UCLA
UCLA Electronic Theses and Dissertations

Title
A Complex Network Method for Traffic Modeling and Control

Permalink
https://escholarship.org/uc/item/5ct600k6

Author
Yeung, Fiona

Publication Date
2017

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
Los Angeles

A Complex Network Method for Traffic Modeling and Control

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Statistics

by

Fiona Chehong Yeung

2017

© Copyright by

Fiona Chehong Yeung

2017

ABSTRACT OF THE THESIS

A Complex Network Method for
Traffic Modeling and Control

by

Fiona Chehong Yeung
Master of Science in Statistics
University of California, Los Angeles, 2017
Professor Mark Stephen Handcock, Chair

Transportation systems are the economic foundation of any regional development. Our reliance on transportation to move goods and resources and to ensure access to labor to increase productivity, all have tremendous impact on revenue generation and growth. Traffic congestion is an inevitable byproduct of economic growth; the costs of traffic is not just time wasted, but also include the financial loss and environmental impacts of fuel being wasted. As an effort to understand congestion formation, this project investigates modeling traffic as a network and uses a percolation model to identify a normal traffic pattern as exhibited by the inhabitants of the region. Using real street maps from the OpenStreetMap project, morning, noon, and evening rush-hour traffic zones in Westwood Village were created to simulate the travel behavior of the inhabitants. The street bottlenecks identified for a 24-hr period were then compared to those formed from a uniform traffic flow. The results from this study may provide the foundation for a reasonable starting configuration for a self-organizing traffic light network that can dynamically adapt to unexpected demand in real-time.

The thesis of Fiona Chehong Yeung is approved.

Arash Ali Amini

Chad J Hazlett

Mark Stephen Handcock, Committee Chair

University of California, Los Angeles

2017

TABLE OF CONTENTS

1 Introduction **1**

2 Data Selection **3**

2.1 Synthetic Data Generation for the Vehicular Traffic Scenario 3

2.2 Overview of Simulation of Urban Mobility (SUMO) 3

2.2.1 User Inputs 4

2.2.2 SUMO Platform 8

2.2.3 SUMO Output 10

3 Traffic Dynamics Modeled by Percolation Transition **12**

3.1 Building a Functional Network from a Structural Network 12

3.2 The Application of Percolation Theory in Identifying Traffic Pattern and Bottlenecks 14

3.3 Finding Percolation Transitions Using Simulated Traffic Flow 15

3.3.1 Discussion of the Bottlenecks Identified During Rush-Hour Periods . 17

3.4 Significance of the Rush-Hour Bottleneck Distributions Compared To Randomness and/or Extraneous Factors 18

3.4.1 Synthetic Data Generation for the Reference Null Distribution 18

3.4.2 Bottleneck Distributions Similarity Measure 18

3.4.3 Within- and Between- Group Comparison of the Similarity Measure . 19

3.5 Validation of Modeling Traffic Dynamics as a Percolation Process 20

4 Conclusion **35**

4.1 Summary 35

4.2 Shortcomings of the Percolation Theory Approach 35

5 Future Work **37**

References **38**

LIST OF FIGURES

2.1 Data Generation Overview 4

2.2 Westwood Village Street Map 5

2.3 Westwood Village TAZ 6

2.4 Origin-Destination Matrices 7

2.5 SUMO Simulation 10

3.1 Road Network Relative Velocity 22

3.2 Cluster Sizes During Morning Rush Hour 23

3.3 Largest Clusters During Morning Rush Hour 24

3.4 Cluster Sizes During Noon Rush Hour 25

3.5 Largest Clusters During Noon Rush Hour 26

3.6 Cluster Sizes During Evening Rush Hour 27

3.7 Largest Clusters During Evening Rush Hour 28

3.8 Line Plot of Mean q_c Values in the 24-hr Rush-Hour Scenario 29

3.9 Most Frequent Bottleneck Streets 30

3.10 Bottleneck Streets Geographical Locations 31

3.11 TV-norm for Rush-Hour and Uniform Traffic Scenarios (Within Group) 32

3.12 TV-norm for Rush-Hour and Uniform Traffic Scenarios (Between Groups) . . . 33

3.13 Boxplot of Mean q_c Values in the Uniform Traffic and Rush-Hour Scenario . . . 34

CHAPTER 1

Introduction

Regional economic competitiveness is heavily dependent on an efficient and reliable transportation system. Studies have shown that even with planned improvements, our transportation system would not keep pace with projected increases in freight and general traffic. Built up over 30 years at a cost of $400 million and completed in 2013, the Automated Traffic Surveillance and Control (ATSAC) system in Los Angeles uses magnetic sensors on roads that measure the flow of traffic, hundreds of CCTV cameras, communication equipment, and a centralized computer system that synchronizes the traffic lights to keep cars moving. With an estimated deployment cost of $78,000 per intersection [SS10], a 5-mile drive on Los Angeles streets slightly improves to 17.2 mins from 20 mins, with average speed of 17.3 mph, up from 15mph. This is, inarguably, an improvement, though insignificant for most commuters.

The effectiveness of any traffic control system ultimately depends on the predictability, adaptability, and scalability of the algorithms used. With its high cost and the need of human intervention to synchronize the traffic lights in the network, the current centralized system cannot be scaled quickly and economically to satisfy the changing and growing demand of highly populated cities. As such, this project aims to examine self-adaptive methods in literature for their applicability in improving traffic congestion. In particular, we will focus on complex system- and network-based methodologies in this thesis. In order to validate and compare their performance, we emulated rush-hour traffic by setting up a vehicular flow scenario with congestions taking place at specific times and locations. In some cases of algorithmic comparison, such as in assessing the effectiveness of traffic light synchronization schemes, it is critical to use simulated data as opposed to empirical data because we cannot

repeat exactly the same traffic conditions under different algorithms in a real-life setting. In other cases, we validate the algorithms by comparing their results with the true model that generated the data. Since there is no way of knowing the true data generating model of any empirical data, we can only use synthetic data for this purpose. For this study, validation of any descriptive algorithms that characterize traffic patterns is mainly based on comparison against the null distribution generated by the uniform traffic flow scenario which does not include any periods of increased traffic demand.

In this thesis, Chapter 2 discusses the details of the data generating process underlying the simulated traffic scenario used throughout this thesis. Chapter 3 validates a method for identifying street bottlenecks by modeling traffic dynamics as percolation transitions. Chapter 4 presents the conclusion from this study and Chapter 5 suggests future work.

CHAPTER 2

Data Selection

2.1 Synthetic Data Generation for the Vehicular Traffic Scenario

Since the goal of this project is to validate and compare the performance and adaptability of various algorithms under repeatable conditions, a 24-hr simulated traffic data scenario was used throughout this project in lieu of empirical data. We chose the Simulation of Urban Mobility (SUMO) as our simulation generating platform to help with data generation because of its extensive online documentation and readily available basic functionalities such as routing, random vehicle trip sampling, map importing, and traffic visualization. The specifics of the traffic data scenario, such as vehicle flow direction, departure time, and density, can be customized by the user to create any traffic scenario of choice. This section describes in detail the parameters and settings used in the data generating process carried out by SUMO for this project.

