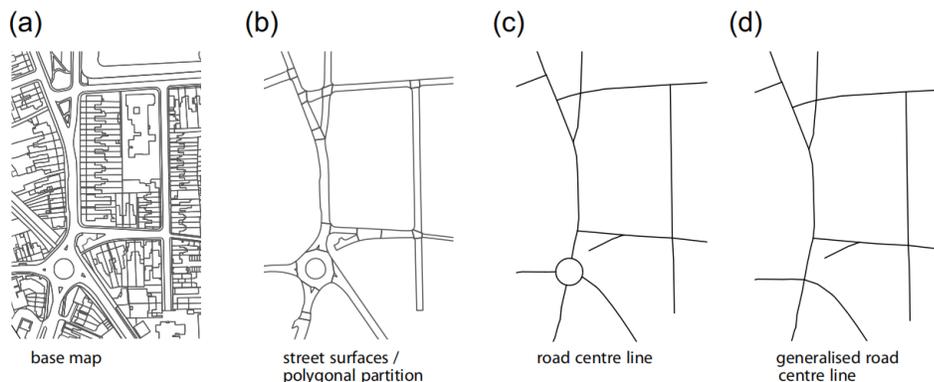


Figure 2.1.: Four representations of the road networks [4]

From the left to the right, the graph tends to be more sparse. Paper [4] found that sometimes the road modeling followed the disciplinary traditions and lacked specific motivations. For instance, the division of different intersections is ambiguous. Two proximate intersections are modeled as an intersection in some cases and separate intersections in other cases. There does not exist a quantitative standard to classify

the basic components in the road network, like junctions, links, and roundabouts. The situation would be worse in metropolitan areas with complex road networks.

2.1.1 Primal models

Generally speaking, most road network models are primal models where the road segments are edges and intersections are nodes. These models are intuitive, straightforward, and are widely used by scholars. One drawback of the primal model is that a long road with a unified property is divided into several parts, bringing about the inconvenience to describe the overall dynamics for the road.

Road network topology researches arise from the network modeling techniques. Paper [5], Paper [6], and Paper [7] used the primal representation to discover the road networks. The first research discussed road network statistics. Using a Python package OSMnx, paper [5] automatically downloaded and analyzed the road map for 27,009 USA street networks from OpenStreetMap, an open road geolocation data source. The author presented comprehensive network statistics: circuitry, clustering coefficient, average street length. It was showed: among the 30 urban areas, New York has the largest urbanized area of $8,973 \text{ km}^2$; the average street in Cleveland is longest (198 m). These results may not be absolutely accurate. Actually, OpenStreetMap receives online submissions of map drawings from personal users. The definition of roads and intersections vary from person to person. Some may incorporate narrow side roads in the submission but others may ignore them.

Besides, using primal representation, paper [6] investigated the centrality of road networks in 18 sampling cities. The authors calculated closeness centrality, betweenness centrality, straightness centrality, and information centrality. It was found self-organized cities belong to different centrality clustering. Paper [7] showed and compared the primal road network in Beijing inside the 2nd, 3rd, 4th, 5th ring road respectively. In the context of global road network, Paper [8] investigated length distribution for urban, cropland, and semi-natural roads. Figure 2.2 displays the

cumulative probability for these three kinds of roads, and the log-log road length distribution. It was revealed that urban, cropland, and semi-natural roads constituted respectively 12%, 51% and 37% of the total road length, and the average road lengths are 1.2 *km*, 7.4 *km*, and 7.0 *km*.

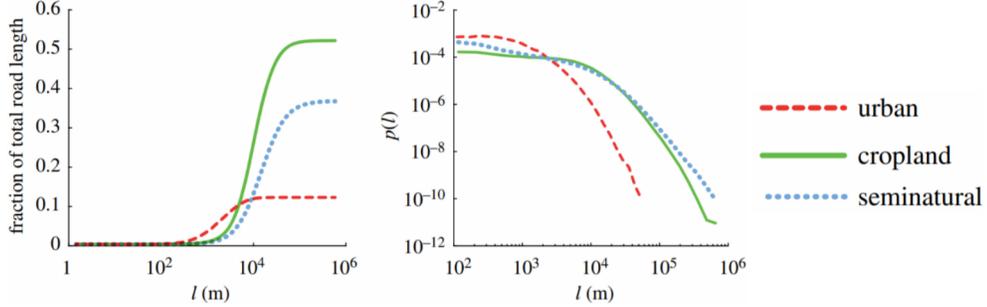


Figure 2.2.: Global road length distribution [8]

2.1.2 Dual models

To maintain the continuity of all segments of a long road, paper [9] proposed a dual presentation model, where the road segments are nodes and intersections are edges and developed the Intersection Continuity Negotiation to generalize the road network. Interestingly, the degree distributions of road network dual representations in several cities (for e.g., Beijing, Shanghai, Ahmedabad) follow the scale-free distribution [9], [10], which fits a wide set of networks like the World Wide Web, citation networks, airline networks [11]. The introduction of dual presentation enriches our modeling methods and also reveals consistent principles as other networks from another point of view. Both the primal and dual models have superiority over the other in certain aspects.

Under the dual representation, plenty of network properties were discovered. Paper [12] built a weighted dual graph for the road network in the UK and defined the weight as the relative angle of road segments. Afterward, the authors evaluated the road importance by calculating a proposed hierarchical index based on the weighted

dual graph. Paper [13] sampled roads in the United States, England, Denmark by querying the postcodes and then formulated the dual road networks. It was revealed that for large regions in the three countries, the degree distribution follows a power-law with the component $2.2 \leq \alpha \leq 2.4$. When it came to road connectivity, Paper [14] confirmed that it did not follow power-law strictly for the three cities, Gävle, Munich and San Francisco, as the log-log plot of the road connectivity versus cumulative probability is not linear. But the average path length between two nodes in the dual graph is not large, confirming the small-world property.

Other than the static road network, some efforts were put into the historical city road network expansion and growth discoveries. These researches made comparisons of previous and current road networks and came up with some patterns about the road networks.

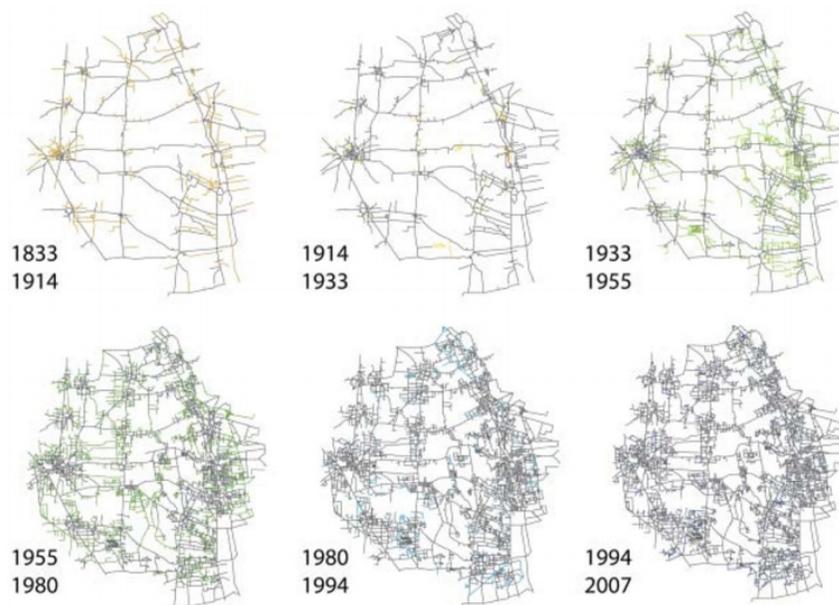


Figure 2.3.: Network expansion in Groane near Milan [15]

The number of roads increased during the urbanization period. Actually, not all the new roads have a unique functionality in the network: some facilitate the current connections for the road network while others enlarge the urban area. In other words,

the first kind of roads increment the density and the second kind of roads increase the urban spatial coverage. To quantitatively clarify these two roads, Paper [15] came up with two processes: *densification* and *exploration* based on results in Groane near Milan City in Italy. Generally speaking, the connectivity of a link e^* in the network is denoted by the amount of shortest paths that contain the link e^* . And the average connectivity $\bar{b}(G_t)$ of all the links in the network measures the importance of the links in the network at an aggregated level. The authors calculated the relative connectivity decrease $\delta_b(e^*) = \frac{\bar{b}(G_t) - \bar{b}(G_t \setminus e^*)}{\bar{b}(G_t)}$. Numerical results indicate that the distribution of the $\delta_b(e^*)$ has two peaks which represent respectively densification and exploration. In addition, the case study of Maynooth in Ireland also confirmed the existence of two peaks in the connectivity decrease distribution [16].

Another way used in road network development pattern discovery is kernel density estimation (KDE). Paper [17] calculated the KDE index for the Beijing road network from 2000 to 2010. The authors defined the *kernel area* as the areas where the road network kernel density is not less than $6 \text{ km}/\text{km}^2$. Results showed that the Beijing road network experienced more rapid growth from 2005 to 2010 than 2000 to 2005 as the kernel area in the year 2000, 2005, 2010 are 356 km^2 , 753 km^2 , 1949 km^2 .

2.2 Traffic Dynamics

2.2.1 Basic traffic dynamic variables

The aforementioned road network models used the road topology information only. In fact, road dynamic researches considering both road topology and traffic dynamics are more closely related to traffic congestion alleviation.

In transportation engineering, average speed v , road density k , and traffic flow q describe traffic for a road segment. A rough result about the interplay between the three variables is $q \sim kv$: the high density and the fast vehicle movement result in large traffic flow. Among the three variables, velocity v is an accessible indicator. Transportation engineers are able to obtain the vehicle velocity information via GPS

devices, like smartphones. At present, when people utilize some routing recommendation system, for example, Google Maps, the location information is recorded by GPS. Later, the system can calculate the average speed of this road segment by integrating the messages from different users traveling on the same road segment. Traffic flow q and road segment density k are not easy to get as the corresponding equipment, like loop detector, camera, is not as prevalent as GPS devices nowadays.

The average vehicle velocity in the city changes periodically. Paper [18] and [19] worked on general pattern based on the evolving velocity data. Specifically, Paper [18] defined the congested road segment whose average velocity is below a threshold p_i and characterized the cluster size as the integration of the number of congested links $S = \int_{t_0}^{t_1} M_s(t)dt$. They found that the cluster size follows the scale-free distribution: $P(S) = S^{-\alpha}$, where α is a positive component. For different cities, the α is different. Similarly, paper [19] classified the roads into functional (uncongested) and disfunctional (congested). The authors found that under certain relative speed threshold q , the functional road segments make up several traffic clusters. And the improvement of the bottleneck roads which bridge the traffic clusters enhances the system performance greatly.

When other variables like the number of vehicles are considered, researchers exhibited more complicated patterns. Paper [20] evaluated the road network system performance by two variables: unloading time τ and demand-to-supply level Γ . Firstly, the authors concentrated on a target group of vehicles entering the road network and tracked the vehicles that still in the network at the following moments. They found that the vehicle number decreases exponentially: $N(t)_{t \geq t_0} = N(t_0)e^{-\frac{t-t_0}{\tau}}$. Here, τ indicates the network response to congestion as high τ means vehicles in the target group arrive at the destination slowly. Secondly, the authors defined the demand-supply ratio Γ as the ratio of vehicle miles travel (VMT) to maximal VMT. The authors found that there is a linear relationship between τ and Γ (see Figure 2.4).

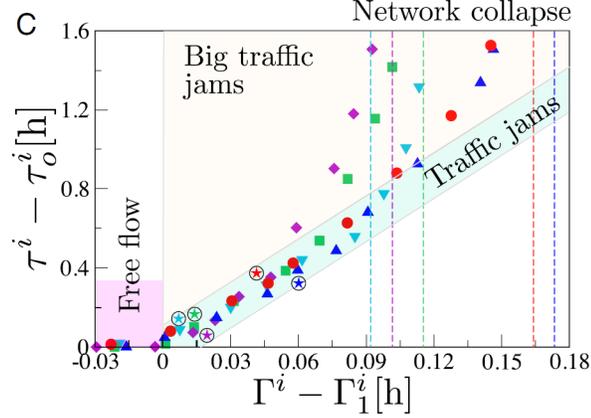


Figure 2.4.: The relationship between Γ and unloading time τ [20]

Some researchers used the mobile phone location data to mine the road congestion or usage characteristics. Mobile phone data, sometimes referred to as call detail record (CDRs), reflects human movement in the urban area with the advantage of its high penetration. Paper [21] tracked the source place of vehicles passing one certain road segment. They built a bipartite network, where the degree of the road segment node K_{road} is the number of major driver sources that produce 80% of the travel demand, and the source degree K_{source} is the number of road segment for which the driver source is a major contributor. One finding was that K_{road} follows a log-normal distribution, and K_{source} is normally distributed with a long tail, implying that only a few driver sources contribute to the majority of the road usage. This result highlights the idea to use economic policies to alert the travel behavior of these driver sources. Also using the mobile phone data, Paper [22] defined the demand-to-supply index Γ , which is the ratio of the total traveling distance of all vehicles over the upper bound value:

$$\Gamma = \frac{\sum_{e \in E} l_e x_e}{\sum_{x_e > 0, e \in E} l_e C_e} \quad (2.1)$$

And it was shown quantitatively the travel time is related to the Γ value.

Traffic flow q is also used to describe the congestion state for the road network in some cases. In the complex network dynamic research, researchers would use the link betweenness centrality as an approximation of the traffic flow for each link, when

lacking the link flow data. When it comes to road networks, this assumption does not hold [23]. The researchers used the road shapefile data, and calculated the traffic flow from the taxi GPS trajectory in Qingdao, a coastal city in China. It was obtained that the correlation coefficients between flow F_c and primal betweenness C_P^B , F_c and dual betweenness C_D^B are only 0.416, 0.186. Afterward, the authors incorporated information using mobile phone Erlang values and distance-decay effect with optimal distance parameter β . The weighted correlation coefficient (WCC) reached to 0.623, which was a huge improvement compared with pure link betweenness centrality. This study showed the weakness of the network topology index in representing flow dynamics. Therefore, estimated or realistic traffic values are essential in congestion research. A typical case was [24]. The researchers measured the link travel time by three steps: 1) assume that the regional traffic demand depends on the population; 2) apply the gravity model to generate the OD; 3) estimate the traffic speed using the cell transmission model [25]. The experiments on 40 U.S. urban areas showed that this process generated the annual trip delay results close to reality.

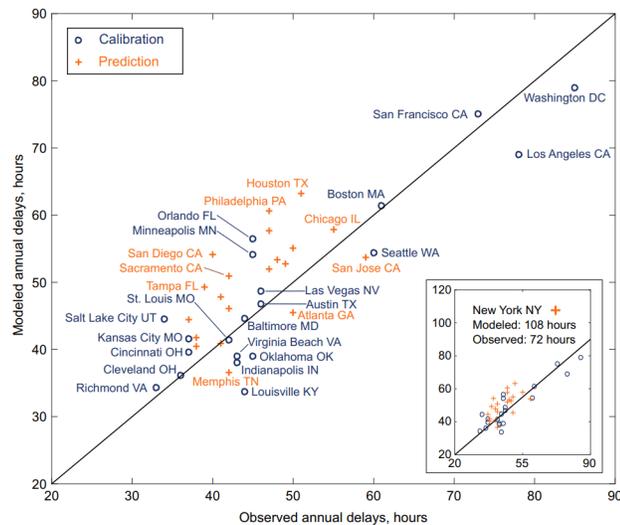


Figure 2.5.: Observed delays and modeled delays [24]

2.2.2 Network-based traffic dynamic models

Network-based traffic dynamic models are also prevalent among researchers. As mentioned above, in real road networks, the link state is depicted by three continuous variables: speed v , density k , and traffic flow q . Sometimes scholars used the relative speed ratio, to describe the traffic state. It is defined as [26]:

$$r_i(t) = \frac{v_i(t)}{v_i^{max}} \quad (2.2)$$

Here, v_i^{max} is the speed limit of the road segment or the 95 percentile largest speed. Hereafter, the binary congestion index is defined as:

$$c_i(t) = \begin{cases} 1, & r_i(t) < \rho \\ 0, & r_i(t) \geq \rho \end{cases} \quad (2.3)$$

where ρ is the predefined threshold. $c_i(t) = 1$ implies that it is a congested road and $c_i(t) = 0$ means that it is not congested. We can characterize whether the road segment is congested via $c_i(t)$. This congestion definition method is used in Paper [19], [27], [28].

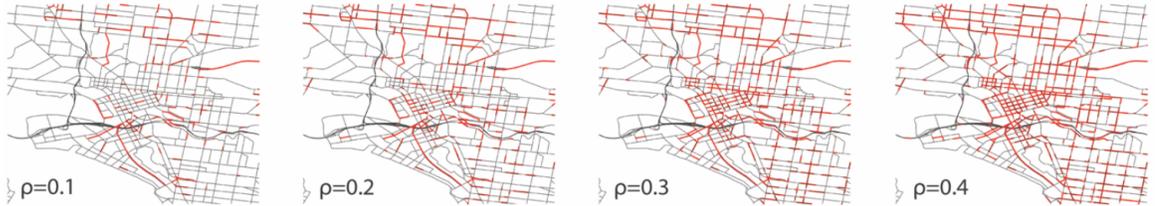


Figure 2.6.: Congestion links under different ρ at 8:30 AM in Melbourne, Australia [26]

For the road network system, the link interactions happen at the intersections, including signalized and unsignalized. An intuitive and widely used assumption of road congestion dynamics is that: the roads near a congested road segment have a high probability to become congested. Based on this assumption, Paper [26], [27], [28], formulated the congestion dynamics models for road networks.

Specifically, Paper [27] built the ordinary differential equation (ODE) for road segment density dynamics, where the change rate of density ($\dot{x}_j(t)$) is the result of incoming flows minus outgoing flows. In this model, the stochastic matrix $\Pi = [\pi_{kj}]$ contains link segment interaction information.

$$\dot{x}_j(t) = \sum_k \theta_\gamma (1 - x_j) \pi_{kj}(t) x_k(t) (1 - x_k(t))^\alpha - \sum_k \theta_\gamma (1 - x_k) \pi_{jk}(t) x_j(t) (1 - x_j(t))^\alpha \quad (2.4)$$

Paper [28] converted the dynamics in the road network into the structural dynamics under the dual presentation: a long road splits into several road segments as the congested road segments *cut* the whole road. The author developed a macroscopic model to describe the number of splits for a road segment, and a microscopic model to reflect where a certain number of splits take place.

Analogously, Ph.D. thesis [26] inherited the idea of an epidemic spreading model, which is a well-known dynamic model in the complex network, to model the onset, spread, and recovery of congestion in the urban area. In this model, each road segment has three states: free-flow, congested, recovered, similar to the three states in the susceptible-infected-recovered (SIR) model [11].

$$\frac{ds}{dt} = -\beta \langle k \rangle i(1 - i - r) \quad (2.5)$$

$$\frac{di}{dt} = -\mu i + \beta \langle k \rangle i(1 - i - r) \quad (2.6)$$

$$\frac{dr}{dt} = \mu i \quad (2.7)$$

They found that, when the congestion threshold ρ is low, different cities have nearly the same spreading rate $\lambda = \frac{\beta}{\mu}$.

2.2.3 Dynamic packet routing road network

Compared with the complex road network, the investigation of the general complex network has a long history. In the general Internet network model (IM), at each time moment, each node is assigned several particles with different destinations. The

release of particles follows *first-in, first-out* (FIFO) strategy. And the particles travel to the destination node under a specific routing strategy. A widely used network performance measurement η [29] is:

$$\eta = \lim_{t \rightarrow \infty} \frac{\langle \Delta\Theta \rangle}{\lambda\Delta t} \quad (2.8)$$

where $\Delta\Theta = \Theta(t + \Delta t) - \Theta(t)$, and Θ denotes the number of packets in the network, λ is the packet generation rate. In fact, the flow dynamics in the road network is not exactly the same as the particle transmission in the computer network, because: 1) The flow in road network proceeds in the links and travel time varies among different links; 2) Each link has a flow capacity. In the Ph.D. thesis [30], Serdar Colak discussed the point-queue model (PQM) to address the first gap and then proposed a spatial point-queue model (SPQM) to consider flow capacity constraints. And the critical packet generation rate λ for the three models are respectively:

$$\lambda_c^{IM} = N(N - 1)(C_{max}/B_{max}^N) \quad (2.9)$$

$$\lambda_c^{PQM} = N(N - 1)(C_{max}/B_{max}^{E*}) \quad (2.10)$$

$$\lambda_c^{SPQM} = N(N - 1)\min\left(\frac{C}{B_{max}^{E*}}, \frac{V}{(\tau B^E + B^{E*})_{max}}\right) \quad (2.11)$$

where N is the node number, B_{max} is the betweenness of nodes, and C_{max} is the node flow capacity, B^E is the edge betweenness. The author conducted a numerical

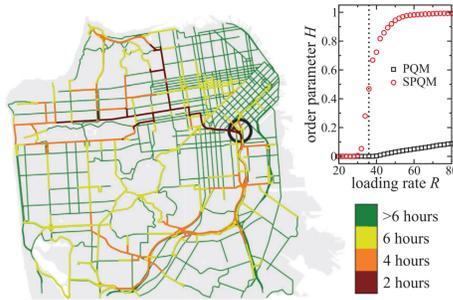


Figure 2.7.: Congestion spreading in San Francisco [31]

analysis for San Francisco. The critical loading rate for PQM and SPQM models are

respectively: 40 (14,400 vehicle⁻¹) and 30 (10,800 vehicle⁻¹). When the real vehicle number exceeds the critical loading rate, the congestion occurs in the black circle region and then spreads over the whole network.