2.2 Overview of Simulation of Urban Mobility (SUMO)

SUMO is a free and open-source traffic data generator that creates vehicular and pedestrian mobility scenarios in lieu of empirical data when such data is not readily available or when experimental studies cannot be conducted in a real setting. As such, the SUMO simulation is often used by traffic engineers to evaluate infrastructure performance and traffic policy changes before deploying them on the road. Research groups have also used this simulation platform to evaluate the effectiveness of environmental zoning and optimize traffic light control algorithms. This simulation includes many features for traffic modeling including

microscopic simulation of individual road vehicles, public transport, and pedestrians. The SUMO software suite also comes with basic capabilities such as route finding, traffic visualization, importing road network from OpenStreetMap, and customization of traffic light time schedules which is necessary for assessing the efficiency of any traffic light synchronization schemes. Details of the modules relevant to this project are discussed in the following subsections and an overview flow diagram of these components is shown in Figure 2.1.

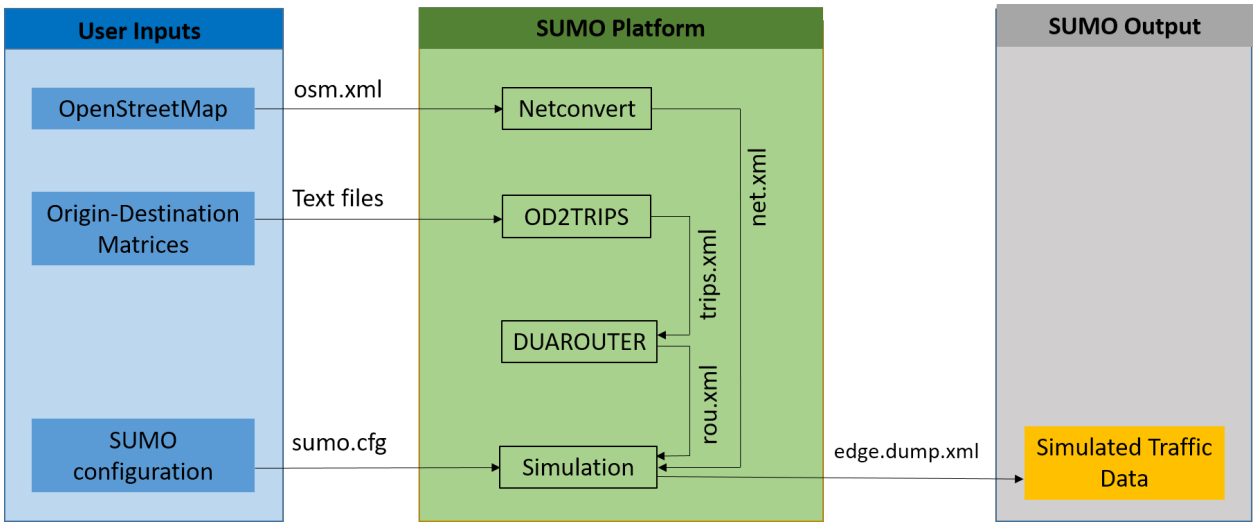


FIGURE 2.1: An overview of the data generation flow of this project. The user input is a set of static files that is read in by SUMO. The SUMO platform includes many functional modules including Netconvert, OD2TRIPS, DUAROUTER, and the simulation engine, which is the actual data generating step for the traffic data used in this study. The user input files provide the specifics of the data scenario generated. The resulting simulated traffic data is then used to build a traffic network, which is the starting point of the experiments conducted in this thesis.

2.2.1 User Inputs

The user input to the SUMO simulation is a set of static text files created by the user according to the specifics of the desired traffic scenario. The specifics of the file format are well documented at SUMO's website located here. The following subsections explain each type of input in detail.

2.2.1.1 OpenStreetMap

Westwood Village was chosen as the area of study for this project. Figure 2.2 shows the street map of the area used in this study. The region spans 34.05809°N to 34.06551°N in latitude and 118.43711°W to 118.45224°W in longitude. Maps exported from OpenStreetMap are in XML format, where road junctions and segments are represented as nodes and edges, respectively, in a directed graph. In addition to the structural connectivity of the road network, every road segment also include features such as number of lanes, geometric shape, street names, street type, segment length, turn lanes, and the speed limit of every lane. After obtaining the XML file for the region of interest (denoted as osm.xml in Figure 2.1) , SUMO's Netconvert is used to convert it to an another format (net.xml in Figure 2.1), which is used in the simulation step in Figure 2.1.

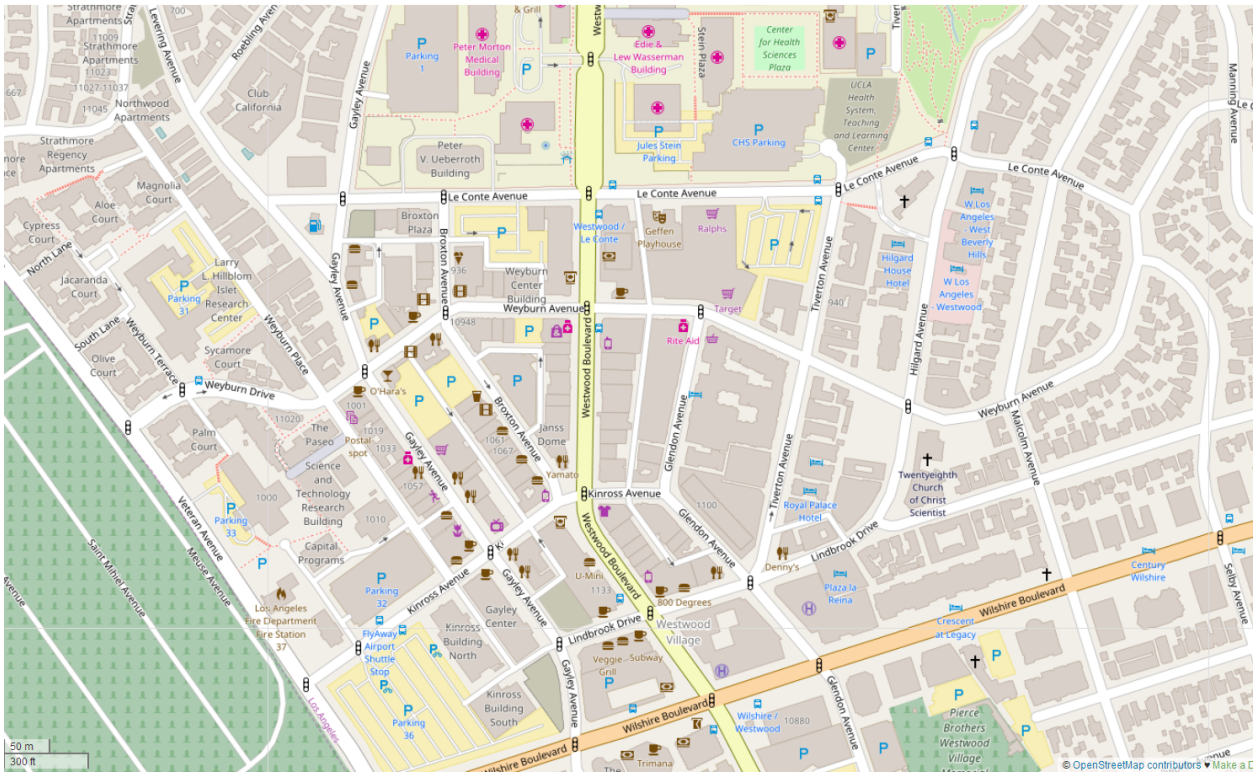


FIGURE 2.2: The area of Westwood Village used in this study as exported from OpenStreetMap.org. The textual representation of the region is expressed as nodes (road junctions) and edges (road segments) in a directed graph in an XML file.

2.2.1.2 Origin-Destination Matrices

In this study, three traffic assignment zones (TAZ) were set up to create the distinct geographical communities in Westwood Village. Directional traffic demand to and from each zone at specific times emulates the rush-hour traffic flow in the simulated data scenario used in this project. Figure 2.3 illustrates the three traffic assignment zones. Street segments in red, purple, and blue were designated as TAZ 1, 2, and 3, respectively. The remaining street segments were assigned to TAZ 4. The traffic demand to and from each TAZ was specified in a set of text files in the format of origin-destination (OD) matrices. Each entry in a matrix is interpreted as the number of vehicles that travel from one TAZ to another within a certain time within a 24-hr period. An example of the content of such a set of matrices is shown in Figure 2.4. The exact file format is documented at SUMO's website located here.

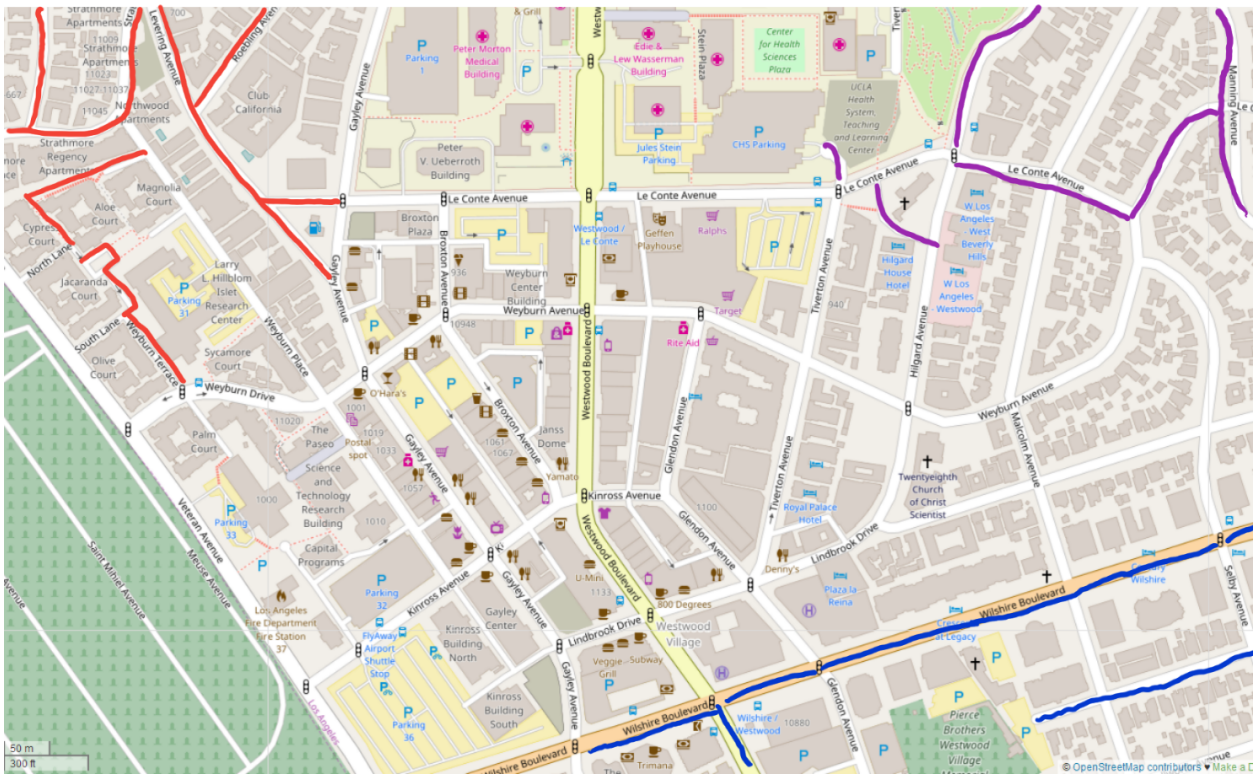


FIGURE 2.3: Westwood Village was divided into three traffic assignment zones (TAZ) in this simulation study. Street segments from TAZ 1, 2, and 3 are colored in red, purple, and blue, respectively. Morning rush-hour traffic primarily consists of vehicles going from TAZ 1 to 2; afternoon rush-hour traffic primarily from TAZ 2 to 3; evening rush-hour traffic primarily from TAZ 3 to 1.

TAZ		Time							
origin	destination	0.00-5.00	5.00-7.00	7.00-9.00	9.00-11.30	11.30-13.30	13.30-16.00	16.00-19.00	19.00-24.00
1	1	0	0	0	1	1	1	1	1
1	2	2	6	60	6	6	3	3	3
1	3	1	4	5	3	3	3	3	3
1	4	3	3	3	2	2	2	2	2
2	1	1	3	3	2	2	2	2	2
2	2	0	2	2	1	1	1	1	1
2	3	1	3	3	5	60	5	6	5
2	4	3	3	3	3	3	3	3	3
3	1	1	2	2	2	2	6	60	6
3	2	2	2	2	2	2	2	2	2
3	3	0	1	1	1	1	1	1	1
3	4	2	4	4	2	2	2	2	2
4	1	0	2	2	2	2	2	2	2
4	2	0	2	2	2	2	2	2	2
4	3	1	2	2	2	2	2	2	2
4	4	3	6	6	4	4	3	3	3
		20	45	100	40	95	40	95	40
total	475								

FIGURE 2.4: The complete Origin-destination specification of the traffic flow to and from TAZ 1-4 in a 24-hr period used in this study. The time period in orange are the rush-hour periods. The entries highlighted in yellow illustrate the increased flow between certain TAZs to simulate rush-hour traffic.