3. COMPLEX ROAD NETWORK STRUCTURAL MODELS

3.1 Introduction

Metropolitan cities have complicated road networks where the traffic jams cause considerably negative effects on both the individuals as well as the whole society. To alleviate traffic congestion and make better use of urban road infrastructures, it is needed to model the road system and uncover the topology property of the road network properly. A typical complex network comprises two elements: nodes (vertices) and edges (links), whose interaction is analogous to the interplay of intersections and road segments in urban areas. In this chapter, we utilize the complex network method to investigate the property of road networks in metropolitan cities. We develop a road shapefile data preprocessing algorithm with complexity $O(n)$, where n is the number of intersections. Hereafter, we build a complex network for the real road network and define the in-degree and out-degree. Furthermore, the weighed degree distribution model is proposed when the length and width of road segments are taken into consideration. Numerical experiments on the road networks in Beijing and Shanghai confirm that the proposed algorithm merges the nodes accurately and efficiently. We also compare the road network in Beijing and Shanghai based on the obtained degree and weighted degree distribution. Finally, we examine the intersection degree distribution with varying parameter ω , and reach the conclusion that under large ω the degree distribution is fitted well by the power-law. This research provides a better description of road networks in cities.

The number of vehicles in the world has increased rapidly during the last tens of years. Many metropolitan cities, especially where the travel demand increase greatly, suffer from severe traffic congestion. For U.S. urban regions, the average delay per commuter has increased from 20 hours in the year 1982 to 54 hours in the year

2017 [1]. Traffic jams result in working hours decline for citizens, more gasoline energy consumption, and more air pollution. To alleviate the congestion and make citizens live better, it is quite needed to prioritize modeling both road networks and further congestion dynamics in metropolitan cities in an explanatory and applicable way.

Complex network, as one of the most fundamental data structures, serves as a useful tool to model and analyze the large-scale system. A standard network is composed of the nodes and corresponding edges, where nodes and edges represent the individual objects and their iterations, respectively. There are various examples of systems that can be modeled as a complex network. For instance, for human social relationships, each person is modeled as one node, and their friendship (following in Twitter, etc) is modeled as an edge.

As a concise and intuitive model, the complex network offers large convenience to capture both the structure and dynamics features for complex systems. The structure here implies the relationship between nodes and edges, and the dynamic features rely on the network structure and govern the information, traffic, message flow from node to node in the network.

In particular, the road system in a metropolitan city is a typical example that can be modeled as a complex network. The road system has two components, road segments, and intersections (including both signalized and unsignalized ones). The characteristics of car-following and lane-changing behaviors within road segments, vehicle routing strategies in intersections, make up the backbone of dynamics in the road systems, which is relatively analogous to node-edge iterations in the complex network. Accordingly, the complex network serves as a proper tool to model realistic road networks. Utilizing the complex network to investigate the structure of road networks and congestion dynamics helps us to manage the traffic in large cities better.

3.2 Background

3.2.1 Complexity of road network

In modern society, roads in metropolitan cities are more intricate than before. Existence of overpass, tunnels, roundabouts, ramp roads, one-way roads causes some barriers to characterize the road networks as it lacks a unified standard to define the intersections and road segments quantitatively. A small T-junction joining an avenue may be considered as an independent intersection or may not. Four-direction roundabout can be modeled as one node or four different nodes. What's more, the resolution of the road dataset varies from one city to another. In some shapefiles, the ramp roads are stored as independent road segments. But in other cases, they are part of the main roads. Ambiguous classification of road infrastructures brings difficulties to road network modeling. For the aforementioned two reasons, it is essential to define the intersection and road link standard properly. Otherwise, the comparison between cities using different types of shapefiles is not convincing.

3.2.2 Related work

There is a collection of literature focusing on road network modeling. G. Boeing adopted the OSMnx software to automatically download the road shapefiles of US cities from the OpenStreetMap [5]. He conducted a comprehensive statistical analysis of the road using the indexes like intersection density, average street per node, average street segments. Using the historical street data in *Groane* area, a town near Milan in Italy, E. Strano et al. classified the new roads during the urbanization into two categories (densification, exploration) based on the betweenness centrality [15]. Moreover, S. Porta et al. surveyed a wide range of previous modeling methods for road networks and proposed the dual representation, in which the road segments are modeled as nodes, while the intersections are modeled as edges in the network [9]. When it comes to the traffic states in the road network, Daqing Li et al. character-

ized the velocity data in the area surrounding Beijing West railway station. They discovered the relation between percolation and congestion index which is defined as the average speed of vehicles divided by the 95th percentile of vehicle speed for the road segment in a day [19].

However, the existing research on the road network modeling does not offer the definition of road intersections and segments explicitly. Most researches mention the road dataset they use but do not emphasize the data preprocessing process in the paper. Without a reasonable definition and a unified data preprocessing technique, the road network and dynamic research may induce unpredictable bias. For instance, in some shapefiles, several nodes represent one intersection in reality. Analysis of velocity for the small edges between these nodes is meaningless as they are not the actual links but part of an intersection. Consequently, it is necessary to conduct data preprocessing on the road shapefile before road network modeling. Our analysis is as follows: we propose a δ -based node merge algorithm. After that, we define the degree distribution and weighted degree distribution when considering the length and width of the road segments.

3.3 Node Merge Algorithm

City road shapefiles are widely used in road network researches. Primitive shapefiles from open websites sometimes have *duplicate node problem*: one intersection is represented by multiple nodes in the shapefile. This phenomenon is likely to happen for large-scale intersections in metropolitan cities. Figure (3.1) is a sample network in Beijing downloaded from OpenStreetMap via OSMnx Python package ¹ where the duplicate node problem exists. Without data preprocessing, we will get the wrong results about intersections and road segments, as well as other network properties like connectivity and closeness, because it is actually one node instead of several dif-

¹<https://geoffboeing.com/publications/osmnx-complex-street-networks/>

ferent nodes. Hence, we are going to cluster different nodes that represent the same intersection in reality.

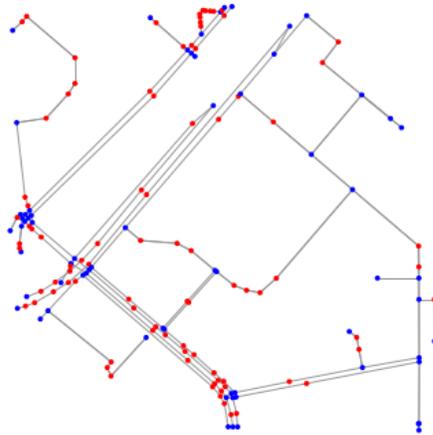


Figure 3.1.: A sample road network in Beijing from OpenStreetMap

The problem is described as follow: given a graph $G = \{V, E\}$ which has n nodes with longitude and latitude (x_i, y_i) and given a distance threshold δ , find all pair of close nodes $v_1, v_2 \in V$ such that

- 1) has a distance less than δ : $d(v_1, v_2) < \delta$ or
- 2) has a k -element path ($k > 2$) with all segment length smaller than δ : $d(v_1, v_3) < \delta, d(v_3, v_4) < \delta, \dots, d(v_{k-1}, v_k) < \delta, d(v_k, v_2) < \delta$.

As mentioned before, there might be several nodes in the shapefile representing one intersection in reality. Here, we define one intersection as a connected component of close nodes and the geolocation of this intersection is determined by the average of the close nodes. This definition makes sense as distances for two different intersections (more than 100 meters) are always larger than the intersection diameters (less than 100 meters).

The silly algorithm to solve the node merge problem is to compare distances of all node pairs in G with δ , which is computationally expensive (complexity $O(n^2)$). And a metropolitan city like London, Beijing has more than 10,000 nodes in the city shapefile. Here, we propose the following node merge algorithm:

Algorithm 1 node merge algorithm

Input: $(x_1, y_2), (x_2, y_2), \dots, (x_n, y_n)$, earth radius $R= 6373$ km, distance threshold δ .

Output: $(x'_1, y'_1), (x'_2, y'_2), \dots, (x'_k, y'_k), k \leq n$

1: the edge set $E_\delta \leftarrow \Phi$

2: obtain the minimum and maximum of $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$: $x_{min}, x_{max}, y_{min}, y_{max}$.

3: choose maximal integer n_x and n_y such that:

$$\frac{x_{max} - x_{min}}{n_x} (2R \cos(y_{max}) \pi \frac{1}{360}) \geq \delta \quad (3.1)$$

$$\frac{y_{max} - y_{min}}{n_y} (\frac{\pi}{180} R) \geq \delta \quad (3.2)$$

4: **for** $1 \leq i \leq n$ **do**

5: $n_{i,x} \leftarrow \lfloor \frac{x_i - x_{min}}{\frac{x_{max} - x_{min}}{n_x}} \rfloor$, $n_{i,y} \leftarrow \lfloor \frac{y_i - y_{min}}{\frac{y_{max} - y_{min}}{n_y}} \rfloor$.

6: assign node i into the $(n_{i,x}, n_{i,y})$ -grid

7: **end for**

8: **for** $0 \leq n_1 \leq n_x, 0 \leq n_2 \leq n_y$ **do**

9: $N_{ij} \leftarrow$ all nodes in the (n_1, n_2) -grid

10: $N_{ij}^{neighbor} \leftarrow$ all nodes in the surrounding 9 grids of

11: the (n_1, n_2) -grid (include itself)

12: **for** node u in N_{ij} , node v in $N_{ij}^{neighbor}$ **do**

13: **if** distance(u, v) $< \delta$ and (u, v) not in E_δ **then**

14: $E_\delta \leftarrow E_\delta \cup \{(u, v)\}$.

15: **end if**

16: **end for**

17: **end for**

18: **for** node i in V **do** $Reach(i) \leftarrow \{i\}$

19: **end for**

20: **for** edge (i, j) in E_δ **do**

21: $MergeReach \leftarrow Reach(i) \cup Reach(j)$.

22: refresh the reach set of all elements in $MergeReach$

23: **end for**

24: Find the center of of different Reach sets: $(x'_1, y'_1), (x'_2, y'_2), \dots, (x'_k, y'_k)$

25: **return** $(x'_1, y'_1), (x'_2, y'_2), \dots, (x'_k, y'_k)$

In algorithm 1: node merge algorithm, Line 2 is to get the boundary of the nodes. Line 3 is to choose the proper grid number n_x, n_y such that grid size is not smaller than the distance threshold δ . Line 4-7 is to map each node into the grid. Line 8-17 is to check the distance of nodes in the adjacent grids (if two nodes are not in the adjacent grid, the distance is larger than δ as the grid size is larger than δ). Line 20-23 is to find the component based on the *close* relationship. Line 24 is to obtain the centers of different components. Line 25 returns the results.

Now we analyze the complexity of algorithm 1: node merge algorithm. First, we claim that the number of edges in E_δ is $O(n)$. In fact, node i is adjacent to node j that has a small distance to node i . Due to spatial restriction, the number of node i 's neighbors has an upper bound K . So the number of edges in $E_\delta \leq \frac{1}{2}Kn = O(n)$. Then, we denote that in the unit area, there are u nodes in the city, which is a upper-bounded number. Then the total area of the city is about $\frac{n}{u}$. Then the total number of grids is about $\frac{n}{u*(\delta*\delta)}$. Lines 2 and 4-7 have the complexity of $O(n)$. For lines 8-17, there are $\frac{n}{u\delta^2}$ grids. For each grid, the distance comparison takes $u\delta^2*(9u\delta^2)$ times. So the total comparisons take $O(\frac{n}{u\delta^2}(9u^2\delta^4)) = O(nu\delta^2) = O(n)$ times. And the merging (lines 20-23) and centering (line 24) both take $O(n)$ time. To conclude, the total running time of this algorithm is $O(n)$, which is practically efficient for metropolitan cities.

3.4 Intersection Degree Distribution

3.4.1 Spatial degree distribution

We model each intersection as one node and each road segment as one edge. Then we get the graph $G(V, E)$ where V is the node set and E is the edge set. The n by n dimension matrix A indicates the existence of the edges when $A_{ij} = 1$ implies the

edge from node j to node i and $A_{ij} = 0$ does not. Hence, we obtain the in-degree d_{in} and out-degree d_{out} of each node (intersection) $i, 1 \leq i \leq n$ as follow:

$$d_{in}(i) = \sum_{j=1}^n A_{ij} \quad (3.3)$$

$$d_{out}(i) = \sum_{j=1}^n A_{ji} \quad (3.4)$$

And the degree distribution is the percentage distribution of node in the network that has degree d :

$$P_{in}(d) = \frac{|\{i \in [n] | d_{in}(i) = d\}|}{n} \quad (3.5)$$

$$P_{out}(d) = \frac{|\{i \in [n] | d_{out}(i) = d\}|}{n} \quad (3.6)$$

Degree distributions of the road network reveal both the road-segment interactions and also road hierarchy.

3.4.2 Spatial weighted degree distribution

In reality, long and wide road segments make the road junction that they intersect more critical in the city road networks. If an intersection is attached by a long road, the traffic condition in the intersection is influential for the road segment as all vehicles in the road segment can not leave this road segment without leaving or arriving at this intersection. In addition, wide roads with more lanes always imply that these roads are of high hierarchy or take large traffic volumes. Traditional road network modeling approaches do not distinguish these differences. Therefore, it is reasonable to incorporate the weight of edges (length and width) in the network analysis if we consider the discrepancy of road segment importance.

Previous researches provide us some insights. I.E.Antoniou et al. stated that in a weighted network, the weighted degree or stress of a node is defined as the sum of

weights of edges that have one endpoint as this node [32], which is also mentioned in the work by A.Barrat et al [33].

$$s_i = \sum_{j \in \Pi(i)} \omega_{ij} \quad (3.7)$$

Here, ω_{ij} is the weight of link from node i to j and $\Pi(i)$ represents the links attached to the node i . For our case, we improve the weighted degree model proposed by N. Wan [34] by redefining the contribution of the weight b_j :

$$S_i = \sum_{j=1}^{d_i} \frac{b_j}{b_{max}} \left(1 + \omega \frac{l_j - l_{min}}{l_{max} - l_{min}}\right) \quad (3.8)$$

Here, S_i is the defined spatially weighted degree, b_j represents the road width, and l is the length of the link. b_{max} is the maximal width of the road segment in the whole network, and l_{min}, l_{max} serve as the minimal and maximal road segment length, respectively. The introduction of parameter ω makes this model more flexible. A large value of ω means that the link weight is determined more on the link length than the link weight. Until now, we have defined the weighted degree of an intersection node in real road networks.

3.5 Data

This research relies on two kinds of data: intersection data, road segment data. We use the data collected by Zhan [10] of the UMNI Lab at Purdue University. This dataset contains the geo-location and average speed data of roads in Beijing and Shanghai in China.

The spatial ranges of the dataset for the two cities are: within Fourth Ring Road in Beijing and within Middle Ring Road in Shanghai. Using the measurement function in Google Maps, we know the total length of Fourth Ring Road and Middle Ring Road are respectively $65.03km$ and $69.97km$, and the total areas within Fourth Ring Road and Middle Ring Road are respectively $301.14km^2$ and $314.11km^2$.

The temporal ranges of the speed dataset are seven days for Beijing and six days for Shanghai. Evenly coverage of data within a week makes it possible for weekday and weekend comparison.

For Beijing, the seven days are Dec 6 (Sunday), Dec 7 (Monday), Dec 8 (Tuesday), Dec 10 (Thursday), Dec 11 (Friday), Dec 12 (Saturday), Dec 14 (Monday) in the year 2015. For Shanghai, the six days are July 7 (Thursday), July 9 (Saturday), July 10 (Sunday), July 11 (Monday), July 12 (Tuesday), July 13 (Wednesday) in the year 2016.

The longitudes and latitudes of the center points of intersections make up the intersection data. The road segment dataset contains the information for each road segment: length, ID, speed limit, road name, road width. And the speed data includes the average hourly link speed. There are 10,821 and 11,484 intersections in the research area of Beijing and Shanghai, and the number of road segments is respectively 17,147 and 18,173.

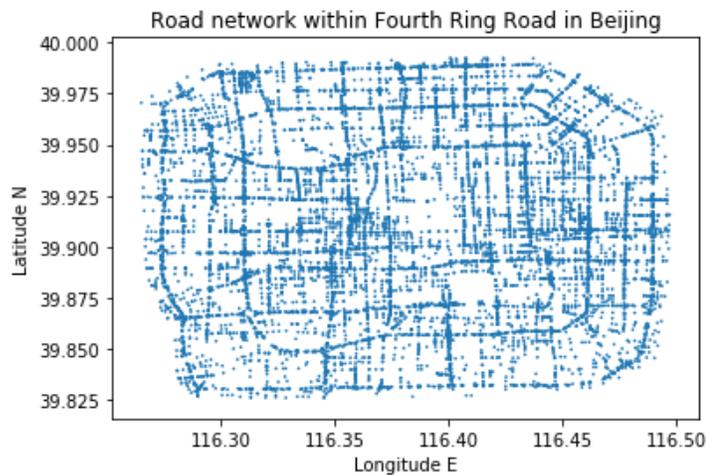


Figure 3.2.: Intersections in the research area in Beijing

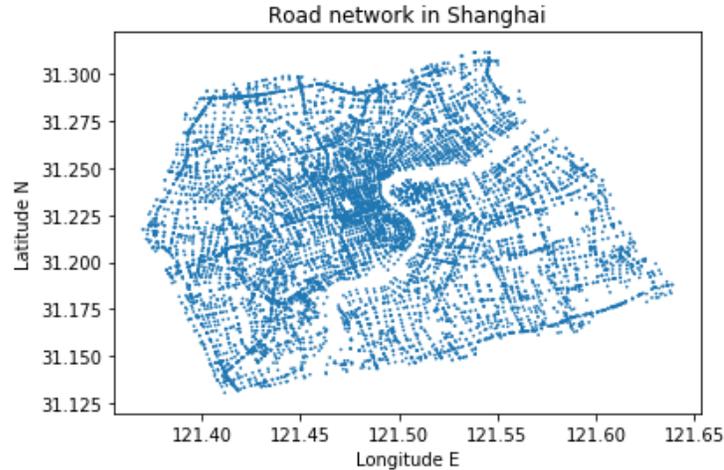


Figure 3.3.: Intersections in the research area in Shanghai

3.6 Node Merge Results

In this subsection, we use algorithm 1: node merge algorithm to merge the intersection nodes for Beijing and Shanghai.

Figure (3.4) is the road network of an area of about $1.5km*1.5km$ near Sanyuan-qiao overpass in the research area in Beijing. From Figure (3.4), we notice that in the dataset, some real intersection consists of several nodes. For instance, node 2210, 6570, 9566 (in the red cycle of Figure (3.4)) represent the same intersection in reality. And their in-coming and out-coming edges represent the road segments that intersect at this location. As our research objective is to model the real road network and further the congestion dynamics, it is reasonable to merge these three nodes together and regard them as one node.

As algorithm 1 indicates, our merging principle is intuitive and simple, every node pair that has a distance of less than δ will be merged into one node. As our researched dataset in Beijing has 10,821 nodes, pair-to-pair distance comparison takes a long time. To solve this problem, we apply algorithm 1. The core idea is to grid the map first, and then compare the distance of nodes from adjacent grids. Table (5.1) indicates, the value of $|E_\delta|$ (number of node pair that has distance shorter than δ),

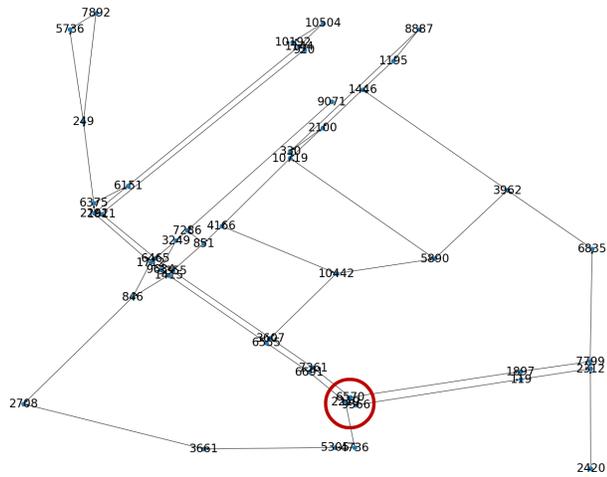


Figure 3.4.: A road network in Beijing before node merging

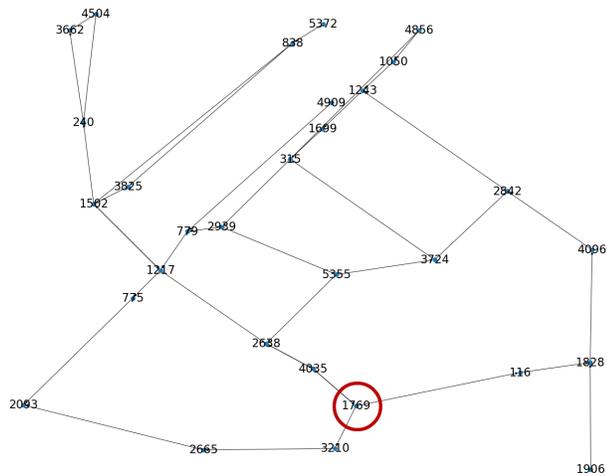


Figure 3.5.: A road network in Beijing after node merging

and the number of nodes after merging under different distance threshold δ . We can find that, as δ increases from 0.01 to $0.20km$, the size of E_δ increments from 830 to $45,785$, and the final node number after node merging drops from $10,017$ to 982 .