2.2.1.3 SUMO Configuration

The configuration text file, denoted as sumo.cfg in Figure 2.1, contains the location of net.xml and rou.xml files and additional files such as timing cycle definition of the traffic lights, variable speed limit by time (e.g. for simulating accident delay), visualization properties, time interval for macroscopic values calculation for each edge (more discussion in the Simulation subsection below), and other processing parameters pertaining to the data generation process carried out by the Simulation engine in SUMO (discussed in section 2.2.2.4). This configuration file also specifies the type of output file, whether lane- or edge-based data dump file, to be generated by the Simulation. We chose the edge-based data dump at a time interval of 15 minutes as the output for this project.

2.2.2 SUMO Platform

The SUMO platform consists of many modules, of which, only Netconvert, OD2TRIPS, DUAROUTER, and Simulation were utilized in this project. The following subsections discuss the input, output, and settings for these modules used in this project.

2.2.2.1 Netconvert

Netconvert imports road networks from various sources and converts them to a format that can be used within the SUMO platform. It outputs a net.xml file (denoted in Figure 2.1) that can be modified by the user and used in the subsequent data generating process. The parameter settings of Netconvert can be found at SUMO's Wiki page. Typically, this module also generates the junction traffic light programs, right-of-way regulation, and roundabouts priority during the computation of the networks.

Netconvert normally creates the traffic light timing program for each node during the computation of the road network. However, these computed programs are often different from those employed in reality. Users can supply a custom program definition for the traffic lights by using an additional file (see section 2.2.1.3 on SUMO Configuration). Users can also switch the type of traffic light control program according to a pre-specified timeline. The procedural details are available at SUMO's Wiki page . However, for simplicity, the experiment described in Chapter 3 uses the default traffic light programs computed by Netconvert since the timing cycle itself has no bearing on the objective of the experiment.

The SUMO platform also provides a Traffic Control Interface (TraCI) for users to modify the traffic light control online. Future work suggested in Chapter 5 will likely need to extend this capability to make the timing cycle reactive to the traffic dynamics. The details of TraCI are documented at SUMO's Wiki page.

2.2.2.2 OD2TRIPS

After the set of OD matrices are specified, SUMO's OD2TRIPS functionality is invoked to randomly assign a trip for each vehicle specified in the OD matrices, where a trip is defined as an origin-destination pair. OD2TRIPS will insert the number of vehicles to and from each TAZ into the network with random departure times within the time periods indicated by the OD matrices. For this project, we assigned all road segments within a TAZ with the same probability for being chosen as the origin or destination. We also used the default setting where the departure lane is any lane that is free on the road segment, the departure position is the point closest to the start of the departure lane that is available for insertion, and the departure speed is the maximum velocity allowed for the chosen lane. OD2TRIPS outputs a trip file (denoted as trips.xml in Figure 2.1) that records the origin and destination for each vehicle at a specific departure time, lane, and initial speed.

2.2.2.3 DUAROUTER

After OD2TRIPS generates the trip origin and destination for each vehicle, DUAROUTER is then invoked to turn each trip into a route with intermediate road segments assigned. Trips that have unreachable destinations from the origins are discarded. We used the default Dijkstra's algorithm in this study to find the shortest paths between the origin and destination nodes in the road network. The entire route for each vehicle is stored in a static file, denoted as rout.xml in Figure 2.1. Notice that the vehicle routes are pre-computed, therefore, the routes will not change in response to any traffic conditions such as accidents or congestions.

2.2.2.4 Simulation

SUMO's Simulation is the actual data generating process that produces the synthetic data for this study. It inserts vehicles into the road network at the time and location specified in net.xml and rou.xml files from the previous steps and records the macroscopic values for each edge at each time interval such as mean vehicle speed, mean density, travel time, waiting

time, etc. In this study, we only used the mean speed of the edge within the time intervals for the network analysis. The Simulation also includes a visualization tool for users to play back the entire data scenario from the beginning to the end. Figure 2.5 shows a zoomed in area of Westwood Village from SUMO's visualization tool. In this snapshot, three vehicles (represented by yellow triangles) are waiting at red lights (depicted by the red lines). Green lines represent green lights.

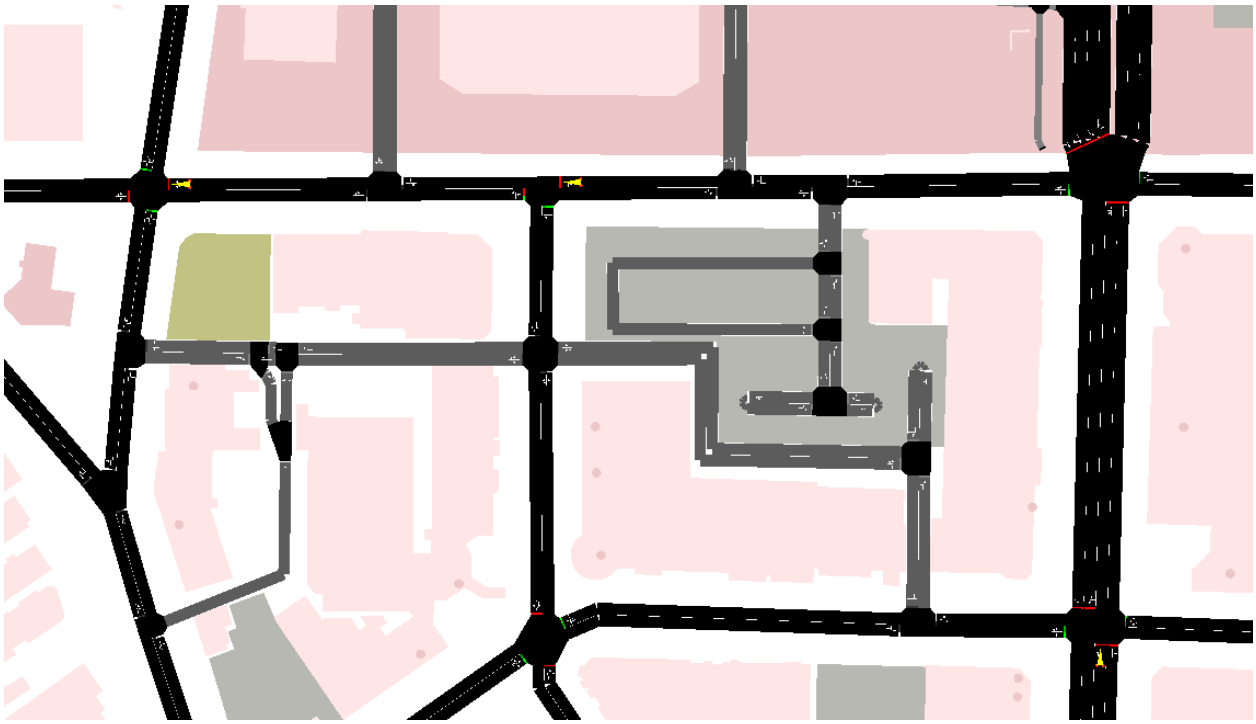


FIGURE 2.5: SUMO's representation of streets and traffic lights. The red and green lines correspond to the red and green signal at the traffic lights, respectively. Vehicles are represented by yellow triangles and the white doted lines represent the lanes of the road segments.

2.2.3 SUMO Output

The SUMO output is the simulated traffic data recorded in edge.dump.data.xml denoted in Figure 2.5 This is the file that contains the synthetic data for the network analysis experiments conducted in this study in lieu of any empirical data.

2.2.3.1 Simulated Traffic Data

This single file contains the aggregated macroscopic data such as mean vehicle speed and mean density at each time interval for all of the road segments. For this study, only the mean speed of the edges was utilized for the experiments described below. The details of the file format are described at SUMO's Wiki page.

CHAPTER 3

Traffic Dynamics Modeled by Percolation Transition

The goal of the experiment in this chapter is to capture the traffic pattern in terms of the time and location of the bottleneck formation in the road network by analyzing the phase transition in the traffic percolation process using the simulated traffic data from Chapter 2. In particular, we focus on understanding the dynamics behind the local traffic clusters that eventually spread to form global traffic congestion. Theoretically, improving the throughput of these critical bottlenecks should lead to improvement of the efficiency of global traffic throughout the road network. The following sections describe the process of the experiment and validate the result. Chapter 5 suggests practical means to achieve such improvement in an urban setting where options such as expanding the road capacity is often infeasible.

3.1 Building a Functional Network from a Structural Network

In order to analyze the efficiency of the traffic flow, we need to convert the structural road network to a functional network first. For this project, the functional measure of traffic flow was based on a relative velocity measure (shown in equation 3.1), which is the ratio between the current velocity of the road segment and the speed limit specified by the XML file downloaded from OpenStreetMap. For each edge from node i to j, the relative velocity is defined as:

$$r_{ij}(t) = \frac{vel_{e_{ij}}(t)}{vel_{e_{ij_{max}}}} \tag{3.1}$$

where $vel_{e_{ij_{max}}}$ is the speed limit for edge e_{ij} and $vel_{e_{ij}}(t)$ is the average speed on edge e_{ij} during the time interval t

After obtaining the traffic flow measure for each edge in the road network, a dynamical traffic network model was then built to identify the traffic pattern of the commuters in the region. In this study, we followed the algorithm as described in the traffic study by [LFW15], where a percolation process was used to model the traffic pattern for different days in the city of Beijing. Their traffic model was also based on the relative car velocity on each road segment in the network. However, in the event that the speed limit data does not exist, they used the ratio between the current velocity and the 95th percentile of its historical velocity for that day of the week. Figure 3.1 shows the graphical representation of the full network built from the street map of Westwood Village for this project. The red edges have relative velocity of less than or equal to 0.25, yellow edges have relatively velocity larger than 0.25 and less than or equal to 0.75, green edges have relative velocity greater than 0.75.

In this graph, the configuration reflects the directional structural connectivity of the streets. In order to convert this to a functional traffic network, for a given threshold q, the following modification is made:

$$e_{ij}(t) = \begin{cases} 1 & \text{if } r_{ij}(t) \geq q \quad (e_{ij} \text{ at time } t \text{ is left intact}) \\ 0 & \text{if } r_{ij}(t) < q \quad (e_{ij} \text{ at time } t \text{ is removed}) \end{cases}$$

In other words, at time t, if the relative velocity drops below a certain threshold, then the edge that represents this road segment dissolves. Otherwise, the edge exists to connect the two intersections. Since the road network is a directed graph, a connected component is defined as a strongly connected component if all pairs of nodes are mutually reachable along a directed path. The physical meaning of the quantity q is the measure of global efficiency of the traffic flow in the network a vehicle can only travel at or below q within the cluster. Therefore, a large q value associated with a very large connected component that consists of most of the nodes in the network is the optimal efficiency (i.e. no traffic congestion); whereas a large q value with small isolated clusters means that a vehicle is trapped within a small area if it has to operate at that high speed (i.e. free of congestion only within a small

confined area).

3.2 The Application of Percolation Theory in Identifying Traffic Pattern and Bottlenecks

Percolation is one of the simplest models in probability theory that exhibits critical phenomena, where there is usually a natural parameter in the model at which the behavior of the system drastically changes. Percolation theory has a wide range of applications in which the major problems are easily stated, but whose solutions are not immediately obvious. In modeling traffic dynamics, the phenomenon that [LFW15] observed was that there is a critical phase transition between isolated local flows (i.e. when some edges dissolve causing the road network to break down into isolated clusters) and global flows (i.e. when all the nodes are connected as in the structural road network). This critical transition happens when the second largest cluster reaches its maximum size, at which the q value is denoted as the critical threshold, qc. As such, q_c quantifies the organizational efficiency that reflects the maximal relative velocity one can travel over the main part of a network. In their study, q_c was observed to change dramatically during the day coinciding with the rush hours, and that the pattern for the weekdays are different from that for the weekends. They define the bottleneck links as the edges that break the network into clusters at criticality; these links change during the day with different links disappearing in different hours of the day, but the pattern persists across days. Therefore, the pattern found may represent the normal traffic pattern of the region due to the commuting habits of the inhabitants. In addition, they found that these bottlenecks are different from structural bottleneck links found by traditional network analysis because they evolve with time according to the traffic dynamics. In fact, a rather significant finding from this study is that the bottlenecks identified are different from those with the highest degree centrality (i.e. the ones that bridge different topological communities). Their study showed that increasing the velocity of such highly connected links does not improve the global traffic network, whereas increasing the velocity of the bottlenecks identified in the phase transition shows significant improvement (increase)

of q_c.

3.3 Finding Percolation Transitions Using Simulated Traffic Flow

In this part of the project, we reproduced the findings of the study by [LFW15] by using their methods on our simulated traffic data generated by the process described in Chapter 2. Statistical analysis was then used to validate our result in section 3.4. As indicated in Figure 2.4, morning rush-hour is 7am-9am, noon rush-hour is 11:30am-1:30pm, and evening rush-hour is 4pm-7pm. Per the OD matrices, 60 vehicles were inserted into the road network going from TAZ 1 to 2 from 7am to 9am; , 60 vehicles were inserted going from TAZ 2 to 3 from 11:30am to 1:30pm; another 60 vehicles were inserted going from TAZ 3 to 1 from 4pm to 7pm. The rest of the entries, with significantly smaller values, serve as the background level of traffic flow assigned to each incoming and outgoing TAZ. As indicated in section 2.2.1.3, the mean speed of each edge is computed at 15-min time intervals.