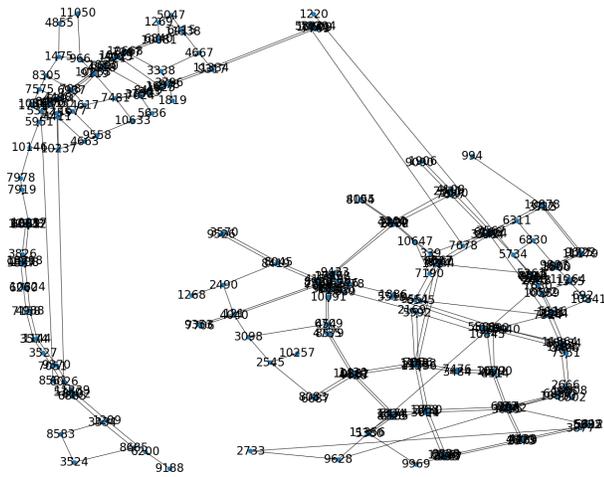


Figure 3.6.: A road network in Shanghai before node merging

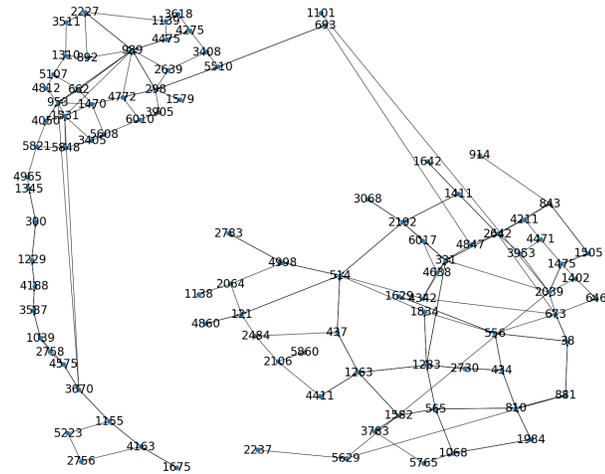


Figure 3.7.: A road network in Shanghai after node merging

In fact, main roads in metropolitan cities have more than 4 lanes and each lane is about 2.7-3.6 meters wide [35], so the size of an intersection is tens of meters. Even though the diameter for a large-scale overpass in metropolitan cities reaches to hundreds of meters, such enormous overpass consists of main roads and auxiliary

Table 3.1.: Node number under different distance threshold δ

$\delta(\mathbf{km})$	$ E_\delta $	final node number
0.01	830	10,017
0.02	3,631	7,897
0.05	9,555	5,457
0.10	19,411	3,253
0.20	45,785	982

roads as well as many ramp junctions. So it is superior to model gigantic overpass as several nodes instead of one node. Otherwise, over-simplification loses the features of the road network in megacities.

For the above two reasons, we adopt 50 meters as the distance threshold δ . When the threshold is $\delta = 0.05km$, the implementation of codes of line 8 to line 17 takes 10.54 seconds. And the whole node merging process takes less than 15 seconds, which is practically efficient. The computer runs on a Windows system and Intel(R) Core(TM) i7-8550U CPU @1.80GHz and 8.00 GB memory.

Figure (3.5) is the network representation after data preprocessing for the same area as Figure (3.4) in Beijing. There is one node (node 1769) in the red cycle, implying that original nodes that represent one same intersection have been merged together properly. Furthermore, along a road segment, the nodes that represent different intersections separate from each other clearly, indicating that our distance threshold is not as high as to merge two adjacent intersections together. And Figure (3.6) and (3.7) about Shanghai provide another evidence that nodes are merged properly using this algorithm. Therefore, our node merge algorithm is valid for this problem and $\delta = 50$ m is a proper parameter setting.

3.7 Degree Distribution Results

3.7.1 Degree distribution

After the node merging, we obtain the new directed graph $G(V, E)$ for Beijing. This graph has 5,457 nodes and 10,191 edges. Both the average in-degree and out-degree is $\frac{10,191}{5,457} = 1.868$ and the average total degree is 3.735. For Shanghai, the node and edge numbers are respectively 6,285 and 12,276. And the average total degree is $\frac{2*12,276}{6,285} = 3.906$. The statistics are shown in Table (3.2).

Table 3.2.: Node number changes for Beijing and Shanghai after node merging

City	Beijing	Shanghai
node	10,821	11,484
edge	17,147	18,173
Average degree	3.169	3.165
node (new)	5,457	6,285
edge (new)	10,191	12,276
Average degree(new)	3.735	3.906

Assume in the primitive network, there is a directed edge from node i to node j . And in the new network, these two nodes are replaced with node i_{new} and j_{new} , respectively. And the old edge (i, j) is replaced with the new edge (i_{new}, j_{new}) . For the special case when the node i and j are merged to be the same node $i_{new} = j_{new}$, we **do not** generate the self-loop for the node i_{new} . This is because the edge (i, j) is a very short road segment in reality or there is no real road segment between node i and node j at all. On the other hand, if there are multiple edges from node i_{new} to j_{new} , we will only maintain one edge, to avoid multiple edges. With the above operations, we obtain the out-degree, in-degree, and total degree distribution for the network.

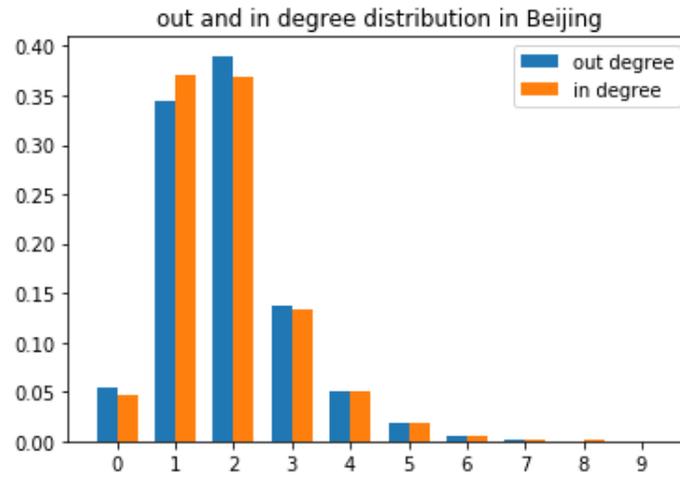


Figure 3.8.: Out and in degree distribution in Beijing

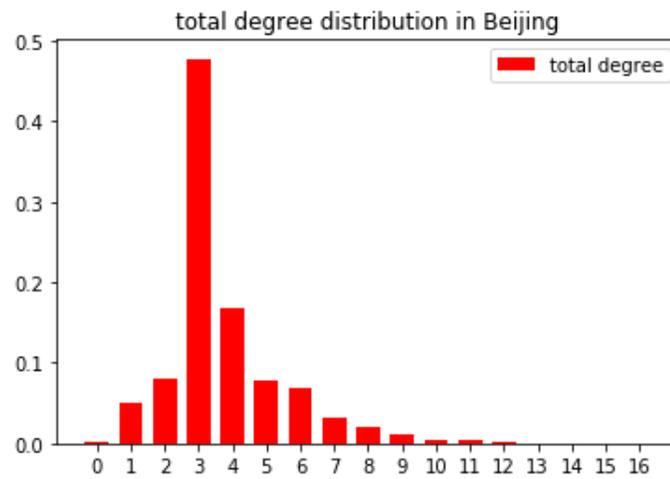


Figure 3.9.: Total degree distribution in Beijing

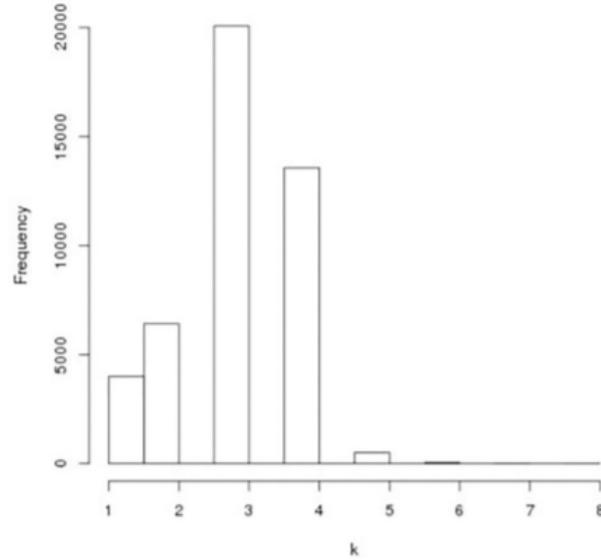


Figure 3.10.: Degree distribution in Beijing (within 6-th ring road) [36]

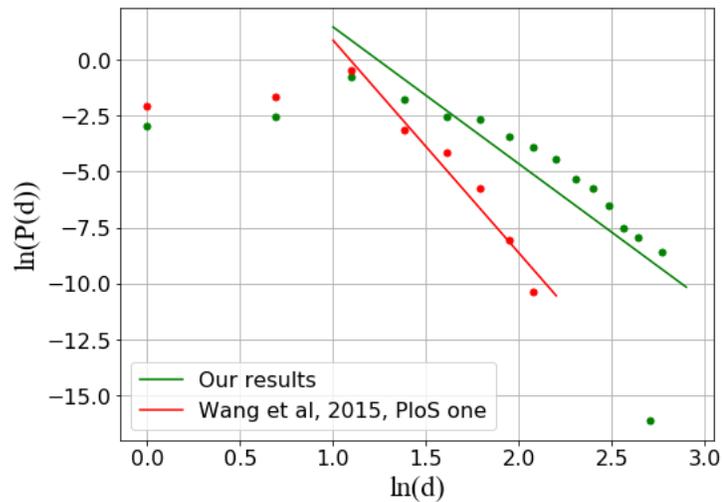


Figure 3.11.: Degree distribution comparison

We compare our results with the existing research by J. Wang et al. [36]. This research also focused on the road network in Beijing. And the researchers found that the average degree is 3.035, which is smaller than 3.735. The difference lies in the fact that their research area was larger than us (within the 6th ring road VS within

the 4th ring road). Figure (3.10) shows that the degree distribution in our results is similar to their results. In addition, Figure (3.11) indicates that if the small degree ($k = 1$ and $k = 2$) intersections are not considered, both the intersection degree follow power-law distribution. And the difference is the γ value.

For Shanghai, we also attain the out, in, and total degree distributions. The average degree is a little bit larger in Shanghai than Beijing (3.906 VS 3.735), which means that the intersections in Shanghai are attached by more road segments than Beijing in average. And an interesting result is that: Shanghai has a higher percentage of the intersection with the degree of 1. These one-degree intersections are the endpoints of the dead-end roads. It implies that roads in Shanghai extend more deeply into the residential and business areas. This result is related to the *last mile problem* which refers to the coverage of transportation facilities near the office or home [37]. Previous researchers used queuing theory and optimization methods to allocate transportation resources to meet the travel demand for the last mile trips. Our road network degree analysis quantifies the road resources near home or office and can be developed to guide the policy-making for last-mile transportation.

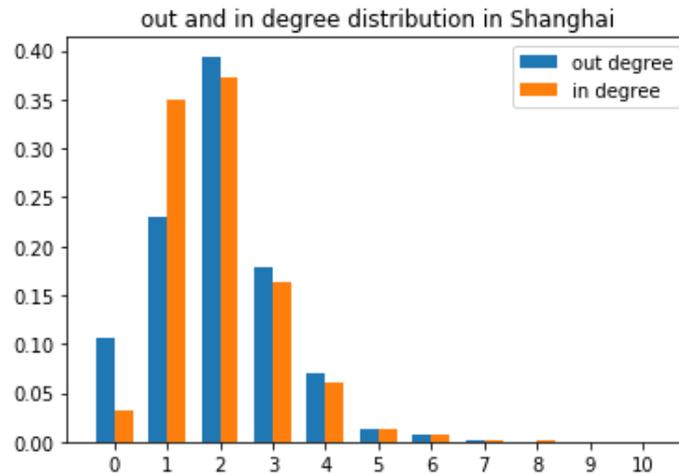


Figure 3.12.: Out and in degree distribution in Shanghai

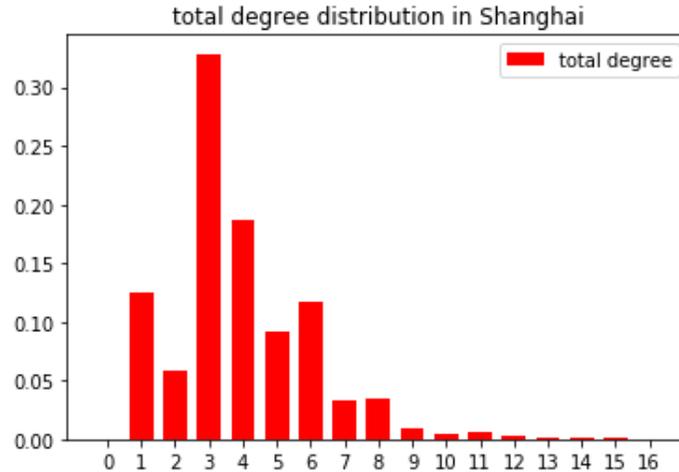


Figure 3.13.: Total degree distribution in Shanghai

3.7.2 Weighted degree distribution

The spatial weighted degree distributions are shown in Figure (3.14). **Note that the road segment width in our dataset is abnormal, we regard the width to be a constant for all road segments in this experiment. Road width information can be considered when accurate and valid road segment width data is available.** As the parameter w increases, the weighted degree distribution is more analogous to the power-law distribution. It means that if the importance of the intersection is defined based more on the road segment length than the road width, most intersections have small weights while there exists some *hub* intersections having large degrees. Former researchers found that under primal representation (intersection: node; road segment: link), the degree distribution of road network is NOT power-law. And according to Sergio Port et al. [9], the degree distribution of road network is power-law under the dual representation (intersection: link; road segment: node). At present, to the author’s knowledge, primal representation dominates road network researches and dual representation is not widely used. Our results offer a novel idea that the consideration of road segment length makes the degree distribution under primal representation to be power-law, which might be useful to discover

more properties of the road network, as power-law is prevalent in the world and has been deeply investigated by network researchers. If such a gap is filled properly, we are able to analyze the road network with existing scale-free network research results.

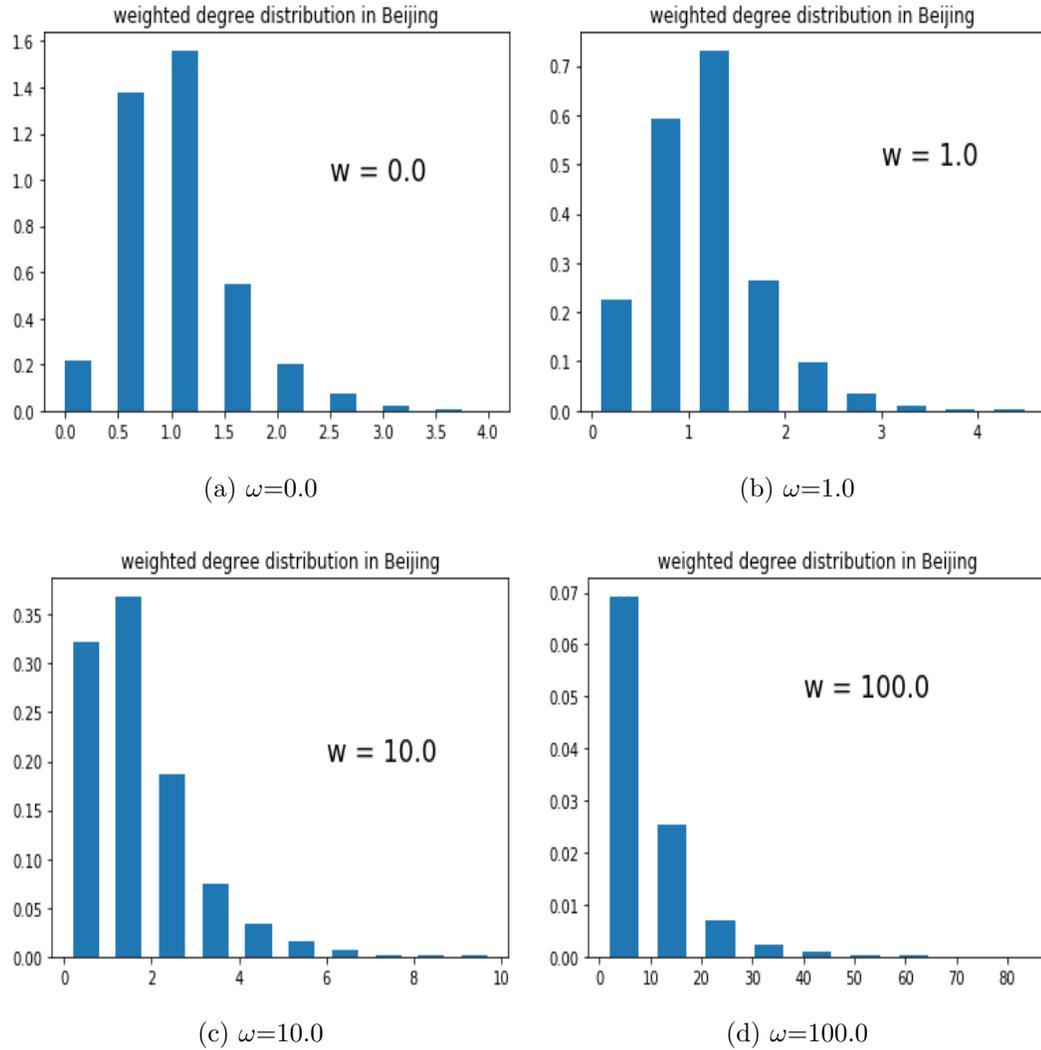


Figure 3.14.: Spatial degree distribution with varying ω

3.8 Conclusion

This chapter investigates the road network topology properties for metropolitan road networks. Different from other researchers, we emphasize the data preprocessing process by proposing an efficient node merge algorithm. The main idea of the algorithm is to grid the network and assign the nodes into the network, which helps to avoid the distance calculations for the nodes with large distances. Based on the node merge algorithm, the classic degree and weighted degree distribution models are built, revealing the topology characteristic of the road networks. The case studies on Beijing and Shanghai demonstrate the strength of the node merge algorithm and the application of degree analysis. And our results can guide the urban network modeling and further traffic congestion modeling and management. In the next sections, the author will work on the traffic congestion dynamics modeling on road networks using the speed data.

4. COMPLEX ROAD NETWORK DYNAMIC MODELS

4.1 Introduction

For cities, road traffic congestion management is a challenging task due to the complexity of travel demand and traffic state interactions. This task is composed of two sub-tasks: 1) adopt the proper method to model the traffic congestion dynamics; 2) propose an operational and valid traffic management plan. For the first task, the percolation-based congestion researches provide us with a concise and accurate network-level description of traffic states from the spatial link velocity data. For the second task, the Macroscopic fundamental diagram (MFD) serves as a potential for regional traffic management methods like gating, route recommendation, and congestion pricing. However, the percolation-based congestion is not a well-done theory and lacks the practical application. Besides, there are few quantitative and universal methods for road network partition, the first step of traffic management through regional traffic control. To fill this gap, this chapter proposes a complete congestion modeling and management methodology. Specifically speaking, we firstly test the correctness of percolation transition results for the road network. Further, we propose a network partition method which maximizes the network modularity based on percolation results. Experiments in Beijing show: 1) the critical drop of percolation curve exists for most regions; 2) our network partition model is able to offer reliable and realistic candidate network partitions.

4.1.1 Previous work

Average speed is a fundamental variable to describe the severity of urban traffic congestion. It is defined as the average or weighted average of the link speed

for all road segments in the city. Usually, during the morning peak (7-9 am) and evening peak (5-7 pm) on weekdays, the average speed is smaller than non-peak hours. Though easy and straightforward, the average speed does not provide the spatial speed distribution information in the city. Moreover, it is an aggregated variable and is not related to the road network structure in a specific region. For these reasons, new variables and merits need to be introduced.

Scholars have investigated road network congestion from different perspectives. In traffic assignment theory, if everyone is selfish and chooses the path with the least travel time, the overall traffic state satisfies the user equilibrium (UE) condition. Compared with UE, the system optimal (SO) traffic state regards the total travel time as the minimization objective and has less or equal total travel time in comparison to UE traffic state [38]. Accordingly, transportation researchers tried to alert personal route behavior toward the system optimal direction. Here, the total system travel time is one of the traffic congestion merits.