First, a directed graph was built from the structural street connectivity as specified by the net.xml file denoted in Figure 2.1. This graph provides the full network configuration. Then the relative speed, r_{ij}, of the edge connecting node i and j, is computed from the edge.dump.xml file shown in Figure 2.1. Then for each time interval, a functional network was built from the full structural network using the relative speed measure discussed in section 3.1. This is repeated at each time step in the edge-dump file until the end of the scenario. The sizes of the largest and the second largest clusters were plotted for visualization of the critical phase transition during morning, noon, and evening rush-hour periods where congestions occur. Figure 3.2-3.7 show such examples of the cluster sizes and the 3 largest clusters formed at the corresponding q_c values, using a single realization of the data generation process described in Chapter 2.

For this method, according to percolation theory, the phase transition of the network from a non-congested phase to a congested phase is demarcated by the critical threshold, q_c, where the second largest cluster reaches its maximal size; examples of which are shown in Figure 3.2 when $q = 0.83$, in Figure 3.4 when $q = 0.78$, and in Figure 3.6 when $q = 0.68$. As

illustrated in these examples, as the q value increases, the giant component is fragmented into pieces. At the same time, the second largest cluster reaches the maximum size at q_c, where the network transitions from the connected state to the fragmented state. Therefore, intuitively, q_c quantifies the global efficiency of the traffic flow in the network: the higher the q_c, the higher the velocity one can travel over the giant connected component of the network (more explanation in section 3.2). Figure 3.8 shows a line plot of q_c, averaged over 10 realizations of the same traffic scenario described in Chapter 2 for a 24-hr duration. In this plot, q_c values were obtained only when the second largest clusters contained at least 5% of the nodes in the network. This plot illustrates that the local minima of q_c during the 24-hr scenario approximately coincide with the morning, noon, and evening rush-hour periods. The minima occurring between morning and noon hours (around 10:30am and 11:00am) are mostly due to the outliers caused by random sampling of the vehicle origin-destination assignment and time of departure of some simulation realizations (see details in the boxplot representation in Figure 3.13).

In percolation theory, the network at criticality behaves as the backbone of the original network [LFW15], at which some links (the red bonds) play a critical role in bridging various connected components of the network. Therefore, by posing the traffic dynamics problem as a percolation process, these bridging links are considered bottlenecks since their velocities are the lowest compared to the backbone piece, and q_c is the threshold value at which these links break down. As a result, the road segments deemed to be bottlenecks are exactly the edges that were dissolved at q_c that cause the largest cluster to break apart to form the second largest cluster. Figure 3.9 shows the frequency of occurrence of the bottleneck links that disintegrate the backbone giant connected component and result in the second largest cluster reaching its maximal size. These links are obtained by comparing the functional network just below and immediately above q_c. Figure 3.10 shows their geographical locations. The occurrence of these bottlenecks are counted within the morning, noon, and evening rush-hour periods separately to generate the histogram. For simplicity and visualization purposes, street segments are grouped together in Figure 3.9 if they share the same street name. However, different segments of the same street often have different frequency distributions

for morning, noon, and evening rush hours. The same frequency plots for the individual edges are shown in Appendix **??**. Only a single realization of the data generating process described in Chapter 2 was used to generate the plots in Figure 3.9 and in Appendix **??**.

3.3.1 Discussion of the Bottlenecks Identified During Rush-Hour Periods

From the overall distribution shown in Figure 3.9, Le Conte Ave, Westwood Blvd, and Wilshire Blvd are the most frequent bottlenecks that arise from the TAZ setup in the rush-hour traffic scenario. This makes sense since both Westwood Blvd and Wilshire Blvd are favored by SUMOs routing algorithm because they serve as the main arterial roads in the area with higher speed limits and more lanes. For non-arterial roads, Le Conte Ave serves as the more direct path that connects TAZ 1 (red) and TAZ 2 (purple) for morning commute. However, as shown in Figure 3.10, certain segments of Le Conte Ave are also frequently used during noon rush hour for traffic leaving TAZ 2 (purple) for TAZ 3 (blue) via Westwood Blvd heading south. Other segments of Le Conte Ave also experience substantial demand from evening traffic leaving TAZ 3(blue) for TAZ 1 (red) via Westwood Blvd. Weyburn Ave is an alternate route that connects TAZ 1 (red) to TAZ 2 (purple), but it may be less traveled than Le Conte Ave because it is to the south of both zones, which would result in a longer path for most origin-destination pairs within the two zones. For similar reason, Weyburn Ave is much less traveled in the evening than in the morning since Weyburn Place offers a better route in combination with arterial roads such as Wilshire Blvd. In addition, there is more traffic demand for north/south bound streets than east/west bound streets during noon and evening rush-hour periods due to the travel pattern defined in the traffic scenario. For example, Gayley Ave and the southern segments of Westwood Blvd both experience higher demand in the evening compared to the morning period. Overall, the bottleneck links identified are consistent with the construction of the simulated traffic data according to the OD matrices indicated in Figure 2.4.

3.4 Significance of the Rush-Hour Bottleneck Distributions Compared To Randomness and/or Extraneous Factors

Although the bottleneck links identified using this method seem to be consistent with the TAZ simulated traffic data scenario that was set up in Chapter 2, this section performs the statistical analysis to validate our result against a reference null distribution. The analysis conducted in this section uses the bottleneck frequency distribution for the individual segments without the aggregation by street name as in Figure 3.9. Using pseudo random seeds, 10 realizations of the simulated TAZ traffic scenario constructed as described in Chapter 2 were used and 10 realization of the reference null distribution scenario were used for comparison.

3.4.1 Synthetic Data Generation for the Reference Null Distribution

In order to verify that these links are the result of the simulated rush-hour traffic (instead of any external factors such as any bias introduced by the routing algorithm), the bottleneck distribution generated by the 24-hr scenario with morning, noon, and evening rush-hour periods is compared with that from a 24-hr scenario with uniform traffic flow where there is an equal number of vehicles going to and from each TAZ, with the same number of vehicles in total as the former scenario. The latter scenario is our reference null distribution used for the rest of this chapter.

3.4.2 Bottleneck Distributions Similarity Measure

In this experiment, the similarity measure between two bottleneck distributions, A and B, is the Total Variation (TV) norm calculated as:

$$TV_{norm_{ij}} = \frac{1}{2} \sum_{t \in T} \sum_{e \in E(g)} |A_{t_i}(e) - B_{t_j}(e)| \tag{3.2}$$

where T is the {morning, noon, evening} time periods defined in Section 3.3

$E(g)$ is the set of edges of the structural road network graph g (i.e. all the road segments in the network)

i and j are the indices of realizations of a data scenario. For this experiment, $i, j = 1, 2, 3, , 10$.

$A_{t_i}(e)$ is the frequency count of edge e occurring as a bottleneck (as defined in Section 3.3) during the time period t of realization i from distribution A. Similarly for $B_{t_i}(e)$.

3.4.3 Within- and Between- Group Comparison of the Similarity Measure

In this subsection, we will visualize the effect randomness has on the bottleneck distributions for both the uniform traffic flow and rush-hour traffic scenarios. The randomness under consideration includes the random selection of the departure and arrival edges and the departure time according to the OD matrices during trip generation (detailed discussion in Sections 2.2.1.2 and 2.2.2.2).

First, we will look at the within-group comparison for the two scenarios. Figure 3.11 shows the boxplot of the TV-norm values for the 10 realizations grouped by the scenario type.

The TV-norm value for each realization of the rush-hour traffic scenario is calculated as:

$$TV_{norm_{R_{ij}}} = \frac{1}{2} \sum_{t \in T} \sum_{e \in E(g)} |R_{t_i}(e) - R_{t_j}(e)| \tag{3.3}$$

where $R_{t_i}(e)$ is the frequency count of edge e occurring as a bottleneck during the time period t of realization i of the rush-hour traffic scenario.

Similarly, the TV-norm value for each realization of the uniform traffic scenario is calculated as:

$$TV_{norm_{U_{ij}}} = \frac{1}{2} \sum_{t \in T} \sum_{e \in E(g)} |U_{t_i}(e) - U_{t_j}(e)| \tag{3.4}$$

where $U_{t_i}(e)$ is the frequency count of edge e occurring as a bottleneck during the time period t of realization i of the uniform traffic scenario.

The result in Figure 3.11 shows that within-group difference is smaller for the rush-hour scenario than that of the uniform scenario, and that the uniform scenario generally has larger variance and higher mean for the TV-norm. In other words, the bottleneck distributions from the realizations of the rush-hour scenario are more similar than those from the uniform traffic flow scenario. This observation is consistent with the construction of the rush-hour traffic scenario since the OD matrices dictate the direction of the majority traffic flow; on the other hand, the uniform traffic flow scenario has a much larger range of selection in the origin/destination and the departure time for each vehicle.

To confirm that the TV-norm is larger between the rush-hour scenario group and the uniform traffic flow scenario group, the TV-norm between each realization of the two scenarios is computed as follows and plotted in Figure 3.12:

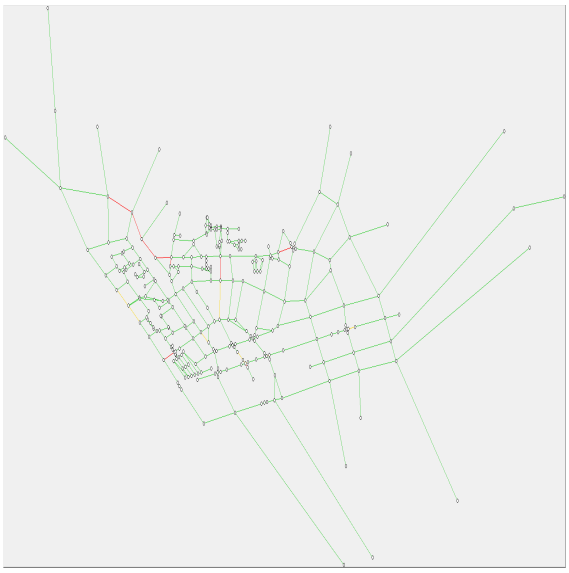
$$TV_{norm_{ij}} = \frac{1}{2} \sum_{t \in T} \sum_{e \in E(g)} |R_{t_i}(e) - U_{t_j}(e)| \tag{3.5}$$

This result shows that the between-group difference is much higher than both of the within-group differences. As such, the links identified as the bottlenecks of the traffic network during rush-hour periods for each realization of the rush-hour scenario are significantly different from their counterparts in the uniform traffic flow scenario. Therefore, these bottlenecks are largely attributed to the majority traffic flow to and from the TAZs specified by the OD matrices instead of any external factors that contribute to randomness.

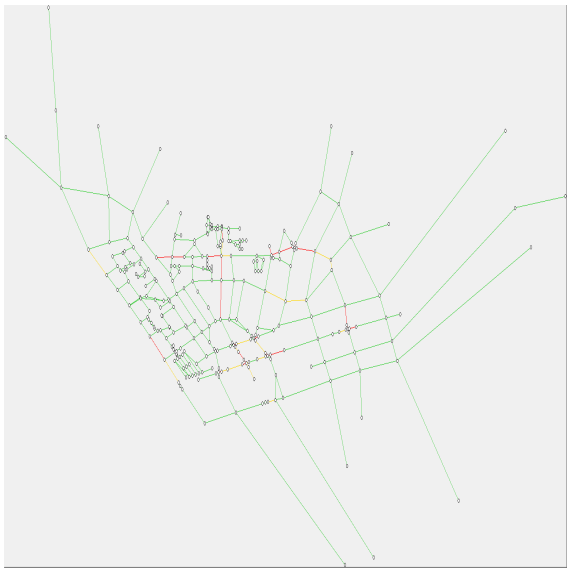
3.5 Validation of Modeling Traffic Dynamics as a Percolation Process

To further validate the application of percolation theory to model traffic dynamics, the critical threshold (q_c) values generated from the 10 realizations of the rush-hour scenario

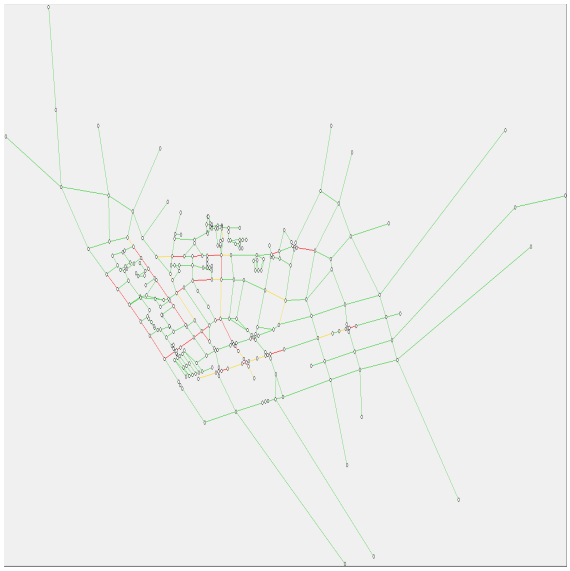
are compared with those from the 10 realizations of the uniform traffic flow scenario at a 15-min time interval for a 24-hour period. As required previously in Figure 3.8, the q_c values here were also obtained only when the second largest clusters contained at least 5% of the nodes in the network. Figure 3.13 illustrates such data comparison as boxplots, where the rush-hour scenario shows three clusters of q_c values which coincide with the time periods for the morning, noon, and evening rush hours. As expected, due to the construction of the rush-hour traffic scenario, the q_c values within these clusters are generally lower. In contrast, the boxplot for the uniform scenario lacks the distinctive clustering pattern. This result shows that there are some noticeable changes in the traffic dynamics during the rush-hour periods where the traffic network changes from a non-congested state to a congested state. Specifically, the disintegration of the giant connected component taking place during the phase transitions signifies the presence of a congestion happening in the bottlenecks (see Section 3.2). This result gives additional evidence to support the use of percolation theory in detecting phase transitions to and from non-congested and congested states.