Besides, *flow* q (vehicle/hour), *density* k (vehicle/mile), and *speed* v (mile/hour) characterize the link traffic state. The fundamental graph clarifies the interaction between these three variables. When the density k increases from 0 to k_{max} , the link flow increases first and then declines. There is a critical point k_c where the flow reaches the maximum. It is assumed that the road is congested when the density k surpasses the k_c .

Some researchers quantified the urban congestion toward a network-scientist's point of view. Zhan et al. adopted the dual road network representation where each continuous road is denoted as a node and an intersection is modeled as an edge [10]. In their congestion model, if one road segment is congested (congestion state is determined by the condition that the relative speed is below a threshold), its affiliated continuous road *splits* into two parts. On the whole, the number of *splits* describes the congestion level. Additionally, in the research about the failure spreading from one node in the spatially embedded network, Zhao et al. used the

$r_c(t)$, average Euclidean radius of the failure region, and $F_r(t)$, the number of node failures to reflect the congestion level in the network [39].

4.1.2 Percolation phase transition

Percolation is a phenomenon that the liquid moves through porous materials. In a flow network, the percolation phase transition describes the network congestion dynamics. Researchers paid attention to the critical point where the network state changes rapidly. It was first researched by S. R. Broadbent in the year 1957 [40]. In recent years, the percolation phase transition was used to describe the packet routing in the complex network [29] [41]. The main feature of the percolation theory is that it quantitatively describes the connectivity of dynamics in the network. The term *percolation* here implies the phenomenon that when the flow in the network becomes large, the used links (whose flow is positive) constitute a large connected component.

In transportation engineering, scholars started to use percolation to study infrastructures and vehicles. For instance, Alireza Mostafizi et al. investigated the mobility effect of connected vehicles with various penetration rates and communication ranges [42]. In the context of percolation-based road congestion, Li et al. defined the critical relative speed q_c for regional road networks to measure the congestion level [19]. Different from the previous road congestion modeling approach, this novel method used the simple velocity data and discovered the general traffic congestion pattern. We would test the validity of the percolation phenomenon in the road network and our partition methodology is an extension of their work.

4.1.3 Macroscopic fundamental diagram

Flow, speed, and density are three major variables to describe traffic. In the year 2008, Nikolas Geroliminis et al. observed a clear relationship between the weighted average flow q^w and weighted average density (occupancy) k^w within one region based

on the traffic data from 500 loop detectors and 140 GPS-equipped taxis in Yokohama, Japan [43].

$$q^w = \frac{\sum_i q_i l_i}{\sum_i l_i}, k^w = \frac{\sum_i k_i l_i}{\sum_i l_i} \quad (4.1)$$

Here, q_i , k_i , and l_i are the link flow, density, and length, respectively. This relationship was named Macroscopic Fundamental Diagram (MFD) as it depicts the traffic state within a spatial area in a macroscopic way. Hereafter, researchers found that a well-defined MFD with little scatter existed when the regional traffic state was homogeneous [44]. The establish of MFD theory shed light on various traffic management methods, like perimeter gating [45], cordon pricing [46] and route guidance [47].

4.1.4 Road network partition

The first step to employ MFD to analyze the traffic state is to decide the areas. For simplicity, transportation engineers used postcode zones or traffic survey zones. But these divisions may not be sufficient to obtain a well-defined MFD due to lack of evidence regarding the relationship between these predefined areas and the traffic-homogeneous zones. To solve the problem, researchers proposed link density-based network partition methods [48] [49] [50] [51]. They modeled each road segment as one node, and defined the distance between node i and node j based on their link densities: if the density k_i and k_j differ a lot, then the node distance $d(i, j)$ would be large. Totally speaking, these density-based partition methods have the following drawbacks:

- 1) Use the density data which is not accessible for all road segments. Their network partition methods rely on the road segment density, which comes from loop detectors that are installed under the pavements. Loop detectors are expensive so that not every road is equipped with them [52]. People have to deal with the data-lacking problem for the remaining road segments. What's more, there is a severe loop detector failure problem. It was reported by the State of New York that 25% loop detectors did not operate at any given time [53].

2) Do not distinguish loop detector locations. Road flow data from Toulouse, a medium-sized French city revealed that when the distance between the loop detector and the junction was different, the citywide MFD curves were different [54]. In Figure 4.1, the points in one of the three colors denote the records collected by the loop detectors whose distances to the next junction are 0-200, 200-400, ≥ 400 meters. We do observe that the three curves have different peak points. But previous network partition methods do not consider the location of the loop detectors, bringing bias to the road segment density data used to generate MFD.

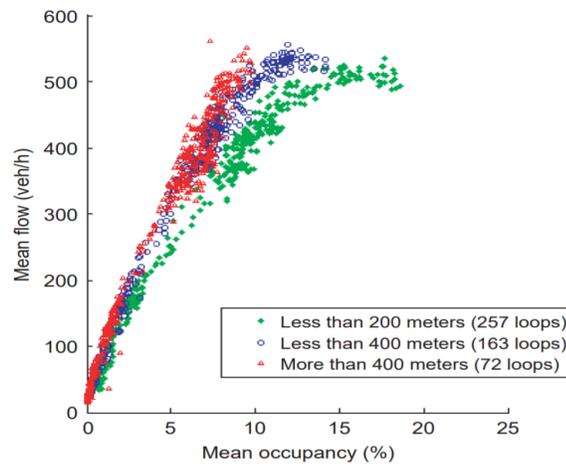


Figure 4.1.: Impact of distance between loop detector and intersection on MFD curve for Toulouse, France, on June 6, 2008 [54]

3) Do not consider the road hierarchy. The research on Toulouse also indicated that MFD curves for highways differ with downtown roads evidently as their maximal speed and road capacity are not the same. However, the aforementioned partition methods do not classify these distinctions.

4.2 Road Network Dynamic Analysis Framework

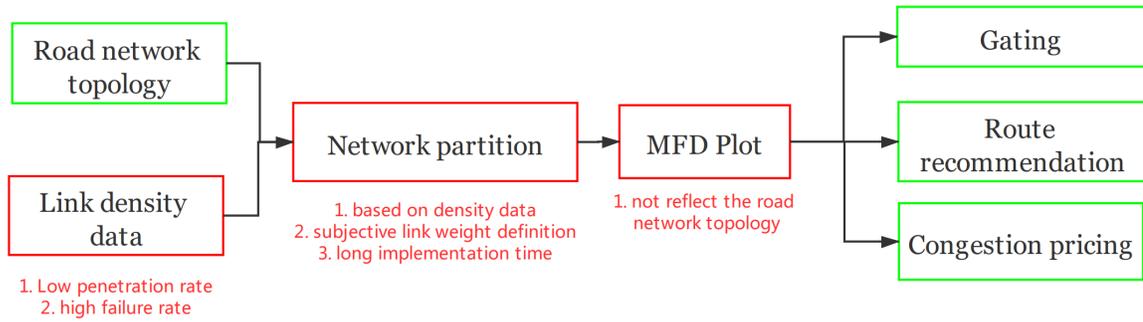


Figure 4.2.: Traditional MFD traffic management technical route

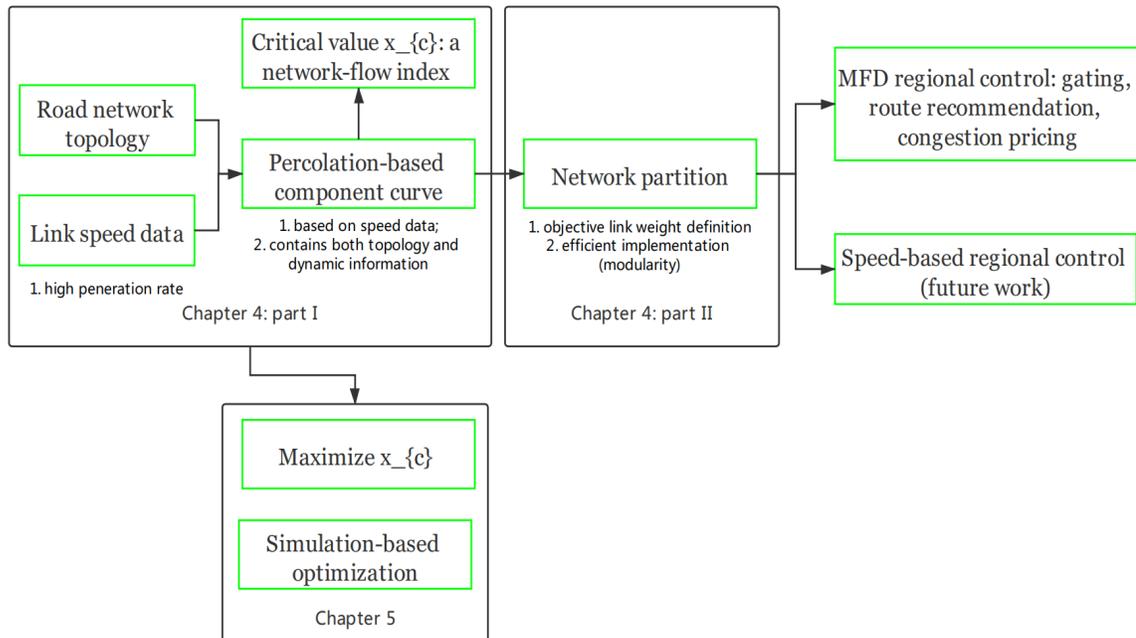


Figure 4.3.: A new traffic management technical route

Our overall procedure to alleviate the traffic congestion consists of two steps: 1) derive the traffic congestion patterns; 2) propose the traffic management strategies based on the traffic congestion patterns. From now, we start to present our traffic management technical route (see Figure 4.3). It has two main components:

1) percolation-based component curve generation and critical value x_c definition; 2) modularity-based network partition. Moreover, the x_c defined in this chapter serves as the objective function of chapter 5. Compared with the traditional technical route (see Figure 4.2), our methods have the following three advantages: 1) rely on the easily-accessed link speed data; 2) preserve both the topology and dynamic features of the road network; 3) the network partition process is efficient.

4.3 Percolation Curve And Critical Value

As mentioned above, the link speed data is an accessible data source reflecting the basic traffic states in metropolitan cities. Hence we try to derive the speed pattern through link speed data. We refer to the road network as $G = G(V, E)$ where V is the node set, and E is the edge set. Our model is a primal representation where the nodes are junctions and the edges are road segments. We classify all links in the road network into two categories: congested and uncongested. For each link $i \in E$, the relative speed r_i at time t is defined as:

$$r_i(t) = \frac{v_i(t)}{v_i^{limit}} \quad (4.2)$$

where the $v_i(t)$ is the average speed of all the vehicles on this road segment at time t , and v_i^{limit} is the speed limit for this link. We assume the road segment i is congested at time t if and only if

$$r_i(t) < \rho \quad (4.3)$$

and $\rho \in (0, 1)$ is a predefined link relative speed threshold. Conversely, the road segment is regarded as uncongested when the relative speed $r_i(t) \geq \rho$. Accordingly, we are able to define the uncongested link set at time t :

$$E_u(t) = \{i \in E | r_i(t) \geq \rho\} \quad (4.4)$$

And the uncongested subnetwork is

$$G_u(\rho) = G(V, E_u(\rho)) \quad (4.5)$$

For a vehicle traveling on the network $G(V, E)$, its trajectory makes up a subset $E_{vehicle} \subset E$. If $E_{vehicle} \subset E_u(t)$, it means that this vehicle travels on uncongested roads all the way and the travel time is not large. Otherwise, when $E_{vehicle} \cap (E - E_u(t)) \neq \Phi$, the vehicles would come across congested links, where the travel time may contribute a lot to the final total travel time. From the driver's perspective, bad traffic conditions and slowing-moving traffic flow in these congested road segments bring unpleasant feelings.

Strictly speaking, the vehicle arrives at different links at different moments, and it is better to profile the congestion level for the link that the vehicle is traveling on at the moment t and t is the real time. Generally speaking, the traffic state in urban cities change during the day. For a trip, if the travel distance is not large and the travel time is not long, we assume the traffic state keeps constant. And this assumption was widely used in the user equilibrium (UE) and system optimal (SO) traffic assignment theory in transportation engineering.

In network science, the notion of *component* is used to describe the integration of a subnetwork [55]. For a directed network $G(V, E)$, a component $G(V', E')$ is a collection of nodes such that for any pair of nodes in V' , there is a path in E' from one node to the other. For any node $u \in V$, we denote the largest component that contains u to be $Comp(u)$. It can be proven that $Comp(u)$ is the intersection of $ReachFrom(u)$ (the set of nodes that are reachable from u) and $ReachTo(u)$ (the set of nodes for which u is reachable).

$$Comp(u) = \{u\} \cup (ReachFrom(u) \cap ReachTo(u)) \quad (4.6)$$

And it is also true that if $v \in Comp(u)$, then $Comp(u) = Comp(v)$. Hence, when we get the $Comp(u)$ for each node $u \in V$, we can figure out the largest, second-largest components of the network, which serve as the indication of the total network performance since both the network topology and the average link velocity are already taken into consideration.

Moreover, the Breadth First Search algorithm (BFS) used in finding the largest component containing node $u \in V$ takes $O(n + m)$ time [56], where $n = |V|$ and

$m = |E|$, which is polynomial with regard to the network size. Even for metropolitan cities with tens of thousands nodes (intersections), the component discovery process would be efficient and is not the computation bottleneck.

4.4 Road Network Partition Algorithm

In this subsection, we would discuss the percolation-based network partition algorithm. We first show the basic idea behind this algorithm. For the road network $G(V, E)$ with the average speed on each road segment, the congestion curves from percolation components reflect the connectivity of good traffic condition roads. For a road segment $i \in E$, if the average speed of link i becomes larger, then with a high probability, the connectivity of good traffic condition roads would become stronger. We refer to this lift as the *connectivity benefit of segment i* toward the road network $G(V, E)$. When it comes to two adjacent roads i, j (that have an intersection in common), if the connectivity benefit of improving i, j is large, then we try to put roads i and j into the same traffic partition.

4.4.1 Modularity maximization and Louvain algorithm

Modularity maximization is the best known and most used network partition criterion [57]. For a network $G(V, E)$ with an undirected weighted adjacent matrix $A = (A_{ij})$. Weighted degree of node i is $k_i = \sum_{j=1} A_{ij}$ and S is a partition of the network G , the modularity is defined as

$$Q(G, S) = \frac{1}{2m} \sum_{s \in S} \sum_{i \in s} \sum_{j \in s} (A_{ij} - \frac{k_i k_j}{2m}) \quad (4.7)$$

Here, m is the sum of all edge weight in the network. In other word,

$$m = \frac{\sum_{i=1}^n k_i}{2} \quad (4.8)$$

A large modularity value $Q(G, S)$ indicates the connection within the community is more frequent than the random case. To achieve modularity maximization, researchers devised some approximate algorithms, among which the Louvain algorithm

was state-of-the-art [58]. Even for an Internet network with 70,000 nodes and 351,000 edges, Louvain algorithm is able to output network partition within only 1 second, which is less than CNM [59], PL [60], WT [61] algorithm and achieves the highest modularity of 0.781 among these four algorithms.

Louvain algorithm consists of two phases:

Phase I: the network is partitioned into several parts that no change of one node's membership yields a modularity score increase.

Phase II: each community is represented as a *super* node and the interactions of super nodes are defined.

Figure 4.4 presents two iterations in the Louvain algorithm.

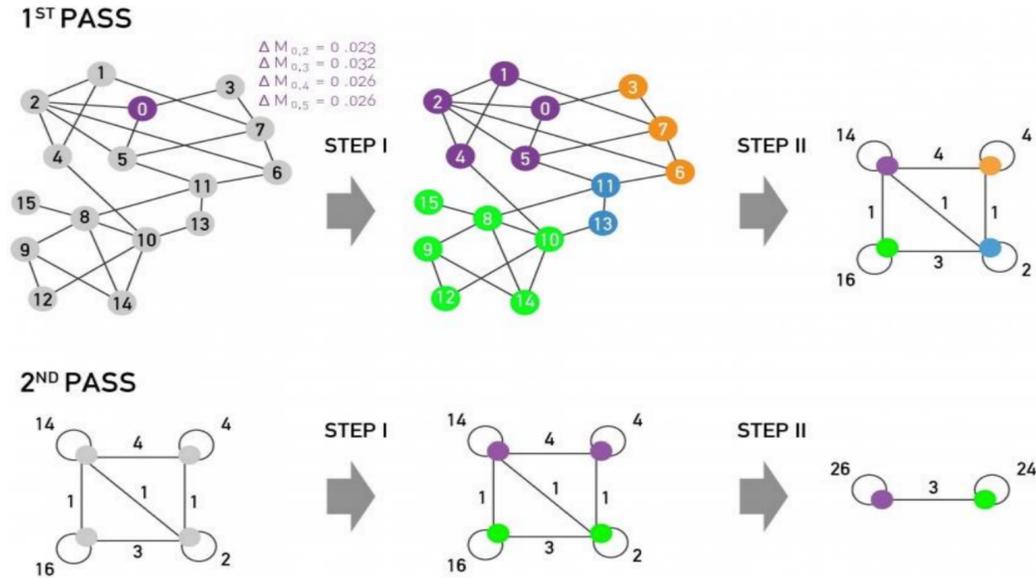


Figure 4.4.: Two iterations of Phase I (Step I) and Phase II (Step II) in Louvain algorithm [62]

4.4.2 Percolation-based road network partition

Now we start to devise a percolation-based road network partition algorithm extended from the Louvain algorithm. We first define the critical value x_c and edge weight A_{ij} for any two road segments in $G(V, E)$ as follows.

At time moment t , we have the road network $G(V, E)$ and the relative speed information for each link. To partite the network based on percolation information and the modularity objective, we build the road segment network $G^d(V_d, E_d)$ which is a dual representation (each node in G^d denotes a road segment, each edge in E_d denotes the adjacency relationship between the two road segments). Assume there are m road segments in the network, and all road segments constitute the set

$$V_d = \{v_1, v_2, \dots, v_m\} \quad (4.9)$$

For every two nodes $v_i, v_j (1 \leq i < j \leq m)$ in G_d , if road segment i and j do not intersect at the same junction, we define the edge weight to be

$$A_{ij} = A_{ji} = 0 \quad (4.10)$$

Otherwise, we use the congestion percolation to define the edge weight as follows.

Assume road segment v_i and v_j intersect at junction v . And the longitude and latitude of junction v is $Lo(i, j)$ and $La(i, j)$. Consider the square region centering at location v with the side length $2 \times r$ to be the region $\Gamma(i, j)$. Notice that we have the location information about all road segments and intersections, as well as relative speed for all road segments. In the following experiment section, we will show the largest component curve is a monotonically decreasing function.

$$L(x|i, j, r_i, r_j) : [0, 1] \rightarrow [0, 1] \quad (4.11)$$

$$x \mapsto L(x|i, j, r_i, r_j) = \frac{|lc(G_u(x))|}{|G_u(x)|} \quad (4.12)$$

Here, $lc(\cdot)$ and $|\cdot|$ return the largest connected component and the number of nodes in a network. And r_i and r_j is the relative speed of road segment i and j (see equation 4.2). We define the critical value $x_c(i, j, r_i, r_j)$ to be the relative speed whose slope is largest:

$$x_c(i, j, r_i, r_j) = \operatorname{argmax}_{x \in [0, 1]} \frac{-\partial L(x|i, j, r_i, r_j)}{\partial x} \quad (4.13)$$

The critical value $x_c(i, j, r_i, r_j)$ quantifies the congestion level of the neighborhood area of road segment v_i and v_j . We denote the weight of edge $i - j$ to be

$$A_{ij} = A_{ji} = x_c(i, j, r'_i, r'_j) - x_c(i, j, r_i, r_j) \quad (4.14)$$

where r'_i, r'_j represent the average speed after improvement.

We show the reasons for the aforementioned definition of A_{ij} . Firstly, if the road segment i, j are not adjacent, then the edge weight is set to be 0. It is because the two roads are not directly related. The second case is that road i and j share the same intersection. $x_c(i, j, r_i, r_j)$ represents the connectivity of good traffic condition road for the surrounding area $\Gamma(i, j)$. We showed that local connectivity of the road network is essential to enhance driving comfort and regarded the local connectivity maximization as our regional traffic control purpose. Therefore, a large value of A_{ij} means that the benefit of road connectivity for improving traffic conditions of road segment i and j simultaneously is significant, and it is recommended to put road i and j inside the same traffic management region.

Algorithm 2 Percolation based network partition algorithm

Input: A primal road network $G(V, E)$, $|V| = n$, $|E| = m$, a time moment t , relative velocity $r_i(t)$ for each link $i, i \in E$

Output: A dual weighted road network $G^d(V_d, E_d)$, $|V_d| = m$, $|E_d| = m^2$, a network partition S of $G^d(V_d, E_d)$

- 1: neighbor edge set $N \leftarrow \Phi$
 - 2: edge weight matrix $A_{m \times m}$
 - 3: **for** $v \in V$ **do**
 - 4: find all roads that have an endpoint as v :
 - 5: $E_v := \{e \in E | v \text{ is an endpoint of } e\}$
 - 6: **for** $i, j \in E_v, i \leq j$ **do**
 - 7: $N = N \cup \{(i, j)\}$
 - 8: **if** $r_i(t)(\text{or } r_j(t)) < 1.0$ **then**
 - 9: $r'_i(t)(\text{or } r'_j(t)) = 1.0$
 - 10: **else**
 - 11: $r'_i(t)(\text{or } r'_j(t)) = r_i(t)(\text{or } r_j(t))$
 - 12: **end if**
 - 13: calculate the edge weight between road i and j
 - 14: $A_{ij} = A_{ji} = x_c(i, j, r'_i, r'_j, r) - x_c(i, j, r_i, r_j, r)$
 - 15: **end for**
 - 16: **end for**
 - 17: **for** $(i, j) \in E \times E$ and $i \leq j$ **do**
 - 18: **if** $(i, j) \notin N$ **then**
 - 19: $A_{ij} = A_{ij} = 0.0$
 - 20: **end if**
 - 21: **end for**
 - 22: construct the dual road network $G^d(V_d, E_d)$: $V_d = E, E_d = A$
 - 23: implement the **Louvain algorithm** on $G^d(V_d, E_d)$ and get the k -partition:
 - 24: $S = V_d^1 \cup V_d^2 \cup \dots \cup V_d^k$
 - 25: **return** partition result S
-

4.5 Data

We use the same data of Beijing as in chapter 3. For Beijing, the speed dataset covers 5 weekdays, and 2 weekends. Figure 4.5 shows the average speed for all road segments in the network for weekdays and weekends. We do observe that the morning and afternoon traffic peak happen between 7 to 9 am, 5 to 7 pm respectively. Moreover, compared with weekdays, the morning and afternoon peak during the weekends is not significant. These results agree with the common daily speed variation within one day, which confirms the reliability of our data source.

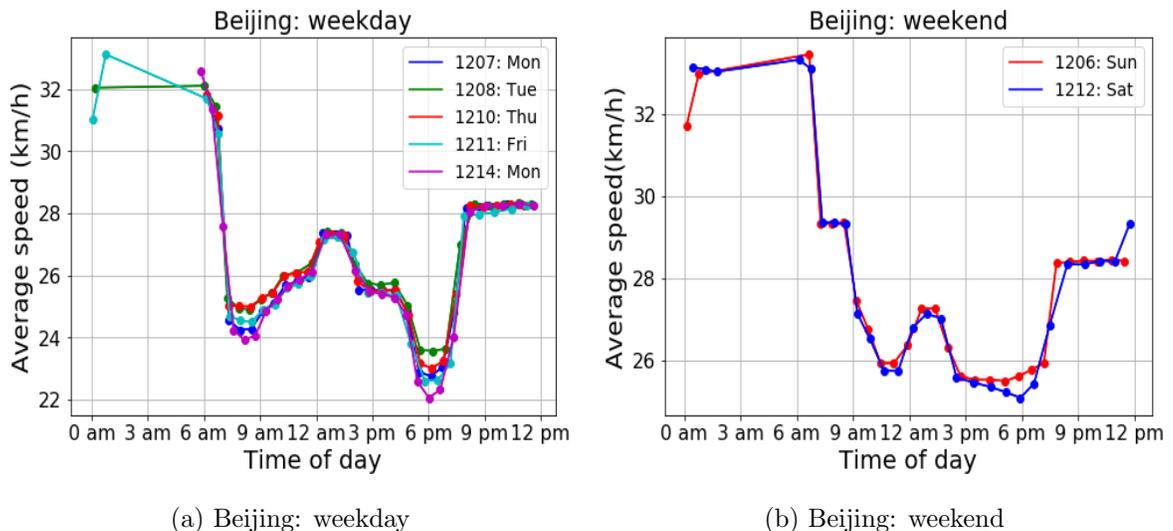


Figure 4.5.: Temporal speed variation

4.6 Percolation Curves And Critical Value Results

4.6.1 Existence

In Figure 4.6, we plot the congested (red) and uncongested (green) links for Beijing. The time was 11:46, December 8, 2015 (Tuesday). And the relative speed threshold is set to be 0.58. We find that traffic congestion occurs among the road network homogeneously.

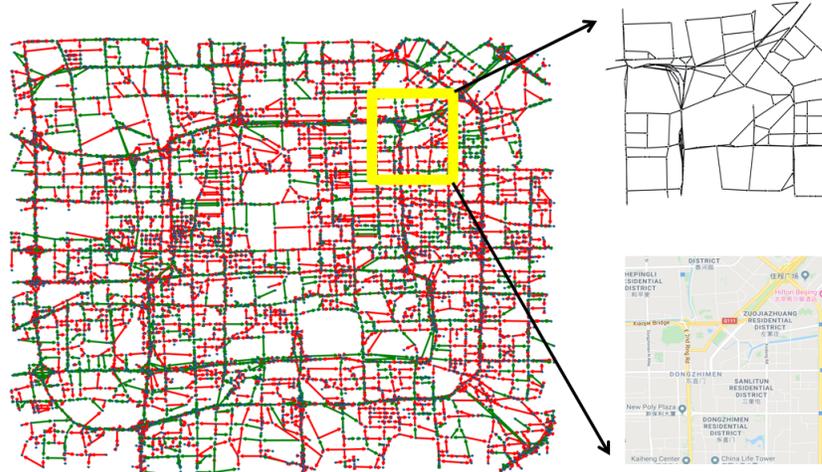


Figure 4.6.: Congested roads in Beijing at 11:46, December 8, 2015 (Tuesday)

Inside the yellow box we display the road network in a region about $3km * 3km$ near Dongzhimen-Sanyuanqiao which is a transportation hub in northeastern Beijing. Hereafter, we consider the area inside the yellow box at that moment with varying relative speed threshold $\rho = 0.30, 0.58, 0.85$. From Figure 4.7 to 4.9, the left sub-figures display the traffic state for each link (the red color represents the congested road segment and the green color represents the uncongested road segment). When the relative velocity threshold is set to be $\rho = 0.30$, we do observe that most road segments within the research area are colored green, making up a huge component. For most trips with the origin and destination located in this area, the driver is able to arrive at the destination with the speed of not less than 30% of the speed limit.

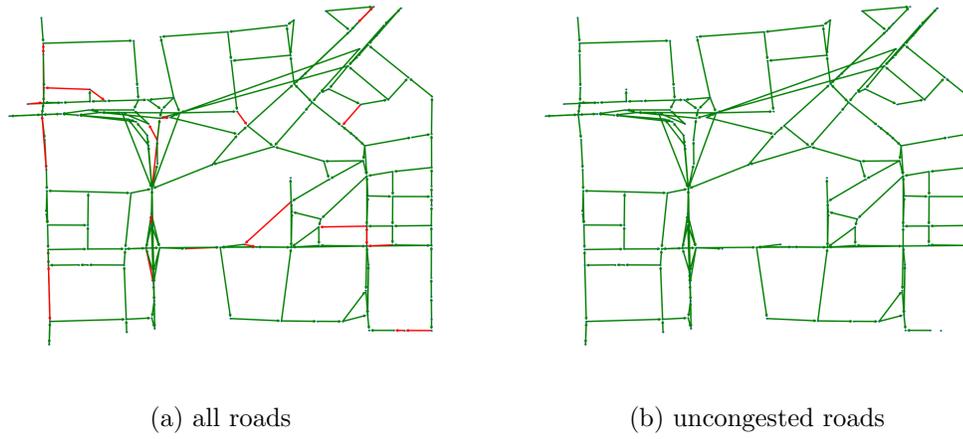


Figure 4.7.: Traffic state of roads with $\rho = 0.30$

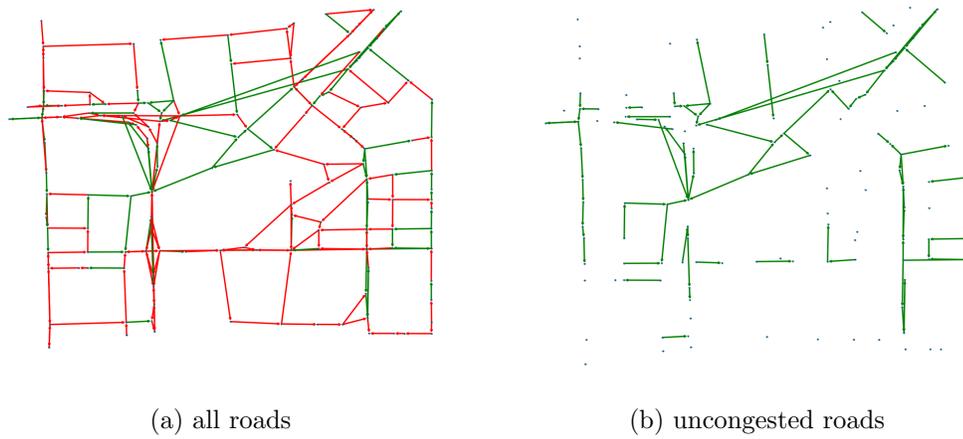


Figure 4.8.: Traffic state of roads with $\rho = 0.58$

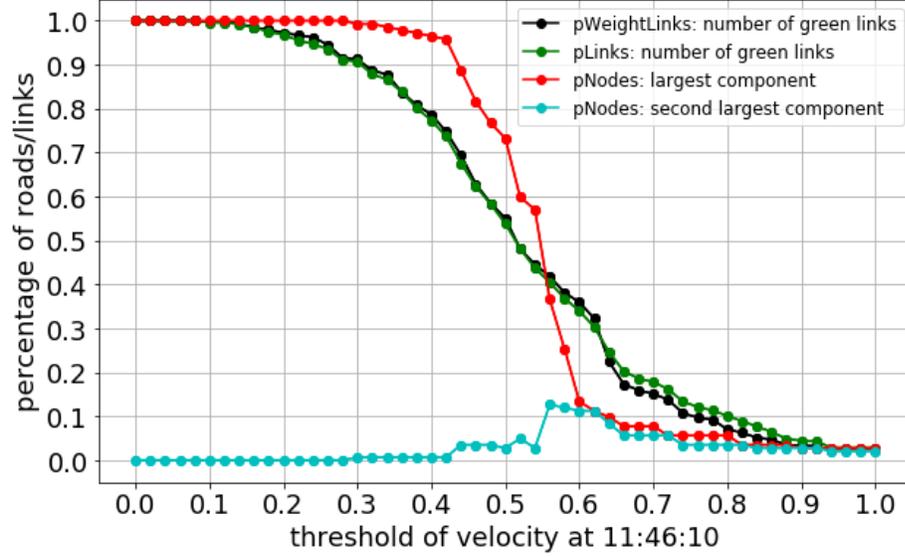


Figure 4.10.: Percolation curve in Sanyuanqiao at 11:46, December 8, 2015 (Tuesday)

Then, we plot the size of the largest and the second-largest components for this region with various thresholds in Figure 4.10. In the Figure, $pLinks$ indicates the percentage of the number of green links within the research area,

$$pLinks = \frac{|E_\rho|}{|E|} \quad (4.15)$$

Different from $pLinks$, $pWeightLinks$ embodies the length of the link,

$$pWeightLinks = \frac{\sum_{i \in E_\rho} length(i)}{\sum_{i \in E} length(i)} \quad (4.16)$$

Here, E_ρ represents links with the relative speed greater than or equal to the given threshold ρ , and E denotes all links in this research area, $length(i)$ is the length of road segment i . More importantly, the $pNodes$ curves (largest component and second-largest component) represent the ratio of one certain type of nodes (largest-component, second-largest-component).

In Figure 4.10, as the relative velocity threshold ρ increases from 0.0 to 1.0, the $pLinks$ and the $pWeightLinks$ curves decline monotonically, and the two curves have steep slopes when ρ is between $[0.3, 0.8]$, which indicates that most of the roads have a relative speed in this interval. In addition, the two curves are close to each other,

demonstrating that the length distribution for the links with one certain speed threshold ρ approximates the length distribution for all links.

More interestingly, the $pNodes$ curve (largest component) drops rapidly when the ρ is between 0.4 and 0.6, where the largest component of green edges split into several sub-components. And the second-largest component reaches its maximum at about $\rho = 0.58$, which matches the rapid drop of the largest component curve. Here, we denote the $\rho = 0.58$ as the critical relative speed x_c . And the critical value x_c is used to describe the overall road network and traffic state features. The larger the x_c , the better the traffic state.

4.6.2 More spatial and temporal experiments

To verify the existence of the critical percolation phenomenon for traffic congestion, we tested different time moments and spatial regions.

Figure 4.11, 4.10, 4.12 relate to the percolation curves at three moments: early morning (06:02), noon (11:46), and evening (18:14) on December 8, 2015. We notice that for the three cases, the largest component shrinks rapidly at one critical velocity threshold. For the second-largest component, there exists a velocity threshold where the second-largest component reaches the largest size. The rapid drop points of the largest components are quite close to the maximum points of the second-largest components.

Besides, the critical relative velocity represents the regional traffic state. From Figure 4.11, 4.10, 4.12, we recognize that the critical value for the morning, noon, and evening are respectively 0.60, 0.58, 0.44. The critical velocity threshold indicates that: 1) the road segments whose relative velocity near the critical velocity determine the road connectivity greatly; 2) The critical relative speed is located at the medium of the relative speed distribution for all links. Figure 4.5 illustrates that for weekdays of Beijing, the traffic state in 06:02 is better than 11:46, and 11:46 is better than 18:14. And this relationship maintains for critical percolation point: 0.60

(06:02), 0.58 (11:46), 0.44 (18:14). This consistency demonstrates that the critical value takes on the role of average speed to reflect the overall traffic state. Moreover, critical percolation points contain the network topology information as we consider the component of the road network.

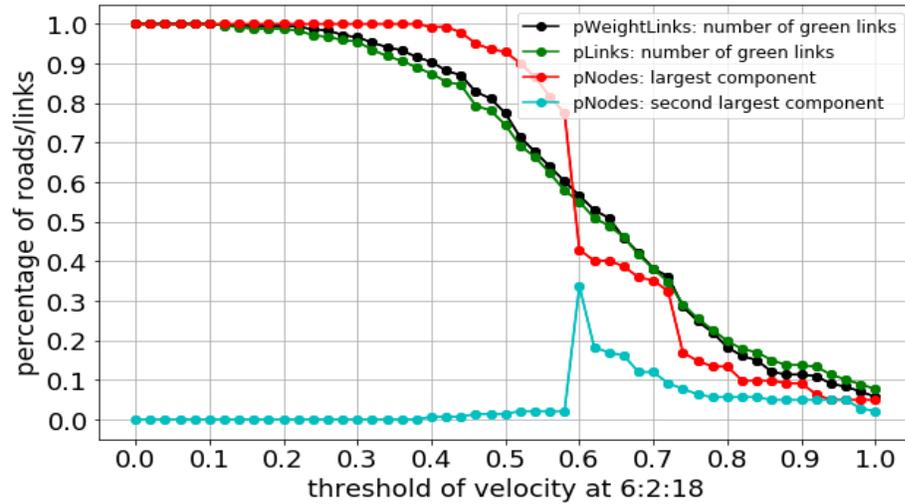


Figure 4.11.: Percolation curve in Sanyuanqiao at 06:02, December 8, 2015 (Tuesday)

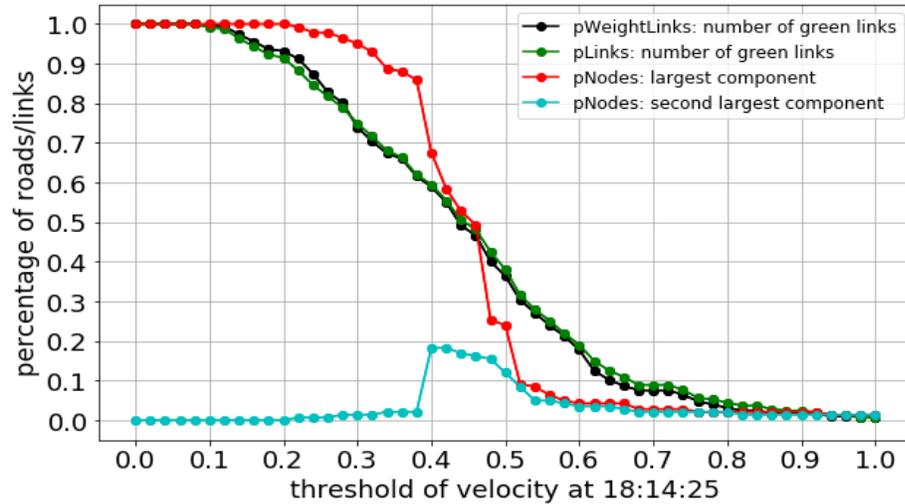
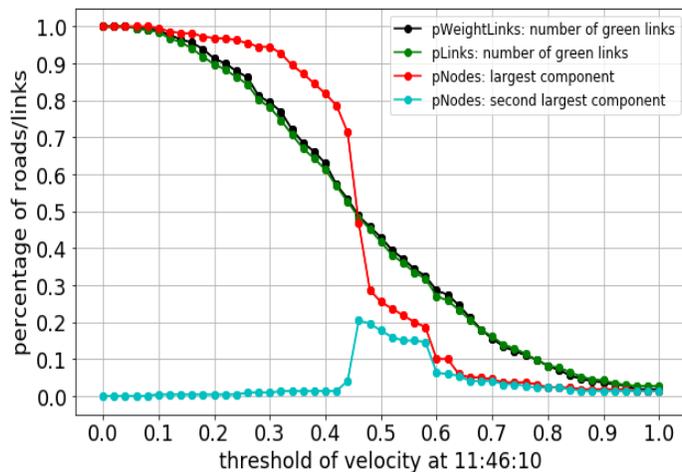


Figure 4.12.: Percolation curve in Sanyuanqiao at 18:14, December 8, 2015 (Tuesday)

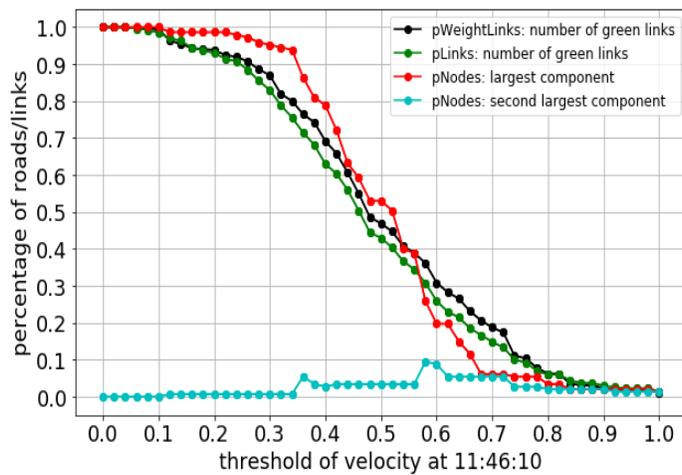
Other than the Sanyuanqiao area, we select two other areas in Beijing: Xidan (see Figure 4.13) and Zhongguancun (see Figure 4.14). The time was 11:46, December 8, 2015. Xidan is a commercial area located within the Second Ring Road in Beijing and Zhongguancun is a technology hub in northwestern Beijing, between the Third Ring Road and Fourth Ring Road. For Xidan (Figure 4.13), the critical relative speed is quite clear (about 0.48). And the curve of $pLinks$ and $pWeightLinks$ are quite close to each other, which indicates that long road segments and short road segments are homogeneous at different congestion levels.

When it comes to the Zhongguancun area (see Figure 4.14), the critical point is not obvious and the largest-component curve decreases nearly linearly from relative speed 0.35 to 0.65. This unique phenomenon indicates that road segments with the relative speed between 0.35 to 0.65 have similar marginal contribution of road connectivity. What's more, different from the other cases in Sanyuanqiao and Xidan, there is an evident gap between $pLinks$ and $pWeightLinks$ curve. In fact, the road hierarchy in the Zhongguancun area is more heterogeneous. Large roads like Fourth Ring Road, Third Ring Road, as well as small roads within the central Zhongguancun area, co-exist in this area. This result reminds us that homogeneous road type may be a pre-requirement to obtain the well-defined percolation curve with a unique critical point. And the relationship between the road hierarchy and existence of abrupt descent in component curves can be a future research.

(a) $\rho=0.50$ 

(b) component curve

Figure 4.13.: Percolation curve in Xidan at 11:46, December 8, 2015 (Tuesday)

(a) $\rho=0.50$ 

(b) component curve

Figure 4.14.: Percolation curve in Zhongguancun at 11:46, December 8, 2015 (Tuesday)

4.7 Network Partition Results

In this section, we implement our network partition algorithm for a $4.8\text{km} \times 7.5\text{km}$ region in Beijing (see Figure (4.15)). This road network has 714 intersections and 1,340 links.

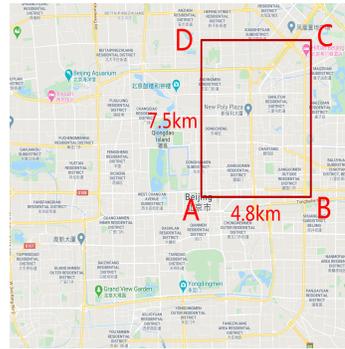


Figure 4.15.: Selected region for network partition

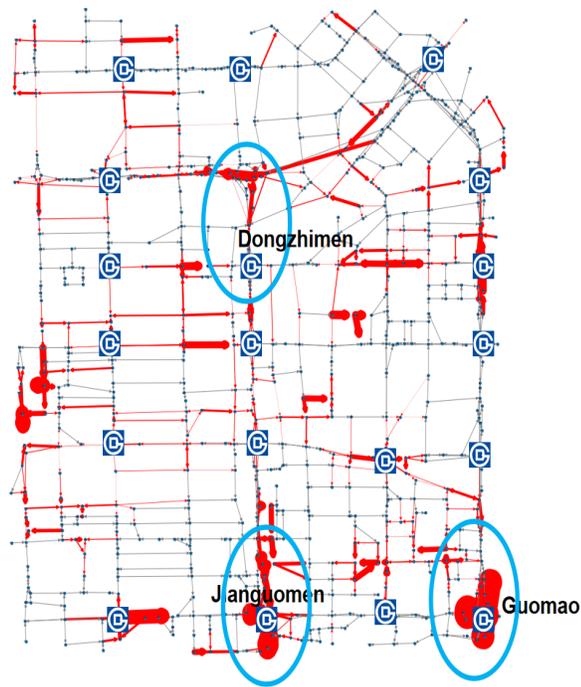


Figure 4.16.: Weighted degree in the dual network $G^d(V_d, E_d)$ in Chaoyang District in Beijing at 11:46, December 8, 2015 (Tuesday)

In algorithm (2), we obtain the dual weighted road network $G^d(V_d, E_d)$ where the node represents the road segment and the edge represents the road segment relationship. Note that this dual road network has weight on edges (road segment relations) but not on nodes (road segments). To better visualize the spatial weight distribution on road segments, we define the weight of node (road segment) $i \in V_d$ as the sum of weights for edge relations that include node i :

$$A_i = \sum_{j \in V_d} A_{ij} \quad (4.17)$$

By this definition, a high value of A_i suggests that road segment i has significant contribution of connectivity of good traffic condition roads when working together with neighborhood road segments. We plot the weight for each node (road segment) in Figure (4.16).

In Figure (4.16), the width of red arrow is proportional of A_i value. In Figure (4.16), we also visualize 19 subway stations on the map. We are aware of some relationships between wide red arrows and transportation hubs. The three largest clusterings of red arrows are at Jianguomen, Guomao, and Dongzhimen, respectively. They are exactly at or near the transportation hubs (subway stations). It informs us that the road segments near the transportation hubs are quite crucial to maintain good connectivity of good traffic condition roads. However, the road segments near some subway stations seem to not show an obvious difference from the roads that are far from subway stations. It is still an open question that whether it is possible to use the subway station location to detect or distinguish the *critical roads* for road connectivity improvement.

In addition, we visualize in Figure (4.17) green edges to represent, the edge relations whose A_{ij} are positive, which means that road segment i and j have a positive benefit toward the connectivity of the surrounding $2km \times 2km$ region. It indicates that there are dense green edges in Jianguomen, Guomao and Dongzhimen regions, which agree with Figure (4.16). Based on our previous explanations, a pair of road

segments i and j with a large value of A_{ij} need to be put into the same transportation management region.

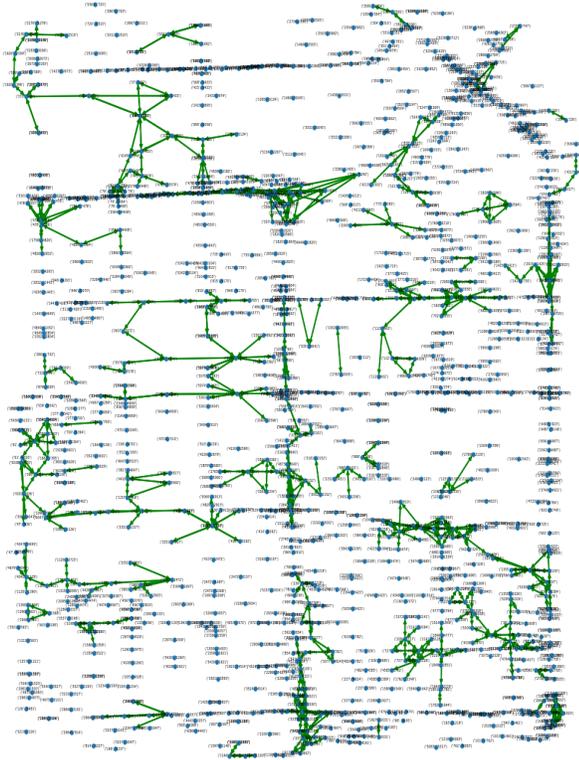


Figure 4.17.: Edges in $G^d(V_d, E_d)$ in Chaoyang District in Beijing at 11:46, December 8, 2015 (Tuesday)

Now we present the network partition results. In realistic road networks, it is common that we put the roads that have short distances in the same partition. After considering the distance effect, we change edge weight in $G^d(V^d, E^d)$ to be

$$A_{ij} \leftarrow \frac{1 + cA_{ij}}{1 + d_{ij}^2}, c > 0 \quad (4.18)$$

where d_{ij} is the distance between the midpoints of road segment i and j , and A_{ij} on the right side is the edge weight we have defined. c is a scalar parameter. We try

different parameters of c and get the network partition with the modularity values shown in Table (4.1). After comparison of the network partition results, we find that the partitions for $c = 1, 5, 10, 50, 100$ are analogous and their partition numbers are all 4. When $c = 500$ and $1,000$, the partition numbers are large.

Table 4.1.: Modularity results

c	1	5	10	50	100	500	1,000
modularity	0.300	0.301	0.287	0.298	0.310	0.351	0.396

Figure (4.18) shows two representative partition plans: $c = 5$ and $c = 1,000$. We notice that when $c = 5$, the three transportation hubs Jianguomen, Guomao, and Dongzhimen are located at respectively red, blue, and green communities. At $c = 1,000$ case, there are totally 9 partitions for these regions. An interesting finding is that Jianguomen and Guomao make up the small yellow, and the black community, respectively. But the Dongzhimen area belongs to the large green community (which is at the northwestern side of the research area). This partition highlights the critical region in transportation management as shown in Figure (4.16) and (4.17). When it comes to real regional transportation management, we recommend to take the $c = 4$ and $c = 1,000$ plan both as the candidate partition plans, even if $c = 1,000$ plan achieves a higher modularity (0.396 vs 0.301). In fact, there is a difference between complex road networks and general complex networks. Road network is likely to be homogeneous in space, making it hard to reach a high modularity score. And in the real application, transportation engineers prefer to define the partition regions of a similar size. Consequently, both of the plans can be taken and more transportation evidence will be used to support further decisions.

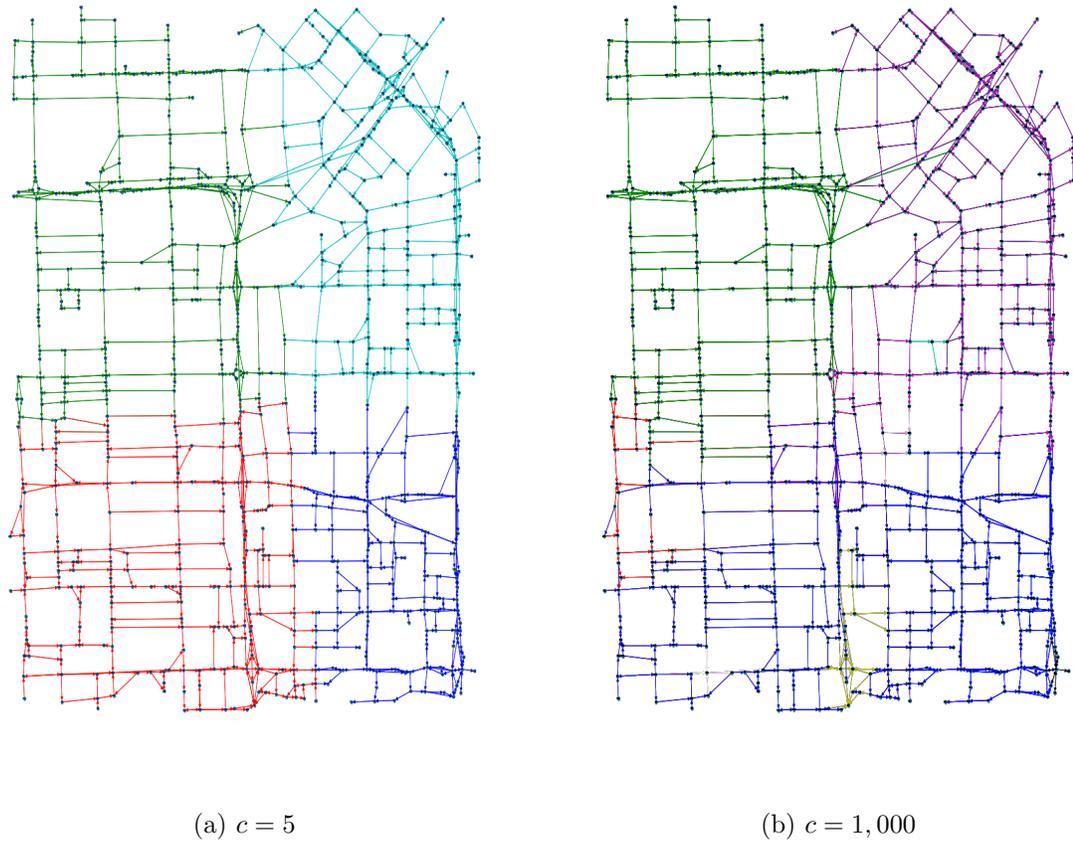


Figure 4.18.: Road network partition results

4.8 Conclusion

In this chapter, we develop a novel regional traffic analysis method using link speed and network topology data. We use the definition of critical value for a road network initially proposed by Li et al. [19]. It is defined as the state where the largest connected component of good traffic roads splits. Furthermore, the percolation-based network partition extended from the Louvain algorithm is proposed. Our results verify that the critical value of the percolation curve exists in the Beijing road network. Furthermore, our partition algorithm generates reliable candidate road network partition results, which helps transportation engineers to better manage the regional traffic.

Three future extensions are: 1) develop analytical models to describe the percolation phenomenon for congested road networks; 2) investigate the iterations between critical value and road hierarchy; 3) extend the partition algorithm to incorporate time information.

5. COMPLEX ROAD NETWORK CONNECTIVITY OPTIMIZATION

5.1 Introduction

Road network optimization, like link improvement, has been a prevalent and applicable transportation engineering research topic. In general, the optimization objective can be total system travel time, total walking distance, et al. And the budget upper bound is one kind of constraint. In the context of metropolitan cities, bottleneck road segments whose average velocities are smaller than the surrounding roads not only decline the road system performance but also boost drivers' road rage. Consequently, the local connectivity of roads with good traffic conditions can also be an alternative optimization objective as a high level of connectivity implies less occurrence of bottleneck road segments. Based on the novel congestion modeling using percolation theory, this chapter formulates the road network improvement problem as a nonlinear simulation-based optimization. The objective is to maximize the local connectivity of good traffic condition roads. And we set the investment budget as one of the constraints. Afterward, we apply the genetic algorithm in Matlab Optimization Toolbox to obtain meta-heuristic solutions and compare them with the simple transportation resource allocation method. Numerical results of different time and regions reveal that our methodology is able to provide the traffic agency with a good road network improvement strategy.

5.2 Background

5.2.1 Road management and maintenance policy

There are multiple road segment management and maintenance policies: 1) Repair the unclear traffic lines caused by rain, snow, tire friction, et al. 2) Monitor reckless road passing behaviors by pedestrians. 3) Assign parking tickets for parking regulation violations. If obsolete traffic marks are replaced by new ones, drivers do not hesitate to conjecture the original traffic signs anymore. If all pedestrians pass roads in a good manner, the drivers do not distract their attention due to any unpredictable surroundings. If the occurrence of illegal parking reduces, then the available space for moving vehicles becomes broad. To conclude, these traffic management policies actually increase the average vehicle speed and improve the efficiency of road utilization.

When it comes to realistic road network improvement, one of the foremost issues to deal with is the traffic management resources allocation problem. The traffic management resources here imply the transportation materials, working labors, administration forces. In reality, the transportation agency has a limited budget and how to use this budget efficiently is their concern. An intuitive idea is to allocate the traffic budget more to the congested region than the uncongested region. Specifically speaking, if the average speed for one road segment is below a threshold, the transportation agency would dispatch management resources to this road. But this simple resource allocation strategy has the following imperfection: it may not achieve a global optimum. It is probable that even road a is more congested than road b , but the marginal benefit of road improvement for b is greater than a . Hence, it is needed to propose an optimal allocation strategy.

5.2.2 Equilibrium-based network optimization

There are plenty of researches focusing on road network optimization problem [63], [64], [65], [66]. Researchers built a bi-level optimization problem where the upper level minimizes the system travel time and the lower level describes the equilibrium of the traffic flow. In their optimization, the budget serves as one constraint. And the investment on one certain road enhances road performance and system performance. However, there are three shortcomings for these equilibrium-based optimizations: (1) User equilibrium (UE) may not accurately describe passengers' actual route choice behavior. Under the UE condition, all selected path has minimal and equal travel time. This assumption may not be true because people may choose sub-optimal routes [67] or familiar routes. Moreover, UE does not model wandering travels, like unoccupied taxi trips. (2) The bi-level optimization problem is NP-hard, which lacks an efficient algorithm with polynomial running time. (3) The objective of the bi-level optimization is the system travel time. When it comes to regional road improvement, the method does not work because there might be some trip with the origin within the research area and the destination outside the area.

5.2.3 General network optimization

There are several researches about routing and robustness improvement in the complex network [41], [68]. The authors developed traffic control methods to enlarge the free-flow phase. In reality, the network here implies the information, communication, and wireless network, where the packets travel from the origin to the destination. And the network congestion level is measured by the change of packets in the network:

$$\rho(R) = \lim_{t \rightarrow \infty} \frac{N(t + \tau) - N(t)}{\tau R} \quad (5.1)$$

where τ is the time interval, R is the packet generation rate, and $N(t)$ is the number of packets in the network. For the road network, change of vehicle number is seldom used to describe the road system performance. Additionally, in the communication

network, each node has a probability r_i , which depends on the bandwidth, to transfer the packet to one of its neighborhood nodes. For road network, the link travel time is monotonically increasing with regard to the link flow and is modeled by the prevalent US Bureau of Public Roads function [69]. Therefore, previous complex network optimization can not be applied to the road network directly.

For the transportation network, David Aldous et al. considered the optimal shape of the network when the length L is given, with the objective as the travel distance to the network [70]. They had some analysis for the star and ring networks. But this model is more suitable for subway or metro line design, other than the road network.

5.2.4 Critical value

In Chapter 4, we model the real traffic state using the percolation theory. At a certain time and region, using the real average speed $v_i(t)$ and the speed limit v_i^{limit} data for one road segment i , we obtain the largest-component curve for the good traffic condition roads whose relative average speed is larger or equal to a predefined threshold ρ :

$$\frac{v_i(t)}{v_i^{limit}} \geq \rho \quad (5.2)$$

We recognize the critical value x_c , which is the x -coordinate of the point on the curve that has the steepest descent in Figure 5.1. x_c contains both the topology and dynamic information of the road network and represents the aggregated traffic condition in this region. A large value of x_c indicates that drivers could drive in a large component of roads while maintaining a relatively high speed. So it is reasonable to regard the x_c as our maximization objective. Compared with total system travel time, which is used in the bi-level network optimization problem, x_c maximization is more closely related to personal driving experiences.

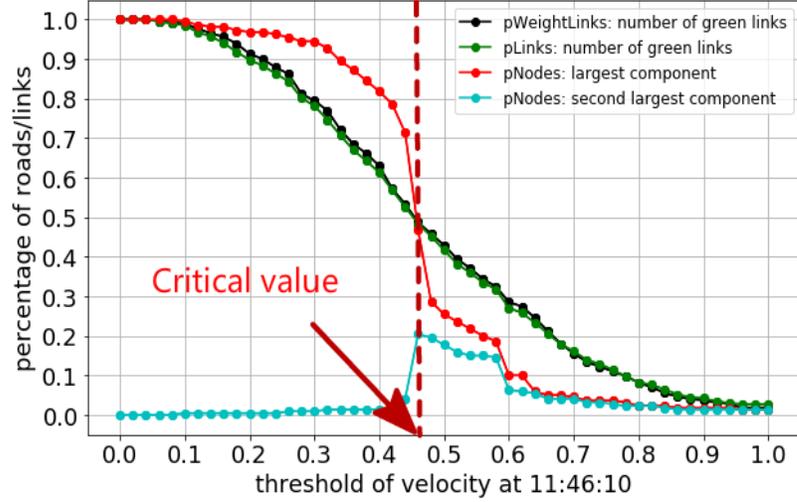


Figure 5.1.: Percolation curve in Sanyuanqiao at 18:14, December 8, 2015 (Tuesday)

5.3 Road Network Improvement Optimization

5.3.1 Optimization problem definition

Now we build the optimization of transportation management resources to maximize the critical value x_c . We refer to the road network as $G = G(V, E)$, where V and E represent intersections and road segments, respectively. There are n roads in this region: $E = \{1, 2, \dots, n\}$. The length of road segment i is l_i , $i = 1, 2, \dots, n$. The relative velocity of road segment i is defined as the ratio of current average velocity divided by the speed limit

$$r_i = \frac{v_i}{v_i^{limit}}, i = 1, 2, \dots, n \quad (5.3)$$

And the current global connectivity of good traffic condition road is described by $x_c = x_c(r_1, r_2, \dots, r_n)$. We would describe x_c as follows. For a predefined threshold $x \in [0, 1]$, the uncongested sub-network is

$$G_u(x) = G(V, E_u(x)) \quad (5.4)$$

where

$$E_u(x) = \{i \in E | r_i \geq x\} \quad (5.5)$$

Note that the relative size of largest connected component in $G_u(x)$ is a monotonically decreasing function with respect to x

$$L : [0, 1] \rightarrow [0, 1] \quad (5.6)$$

$$x \mapsto L(x|r_1, r_2, \dots, r_n) = \frac{|lc(G_u(x))|}{|G_u(x)|} \quad (5.7)$$

The $L(x|\vec{r})$ is displayed by the red curve in Figure 5.1. And the critical value x_c is defined as the x coordinate of the point with steepest descent

$$x_c(\vec{r}) = \operatorname{argmax}_{x \in [0,1]} \frac{-dL(x|\vec{r})}{dx} \quad (5.8)$$

In practice, the transportation agency allocates the human and capital traffic management resources to some link i and increases the relative velocity of link i from r_i to r'_i . It results in an improvement of connectivity of good traffic condition road network. Mathematically speaking, the optimization is cast as:

$$\operatorname{maximize} x_c(\vec{r}') \quad (5.9)$$

subject to

$$\sum_{i \in E'} c_i \leq B \quad (5.10)$$

$$\alpha_i s(l_i) = c_i, \quad i \in E' \quad (5.11)$$

$$r'_i = r_i + \alpha_i * 1\%, \quad i \in E' \quad (5.12)$$

$$\alpha_i \leq \alpha_{max}, \quad i \in E' \quad (5.13)$$

$$r'_i = r_i, \quad i \in E - E' \quad (5.14)$$

$$x_c(\vec{r}') = x_c(r'_1, r'_2, \dots, r'_n) \quad (5.15)$$

Here, we distribute the investment B on a subset of the road segment $E' \subset E$ other than the whole set E . Here,

$$E' = \{i \in E | r_i \leq \beta\} \quad (5.16)$$

Note that we set $\beta = 0.8$ in the following experiments. The introduction of E' indicates that if the road segment has a good traffic condition, we do not assign transportation management resources on it. The objective function (5.9) is to maximize x_c for the road network. Constraint (5.10) represents that the sum of investment on all road segments does not exceed the maximal budget B . In constraints (5.11), $s(l_i)$ is a monotonically increasing function with respect to road length l_i . These constraints mean that the augmentation α_i of relative velocity for the long road under investment c_i is not as substantial as the short road when the investment c_i is the same. Constraints (5.12) imply that the relative velocity of road segment i increases by α_i percentage. We assume that the average speed lift is upper bounded by α_{max} (we set α_{max} to be 20 in the experiments), reflected by constraints (5.13). It is because, in reality, it is not likely to improve the average speed of a road by a large percentage (more than 20% in experiments). Constraint (5.15) calculates the critical value x_c for the new road state: r'_1, r'_2, \dots, r'_n .

In our problem, we discrete the percolation curve and x_c is one of the element in a predefined finite set X with the steepest descent. We set X to be

$$X = \{0.00, 0.01, 0.02, \dots, 1.19, 1.20\} \quad (5.17)$$

In the following subsections, we would describe two resource allocation methods: 1) a simple allocation method; 2) a heuristic allocation method.

5.3.2 Method 1: simple allocation

In the simple allocation, the road investment c_i on road $i \in E'$ is determined by the length l_i and α_{min} . The straightforward idea is that the road investment is proportional to the road length l_i and also has an upper bound. Besides, we try to fully utilize the budget B as much as possible, that is, the constraint (5.10) is binding

or nearly binding. For road $i \in E - E'$, the investment is 0. And for road $i \in E'$, the investment for each road segment is:

$$c_i = \frac{s(l_i)B}{\sum_{j \in E'} s(l_j)}, i \in E' \quad (5.18)$$

Note that by constraint (5.13), we know if the calculated investment is larger than the upper bound $c_i > \alpha_{max}s(l_i)$, we would set

$$c_i = \alpha_{max}s(l_i), i \in E' \quad (5.19)$$

After the resource allocation using equations (5.18) and (5.19), we come up with a feasible solution for the optimization problem. We find that, for $i \in E - E'$,

$$r'_i = r_i > \beta \quad (5.20)$$

For $i \in E'$

$$r'_i = r_i + \alpha_i * 1\% \leq \beta + \alpha_{max} * 1\% \quad (5.21)$$

That is to say, after the network improvement, the uncongested roads are maintained and the congested roads are improved to maximize x_c . Though simple, method 1 incorporates the real-time speed information and provides us with an operational road improvement plan.

5.3.3 Method 2: heuristic optimization

The road network improvement optimization we build is a simulation based-optimization where the objective function (5.9) is a simulation result. As our objective function is calculated by evaluating the size of the largest connected component of $G_u(x)$ so it does not have an analytical form.

For simulation-optimization, there are two major kinds of approaches: 1) gradient approximation methods; 2) heuristic derivative-free methods. The former methods inherit the spirit of derivative optimization, like sequential quadratic programming (SQP) and interior point method (IPM) et al. The difficulty is to manually define

the calculated derivative or use the finite difference method. The latter methods include meta-heuristic methods: genetic algorithm, and heuristic methods: simulated annealing, pattern search algorithm et al. Firstly, for our optimization problem, the objective function has discrete values on X , so for a fixed point $\vec{c} = \{c_1, c_2, \dots, c_n\}$, the neighborhoods have the same objective value as \vec{c} . Hence, it is hard to manually define the derivative. And also, the finite difference method does not work because the difference is often zero. Secondly, in the simulated annealing or pattern search algorithm, each iteration moves from one point to the other and finally converges. Note that the number of variables in our optimization problem is the same as the number of road segments, which is at least hundreds. It is not efficient to move one feasible point sequentially to the optimal point because the feasible point is likely to drop into the local minimum.

In the end, we select the genetic algorithm to solve the optimization problem. In the genetic algorithm, a fixed number of initial feasible solutions are created. For each generation, the *elite* whose fitness function (objective function) is large are maintained. And the *crossover* and *mutation* are conducted to exploit current good results and explore unknown results, respectively.

5.4 Data

We use the same data as Chapter 4. We select a $2.3km \times 2.5km$ rectangle area controlled by the four points A, B, C, D and is belonging to Chaoyang District in Beijing (see Figure 5.2). The investigated time moment was 11:46:10, December 8, 2015, which was a Tuesday.

With the relative velocity threshold 0.54, we obtain the congested and uncongested road map for this region (see Figure 5.3). We do observe that the uncongested roads are homogeneous in space. When we define diverse relative speed thresholds, we are able to get the largest component size curve for the uncongested roads (see Figure 5.4). The curve shows that the rapid drop exists at $x_c = 0.58$ of the relative velocity

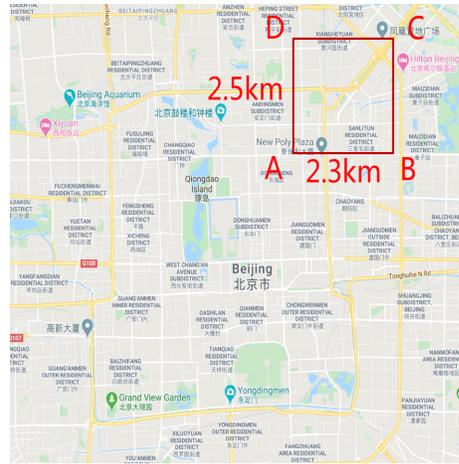


Figure 5.2.: Selected region for optimization

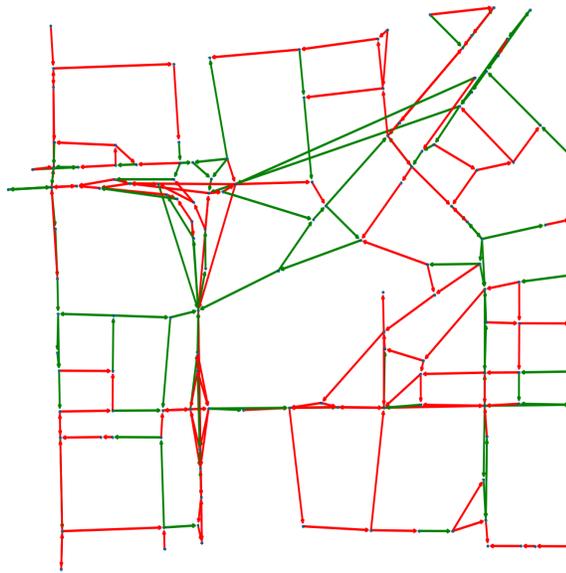


Figure 5.3.: The traffic state with threshold 0.54

threshold. Our optimization objective is to increase the x_c and therefore enhance the local connectivity of good traffic condition roads.

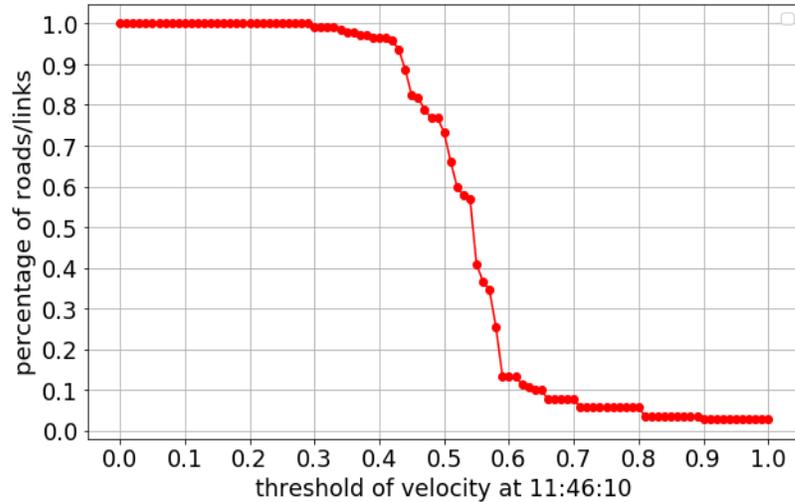


Figure 5.4.: Percolation curve

Now we present some statistical results of the road segments in this area in Figures (5.5) and (5.6). There are in total 247 road segments in this region with the minimal and maximal road segment length as 0.054 and 1.300 kilometers. And the average road segment length is 0.230 kilometers and the total length is 56.93 kilometers. The road length distribution is displayed in Figure (5.5). Most road segments have a length between 0.05 and 0.50 kilometers. From Figure (5.6), we know that during the noon moment, the relative speed distribution is symmetrical to the relative speed between 0.5 and 0.6 and has a long tail. We observe that there are some roads with high relative speed (≥ 1.5). There are two probabilities: 1) our speed limit data is out-of-date; 2) there is a little number of vehicles moving on the street and they move fast. Notice that our optimization objective x_c is the relative speed where the largest components of good traffic condition roads shrink rapidly, the extreme large relative speed data would not affect the x_c . Therefore, we do not clean these large relative speed data.

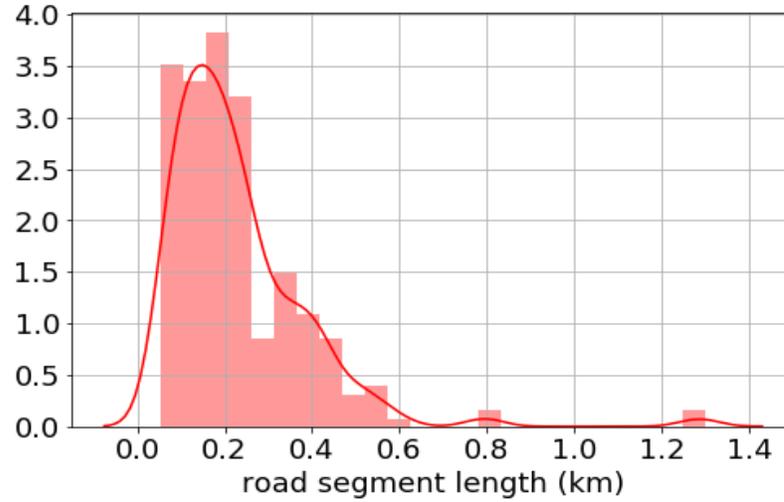


Figure 5.5.: Road segment length distribution

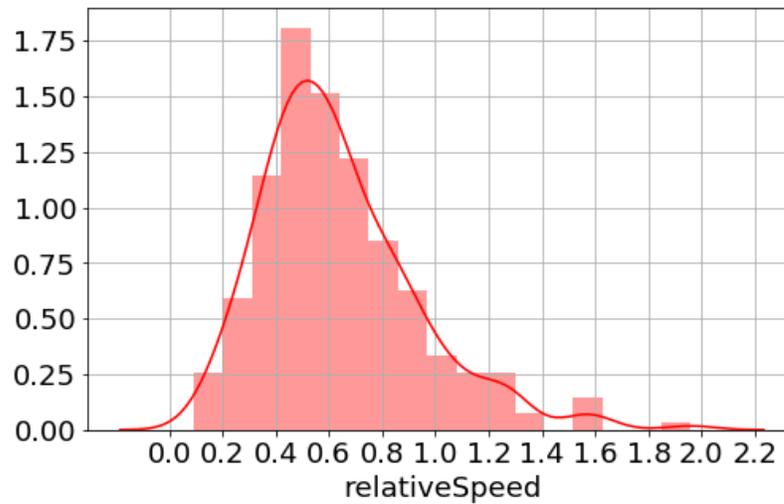


Figure 5.6.: Relative speed distribution

5.5 Experiments

5.5.1 Parameters preparation

In constraints 5.11, s function returns the investment cost given the relative speed lift. In fact, if the road is long and wide, s is large. Because in our dataset, the road

segment width data is not accurate, we only consider length effect. We assume for a 1 km road, 10,000 dollars investment one year improves the road relative speed by 1 percent. That is to say,

$$s(l) = 10,000 \text{ dollars}/\text{km}/\% * l \quad (5.22)$$

Under this setting, if we want to improve the relative speed for each link by 1 %, the annual cost C is

$$\begin{aligned} B &= 1\% * 56.93 \text{ km} * 10,000 \text{ dollars}/\text{km}/\% \\ &= 569,300 \text{ dollars}. \end{aligned} \quad (5.23)$$

Based on this data, we vary the annual road improvement investment B to be from 0 to 4.8 million dollars.

5.5.2 Method 1: simple allocation

In method 1, according to equation (5.18) and (5.19), the transportation agency invests homogeneously based on the road segment length l_i . This investment strategy is not efficient as not all road segment has the same marginal benefit to the connectivity of good condition roads. The lift of critical value x_c is displayed in the black curve Figure (5.7). As the budget increases from 0 to 4.8 million, x_c increments from 0.58 to 0.63. And the increasing trend is quite close to linear.

An interesting phenomenon is that when the budget upper bound rises from 0 to 0.3 million dollars, the x_c decreases from 0.58 to 0.54. In Figure (5.4), the component curve has steep descent from 0.50 to 0.60. The descent is similar to these values. When we invest 0.3 million dollars through method 1, the steepest descent value x_c *jumps* from 0.58 to 0.54. It reveals that we may develop more a robust definition of x_c to consider both the local and global descent of the percolation curve in the future. In the meantime, the drop of x_c demonstrates that irrational road management allocation may make the connectivity of good traffic condition roads worse.

To sum up, method 1 is a simple road improvement investment strategy that serves as the baseline for method 2.

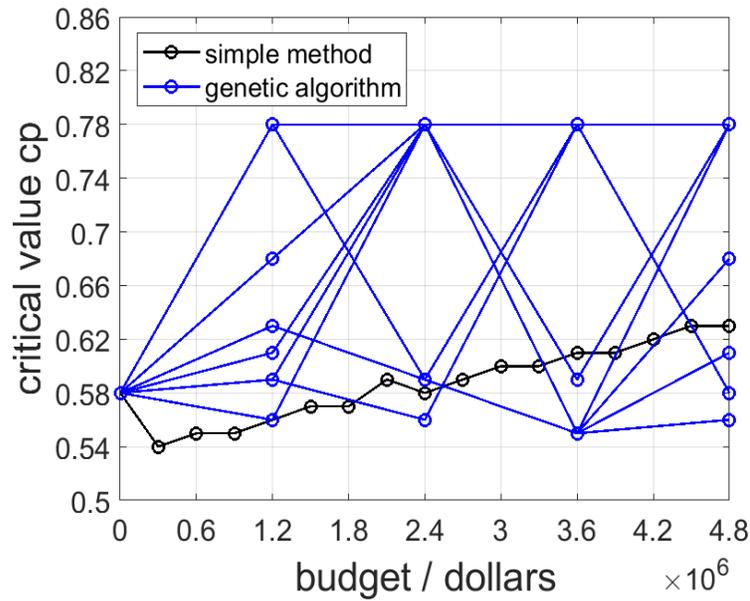


Figure 5.7.: Road improvement for Dongzhimen area at 11:46

5.5.3 Method 2: heuristic optimization

In the genetic algorithm, there are several parameters that we need to decide. As our experimental optimization has 247 decision variables (247 road segments), we set the population size to be 200. The scaling function and selection function is defined as *rank* and *stochastic uniform*, respectively. They are default parameter settings and also have good performance in comparison to other values we try. The elite count is 10, which means that the best 10 feasible solutions in each generation would be passed to the next generation. Finally, *cross* fraction and *mutation* fraction are 0.8 and 0.2. There is a higher probability to exploit the good *genes* than to explore the unknown genes.

We vary the budget values from 0 to 4.8 million, the same as method 1. And we set the maximal running time to be 600 seconds to satisfy real-time traffic management requirements. Since the genetic algorithm produces different results at different

Table 5.1.: Parameter setting for the genetic algorithm

Parameters	Values
population size	200
scaling function	rank
selection function	stochastic uniform
elite count	10
crossover fraction	0.8

implementations, we repeat the process 10 times and get the best of 10 critical value curves (see blue curves in Figure (5.7)).

In Figure (5.7), we observe that the maximal x_c value of 5 instances for the budget 1.2, 2.4, 3.6, 4.8 million are both 0.78, which is evidently larger than the results obtained by method 1 (from 0.54 to 0.63). It demonstrates that as a meta-heuristic method, the genetic algorithm does not guarantee to output the global optimum. But it can offer us a feasible solution much better than method 1.

On the other hand, most blues points in Figure (5.7) are located above the black curve. It shows that even we run the genetic algorithm for one time, we are likely to get a better solution than method 1. So when it comes to real-time traffic management, the transportation agency can run one instance of the genetic algorithm and simple allocation and select the best plan among them.

In Figure (5.8) and (5.9), we compare the spatial traffic management allocation for method 1 and method 2. And the width of the red road segment is proportional to the investment on this road divided by the road segment length. We focus on a budget of 2.4 million dollars. At this budget level, the y -coordinate of the black curve and the best blue curve are 0.58 and 0.78. In Figure (5.8), the investment is homogeneous among different links. But in Figure (5.9), the transportation agency puts the resources at only a proportion of road segments. And these road segments are mainly bottlenecks.

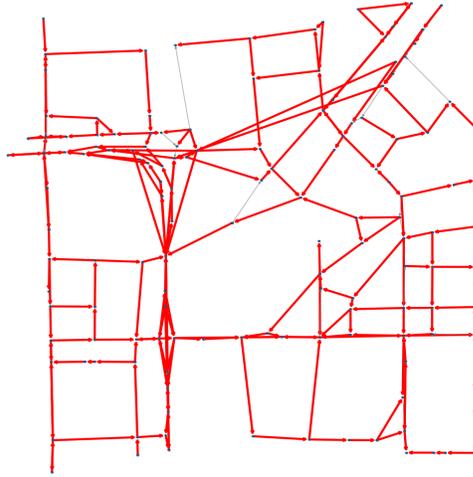


Figure 5.8.: Link investment using Method 1 ($B = 2.4M$)

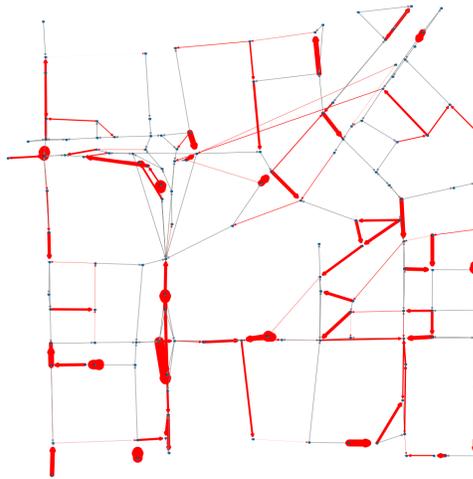


Figure 5.9.: Link investment using Method 2 ($B = 2.4M$)

5.5.4 More experiments of method 2

This subsection verifies that method 2 is still valid for other instances. We vary the spatial and temporal range.

In Figure (5.10) and (5.11), we implement the method 2 for the same region but different time (18:14). This was the evening traffic moment for Beijing and the initial critical value for the region was lower than 11:46 (0.40 VS 0.58). In Figure (5.10), we can find that the genetic algorithm achieves a better objective value than the simple method with a high probability as well. There are only 5 blue points that are situated lower than the baseline black curve. This instance demonstrates that method 2 has a good performance during the traffic peak hours for this region.

In addition, we run the optimization for the Xidan area at 11:46 (see Figure (5.12), (5.13)). The good optimization performance is still obvious shown in blue points in Figure (5.12). But different from the Dongzhimen area which has a large-scale overpass, the Xidan area is a commercial area and has a homogeneous road structure. This is the reason why higher proportion of the road segments are invested (colored red) in Figure (5.13) than Figure (5.11).

Table 5.2.: Budget constraint binding results (unit: thousand dollars)

Region	Dongzhimen	Dongzhimen	Xidan
Time	11:46	18:14	11:46
Method 1	2400	2400	2400
Method 2	1001.8	895.1	1409.2

In the end, we review the three instances that we have tested and present the budget binding statistic in Table 5.2. For the three instances, method 1 utilizes the full budget of 2.4 million, and method 2 only uses nearly half of all the potential budget. It indicates that method 2 gives more efficient transportation resource allocations than method 1 greatly. And these results also remind us that it is probable to discover better solutions than method 2.

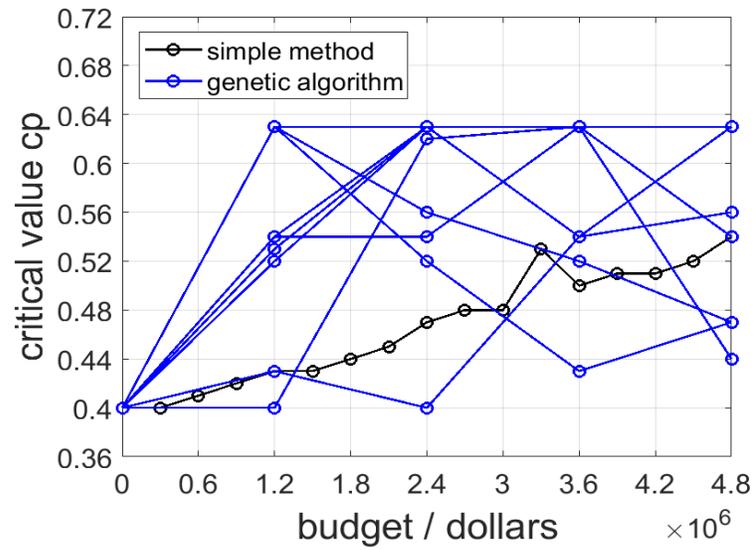


Figure 5.10.: Road improvement for Dongzhimen area at 18:14

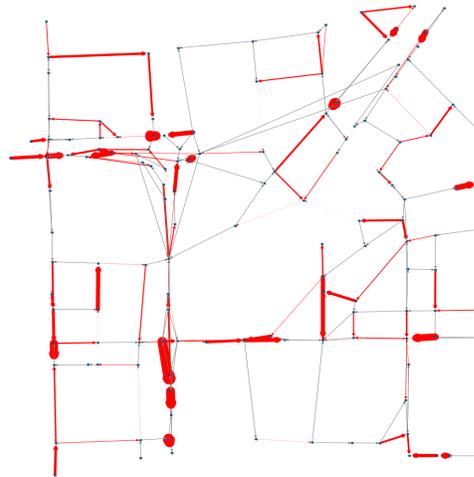


Figure 5.11.: Investment results using method 2 for Dongzhimen area at 18:14

5.6 Conclusion

In this chapter, we establish simulation-based optimization for road network improvement. Our objective is to maximize the critical value of connectivity of good

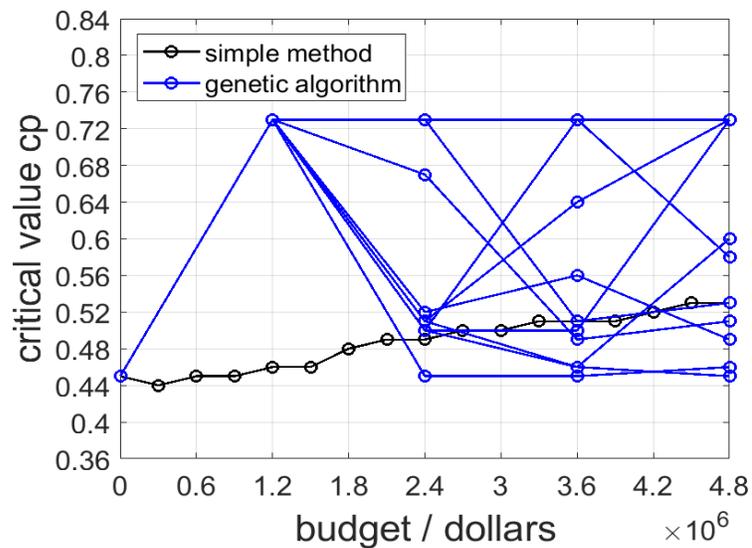


Figure 5.12.: Road improvement for Xidan area at 11:46

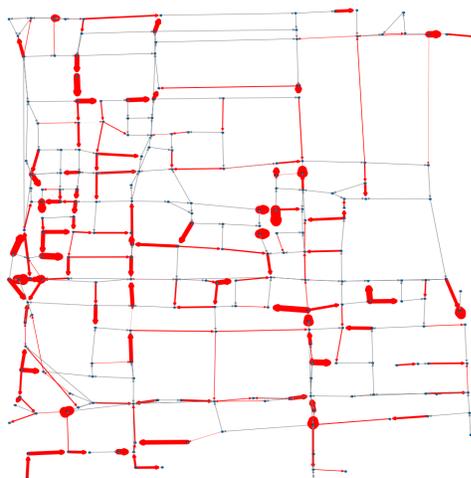


Figure 5.13.: Investment results using method 2 for Xidan area at 11:46

traffic condition roads defined in chapter 4. We propose a simple method and a meta-heuristic method using the genetic algorithm to solve the optimization problem. To demonstrate the method performance, we implement the two methods for the Beijing

road network with varying time and areas. Results show that our second method is able to achieve high objective values in a short time and generate superior road segment improvement plans. We would extend our method 2 to achieve higher objective function and run more configuration cases in the future.

6. CONCLUSIONS

6.1 Summary

This thesis develops structural and dynamic models for complex road networks, formulates and solves network improvement optimization. The novelty comes from the adoption of percolation theory in complex road networks, the overall traffic dynamics analysis, and management framework using link speed data only.

In fact, online road network data provider, like OpenStreetMap, is becoming a popular resource for transportation researchers to access road network data. Researchers built road network models using the data. But most researchers did not highlight the important step of modeling: data cleaning. We have shown that several nodes in shapefiles represent the same intersection in reality. Without pre-processing, road network analysis based on raw shapefile data may lead to errors. To deal with this issue, our node merging algorithm serves as a useful and efficient tool to conduct pre-processing for road network data. In addition, in this thesis, we analyze the distinction of node degree distribution for Beijing and Shanghai, find and explain that Shanghai has more small-degree nodes than Beijing.

In the context of road dynamics modeling, our analysis based on percolation component curves brings fresh air to road network traffic dynamic researches. There is a research gap here that we lack a model to depict the road network congestion considering both the network topology and also traffic dynamics. The largest merit of the percolation model is that it captures both the network topology and speed mechanisms and relies on easily accessible link speed data. Moreover, we inherit the core idea of network partition under the modularity objective and extend the Louvain algorithm to fit the road network partition problem. Our partition is quite efficient and is able to generate the road network partition plan within seconds. Such an

overall framework can be an alternative choice of regional traffic dynamic detection and management.

The last part of the paper optimizes the best allocation of investment distribution over all links in the metropolitan cities. Traditional network optimizations use the total travel time as the objective function and put the flow equilibrium at the lower level. In our optimization, we do not need to make any assumption over people's routing decisions. Our objective is to maximize the x_c , local connectivity of good traffic condition roads for a specific area. The budget is one of the constraints. It is shown that heuristic methods can output high-quality feasible results in a short time. When it comes to realistic transportation control, people could choose the best of the simple allocation method and the heuristic method.

6.2 Future Work

There are three potential future directions:

1) More concise modeling of the overpass, T-intersection, tunnels in the urban road networks. In our models, we use the distance threshold δ to distinguish whether two nodes in the shapefile represent the same intersection, and we refer to 50 meters as the threshold in the experiments. The threshold δ can be extended to consider the aforementioned road network sub-structures more concisely.

2) Detailed design of regional traffic control using speed data only. In chapter 4, we have established the link speed-based traffic monitoring and control system. But the last step of the framework, regional traffic control is not developed. We may get insights from the current MFD control methods like gating, pricing, but we need to avoid using the link density data, which is hard to get.

Figure 6.1 depicts the basic process of gating strategy. In the methodology proposed by [45], transportation engineers detect first the link density and then the number of vehicles at a time step k in the protected network N . Based on the real-

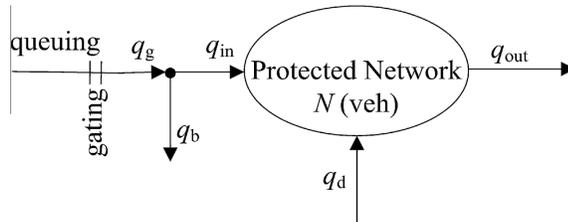


Figure 6.1.: A gating strategy for protected road network N [45]

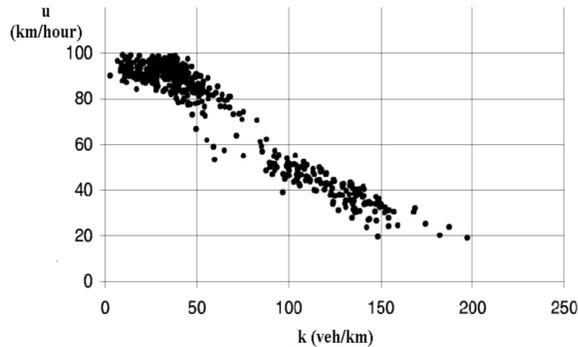


Figure 6.2.: Relationship between speed u and density k for a motorway [71]

time regional density data and expected regional density, they control the inflow $q_g(k)$ to manage the traffic inside the protected network N .

$$q_g(k) = q_g(k-1) - K_p(TTS(k) - TTS(k-1)) + K_I(\hat{TTS} - TTS(k)) \quad (6.1)$$

Under this control strategy, the link density serves as the monitoring variables of the traffic state for network N . On the other hand, the fundamental diagram tells us that when the density is not small, as the link density increases, the average speed decreases monotonically (In Figure 6.2, the record dots can be fitted by a decreasing function.) These results inform us that it is probable to derive some speed-based regional traffic control strategy from traditional density-based strategies.

3) Higher quality optimization solutions. The meta-heuristic genetic algorithm can output a relatively good solution that guides people to allocate transportation resources. We do observe that, when the budget upper bound increases, the best

critical value keeps constant. It motivates us to propose a better method to find near-optimal solutions efficiently.

REFERENCES

- [1] Schrank David, Eisele Bill, and Lomax Tim. 2019 urban mobility report. *Texas AM Transportation Institute*, 2019.
- [2] INRIX. INRIX 2018 Global Traffic Scorecard. <https://inrix.com/scorecard/>. Accessed Jan, 2020.
- [3] Dietrich Braess. Über ein paradoxon aus der verkehrsplanung. *Unternehmensforschung*, 12(1):258–268, 1968.
- [4] Stephen Marshall, Jorge Gil, Karl Kropf, Martin Tomko, and Lucas Figueiredo. Street network studies: from networks to models and their representations. *Networks and Spatial Economics*, 18(3):735–749, 2018.
- [5] Geoff Boeing. A multi-scale analysis of 27,000 urban street networks: Every us city, town, urbanized area, and zillow neighborhood. *Environment and Planning B: Urban Analytics and City Science*, page 2399808318784595.
- [6] Paolo Crucitti, Vito Latora, and Sergio Porta. Centrality measures in spatial networks of urban streets. *Physical Review E*, 73(3):036125, 2006.
- [7] Xin-Ling Guo and Zhe-Ming Lu. Urban road network and taxi network modeling based on complex network theory. *J Inf Hiding Multimedia Signal Proc*, 7(3):558–568, 2016.
- [8] Emanuele Strano, Andrea Giometto, Saray Shai, Enrico Bertuzzo, Peter J Mucha, and Andrea Rinaldo. The scaling structure of the global road network. *Royal Society open science*, 4(10):170590, 2017.
- [9] Sergio Porta, Paolo Crucitti, and Vito Latora. The network analysis of urban streets: a dual approach. *Physica A: Statistical Mechanics and its Applications*, 369(2):853–866, 2006.
- [10] Xianyuan Zhan, Satish V Ukkusuri, and P Suresh C Rao. Dynamics of functional failures and recovery in complex road networks. *Physical Review E*, 96(5):052301, 2017.
- [11] Albert-László Barabási et al. *Network science*. Cambridge university press, 2016.
- [12] Carlos Molinero, Roberto Murcio, and Elsa Arcaute. The angular nature of road networks. *Scientific reports*, 7(1):4312, 2017.
- [13] Vamsi Kalapala, Vishal Sanwalani, Aaron Clauset, and Cristopher Moore. Scale invariance in road networks. *Physical Review E*, 73(2):026130, 2006.
- [14] Bin Jiang and Christophe Claramunt. Topological analysis of urban street networks. *Environment and Planning B: Planning and design*, 31(1):151–162, 2004.

- [15] Emanuele Strano, Vincenzo Nicosia, Vito Latora, Sergio Porta, and Marc Barthélemy. Elementary processes governing the evolution of road networks. *Scientific reports*, 2:296, 2012.
- [16] Padraig Corcoran, Peter Mooney, and Michela Bertolotto. Analysing the growth of openstreetmap networks. *Spatial Statistics*, 3:21–32, 2013.
- [17] Wenbo Mo, Yong Wang, Yingxue Zhang, and Dafang Zhuang. Impacts of road network expansion on landscape ecological risk in a megacity, china: A case study of beijing. *Science of the Total Environment*, 574:1000–1011, 2017.
- [18] Limiao Zhang, Guanwen Zeng, Daqing Li, Hai-Jun Huang, H Eugene Stanley, and Shlomo Havlin. Scale-free resilience of real traffic jams. *Proceedings of the National Academy of Sciences*, 116(18):8673–8678, 2019.
- [19] Daqing Li, Bowen Fu, Yunpeng Wang, Guangquan Lu, Yehiel Berezin, H Eugene Stanley, and Shlomo Havlin. Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proceedings of the National Academy of Sciences*, 112(3):669–672, 2015.
- [20] Luis E Olmos, Serdar Çolak, Sajjad Shafiei, Meead Saberi, and Marta C González. Macroscopic dynamics and the collapse of urban traffic. *Proceedings of the National Academy of Sciences*, 115(50):12654–12661, 2018.
- [21] Pu Wang, Timothy Hunter, Alexandre M Bayen, Katja Schechtner, and Marta C González. Understanding road usage patterns in urban areas. *Scientific reports*, 2:1001, 2012.
- [22] Serdar Çolak, Antonio Lima, and Marta C González. Understanding congested travel in urban areas. *Nature communications*, 7:10793, 2016.
- [23] Song Gao, Yaoli Wang, Yong Gao, and Yu Liu. Understanding urban traffic-flow characteristics: a rethinking of betweenness centrality. *Environment and Planning B: Planning and Design*, 40(1):135–153, 2013.
- [24] Alexander A Ganin, Maksim Kitsak, Dayton Marchese, Jeffrey M Keisler, Thomas Seager, and Igor Linkov. Resilience and efficiency in transportation networks. *Science advances*, 3(12):e1701079, 2017.
- [25] Carlos F Daganzo. The cell transmission model, part ii: network traffic. *Transportation Research Part B: Methodological*, 29(2):79–93, 1995.
- [26] Meead Saberi, Mudabber Ashfaq, Homayoun Hamedmoghadam, Seyed Amir Hosseini, Ziyuan Gu, Sajjad Shafiei, Divya J Nair, Vinayak Dixit, Lauren Gardner, S Travis Waller, et al. A simple contagion process describes spreading of traffic jams in urban networks. *arXiv preprint arXiv:1906.00585*, 2019.
- [27] Eleonora Andreotti, Armando Bazzani, Sandro Rambaldi, Nicola Guglielmi, and P Freguglia. Modeling traffic fluctuations and congestion on a road network. *Advances in Complex Systems*, 18(03n04):1550009, 2015.
- [28] Xianyuan Zhan. *Novel Approaches to Model Congestion Evolution and Dependencies in Complex Road Networks*. PhD thesis, Purdue University, 2017.

- [29] Liang Zhao, Ying-Cheng Lai, Kwangho Park, and Nong Ye. Onset of traffic congestion in complex networks. *Physical Review E*, 71(2):026125, 2005.
- [30] Serdar Çolak. *Navigating congested cities: understanding urban mobility using new data sources*. PhD thesis, Massachusetts Institute of Technology, 2016.
- [31] Serdar Çolak, Christian M Schneider, Pu Wang, and Marta C Gonzalez. On the role of spatial dynamics and topology on network flows. *New Journal of Physics*, 15(11):113037, 2013.
- [32] Ioannis E Antoniou and ET Tsompa. Statistical analysis of weighted networks. *Discrete dynamics in Nature and Society*, 2008, 2008.
- [33] Alain Barrat, Marc Barthelemy, Romualdo Pastor-Satorras, and Alessandro Vespignani. The architecture of complex weighted networks. *Proceedings of the national academy of sciences*, 101(11):3747–3752, 2004.
- [34] Neng Wan, F Zhan, and Zhongliang Cai. A spatially weighted degree model for network vulnerability analysis. *Geo-spatial Information Science*, 14(4):274–281, 2011.
- [35] Chapter 3: The 13 controlling criteria. *US Department of Transportation Federal Highway Administration*, 2019.
- [36] Jiaqiu Wang. Resilience of self-organised and top-down planned cities—a case study on london and beijing street networks. *PloS one*, 10(12):e0141736, 2015.
- [37] Hai Wang and Amedeo Odoni. Approximating the performance of a “last mile” transportation system. *Transportation Science*, 50(2):659–675, 2014.
- [38] Yosef Sheffi. *Urban Transportation networks: equilibrium analysis with mathematical programming models*. Prentice-Hall, Inc, 1985.
- [39] Jichang Zhao, Daqing Li, Hillel Sanhedrai, Reuven Cohen, and Shlomo Havlin. Spatio-temporal propagation of cascading overload failures in spatially embedded networks. *Nature communications*, 7(1):1–6, 2016.
- [40] Simon R Broadbent and John M Hammersley. Percolation processes: I. crystals and mazes. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 53, pages 629–641. Cambridge University Press, 1957.
- [41] Daniele De Martino, Luca Dall’Asta, Ginestra Bianconi, and Matteo Marsili. Congestion phenomena on complex networks. *Physical Review E*, 79(1):015101, 2009.
- [42] Alireza Mostafizi, Shangjia Dong, and Haizhong Wang. Percolation phenomenon in connected vehicle network through a multi-agent approach: Mobility benefits and market penetration. *Transportation Research Part C: Emerging Technologies*, 85:312–333, 2017.
- [43] Nikolas Geroliminis and Carlos F Daganzo. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42(9):759–770, 2008.

- [44] Nikolas Geroliminis and Jie Sun. Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B: Methodological*, 45(3):605–617, 2011.
- [45] Mehdi Keyvan-Ekbatani, Markos Papageorgiou, and Ioannis Papamichail. Urban congestion gating control based on reduced operational network fundamental diagrams. *Transportation Research Part C: Emerging Technologies*, 33:74–87, 2013.
- [46] Nan Zheng, Rashid A Waraich, Kay W Axhausen, and Nikolas Geroliminis. A dynamic cordon pricing scheme combining the macroscopic fundamental diagram and an agent-based traffic model. *Transportation Research Part A: Policy and Practice*, 46(8):1291–1303, 2012.
- [47] Mehmet Yildirimoglu, Mohsen Ramezani, and Nikolas Geroliminis. Equilibrium analysis and route guidance in large-scale networks with mfd dynamics. *Transportation Research Procedia*, 9:185–204, 2015.
- [48] Yuxuan Ji and Nikolas Geroliminis. On the spatial partitioning of urban transportation networks. *Transportation Research Part B: Methodological*, 46(10):1639–1656, 2012.
- [49] Kang An, Yi-Chang Chiu, Xianbiao Hu, and Xiaohong Chen. A network partitioning algorithmic approach for macroscopic fundamental diagram-based hierarchical traffic network management. *IEEE Transactions on Intelligent Transportation Systems*, 19(4):1130–1139, 2017.
- [50] Mohammadreza Saeedmanesh and Nikolas Geroliminis. Clustering of heterogeneous networks with directional flows based on “snake” similarities. *Transportation Research Part B: Methodological*, 91:250–269, 2016.
- [51] Alessandra Pascale, Dimitrios Mavroeidis, and Hoang Thanh Lam. Spatiotemporal clustering of urban networks: Real case scenario in london. *Transportation Research Record*, 2491(1):81–89, 2015.
- [52] Chuishi Meng, Xiuwen Yi, Lu Su, Jing Gao, and Yu Zheng. City-wide traffic volume inference with loop detector data and taxi trajectories. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 1–10, 2017.
- [53] U.S.Department of Transportation Federal Highway Administration. *Traffic Detector Handbook: Third Edition—Volume II*. 2006.
- [54] Christine Buisson and Cyril Ladier. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. *Transportation Research Record*, 2124(1):127–136, 2009.
- [55] Mark Newman. *Networks*. Oxford university press, 2018.
- [56] Jon Kleinberg and Eva Tardos. *Algorithm design*. Pearson Education India, 2006.
- [57] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.

- [58] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.
- [59] Aaron Clauset, Mark EJ Newman, and Christopher Moore. Finding community structure in very large networks. *Physical review E*, 70(6):066111, 2004.
- [60] Pascal Pons and Matthieu Latapy. Computing communities in large networks using random walks. In *International symposium on computer and information sciences*, pages 284–293. Springer, 2005.
- [61] Ken Wakita and Toshiyuki Tsurumi. Finding community structure in mega-scale social networks. In *Proceedings of the 16th international conference on World Wide Web*, pages 1275–1276, 2007.
- [62] Jure Leskovec. Louvain algorithm, cs224w: Machine learning with graphs. 2019.
- [63] Hai Yang and Michael G H. Bell. Models and algorithms for road network design: a review and some new developments. *Transport Reviews*, 18(3):257–278, 1998.
- [64] Satish V Ukkusuri, Tom V Mathew, and S Travis Waller. Robust transportation network design under demand uncertainty. *Computer-Aided Civil and Infrastructure Engineering*, 22(1):6–18, 2007.
- [65] David ZW Wang, Haoxiang Liu, and WY Szeto. A novel discrete network design problem formulation and its global optimization solution algorithm. *Transportation Research Part E: Logistics and Transportation Review*, 79:213–230, 2015.
- [66] Min Xu, Qiang Meng, and Kai Liu. Network user equilibrium problems for the mixed battery electric vehicles and gasoline vehicles subject to battery swapping stations and road grade constraints. *Transportation Research Part B: Methodological*, 99:138–166, 2017.
- [67] Xuan Di, Henry X Liu, Jong-Shi Pang, and Xuegang Jeff Ban. Boundedly rational user equilibria (brue): mathematical formulation and solution sets. *Transportation Research Part B: Methodological*, 57:300–313, 2013.
- [68] Roger Guimerà, Albert Díaz-Guilera, Fernando Vega-Redondo, Antonio Cabrales, and Alex Arenas. Optimal network topologies for local search with congestion. *Physical review letters*, 89(24):248701, 2002.
- [69] DE Boyce, BN Janson, and RW Eash. The effect on equilibrium trip assignment of different link congestion functions. *Transportation Research Part A: General*, 15(3):223–232, 1981.
- [70] David Aldous and Marc Barthelemy. Optimal geometry of transportation networks. *Physical Review E*, 99(5):052303, 2019.
- [71] LH Immers and S Logghe. Traffic flow theory. *Faculty of Engineering, Department of Civil Engineering, Section Traffic and Infrastructure, Kasteelpark Arenberg*, 40:21, 2002.

VITA

Jiawei Xue was born on December 16, 1995 in Shaoxing, Zhejiang province, China. He studied in the Department of Mathematical Sciences at Tsinghua University from 2013 to 2015. Afterwards, Jiawei Xue studied and graduated from the Department of Civil Engineering with a B.E. degree at Tsinghua University in 2018. He pursued his M.S. degree at Purdue University from 2018 to 2020 under the supervision of Prof. Satish Ukkusuri. During the M.S. period, he investigated the epidemic spreading in contact network, spatial electric vehicle charging station optimization, electric bus route design, urban congestion analysis, multi-armed bandit routing, Braess's paradox in scale-free networks. He authored 5 conference and journal paper. He would continue as a Ph.D. student in Prof. Satish Ukkusuri's research group.