(A) 9:15am



(B) 1:00pm



(C) 6:15pm

FIGURE 3.1: The structural road network of Westwood Village at morning, noon, and evening rush-hour periods using simulated data. Edges are colored according to their relative velocity: below 0.25 (red), between 0.25 and 0.75 (yellow), and above 0.75 (green).

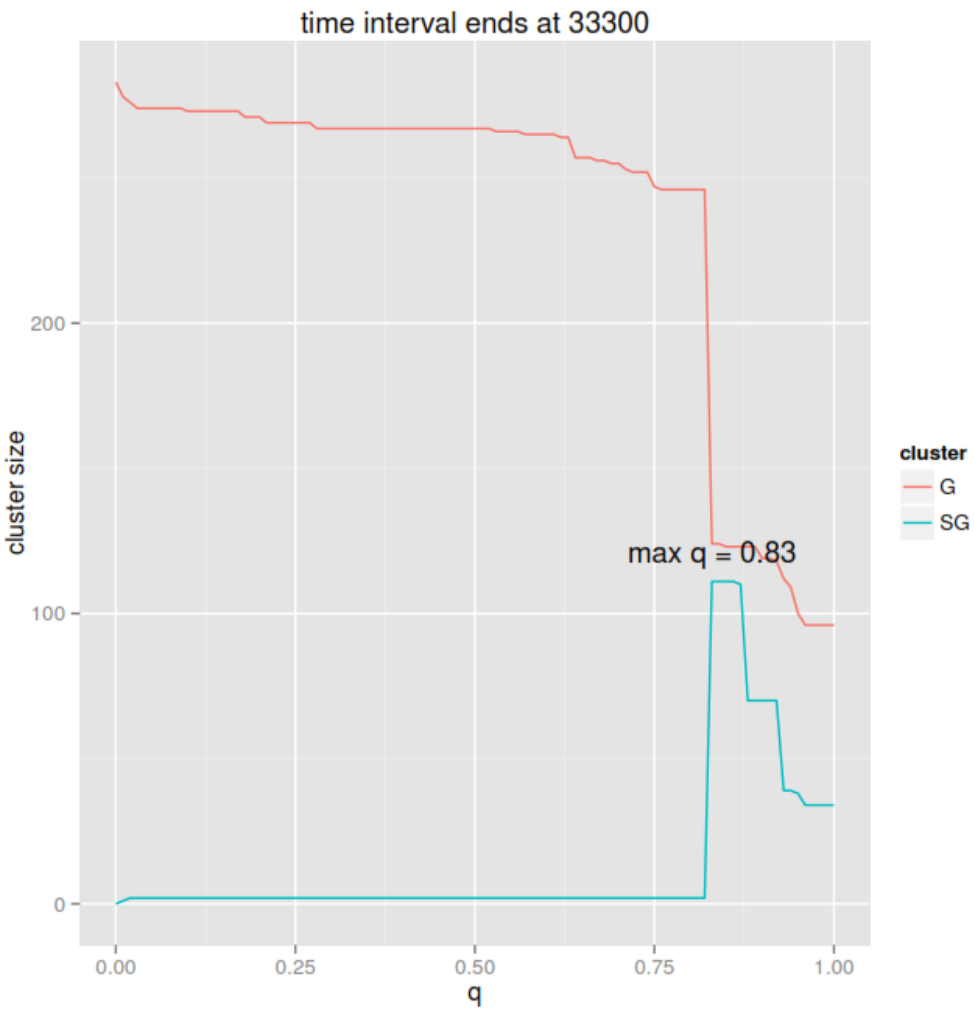
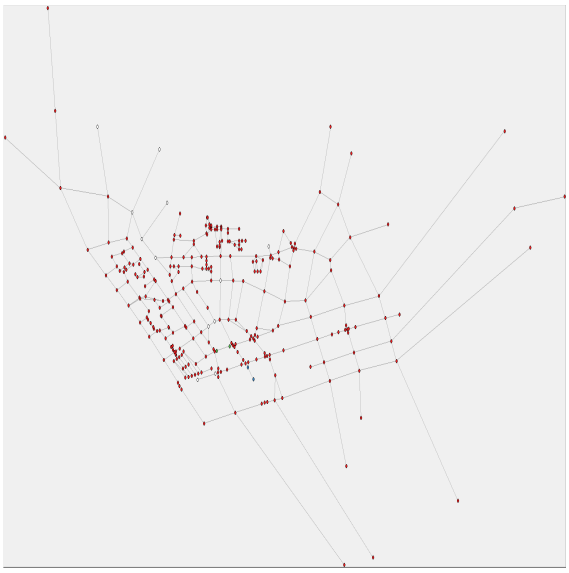
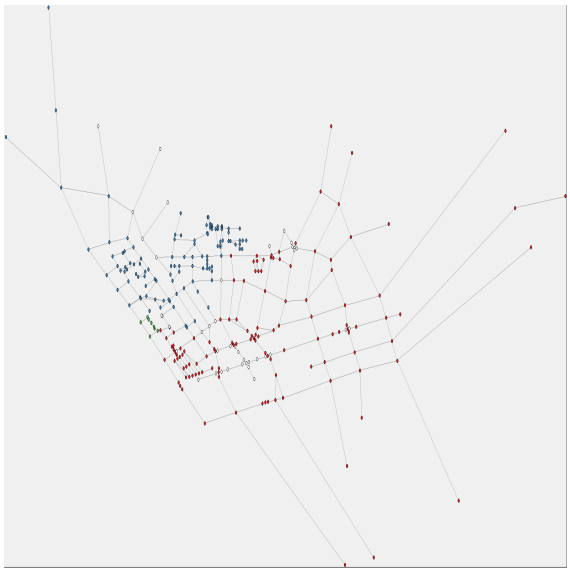


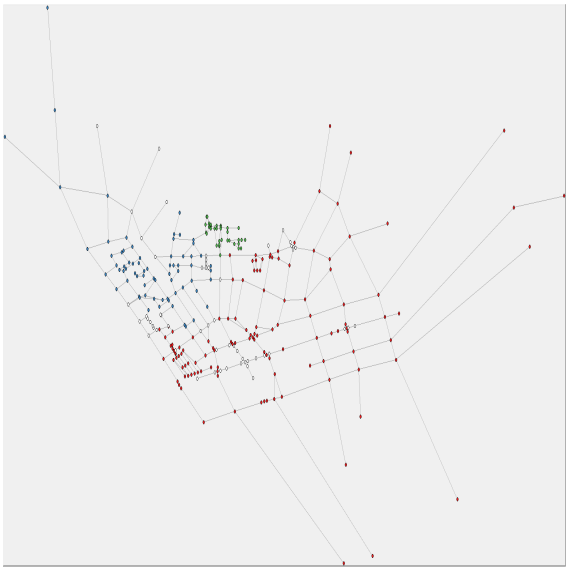
FIGURE 3.2: Size of the largest (G) and the second largest (SG) clusters as a function of q at 9:15am during morning rush-hour.



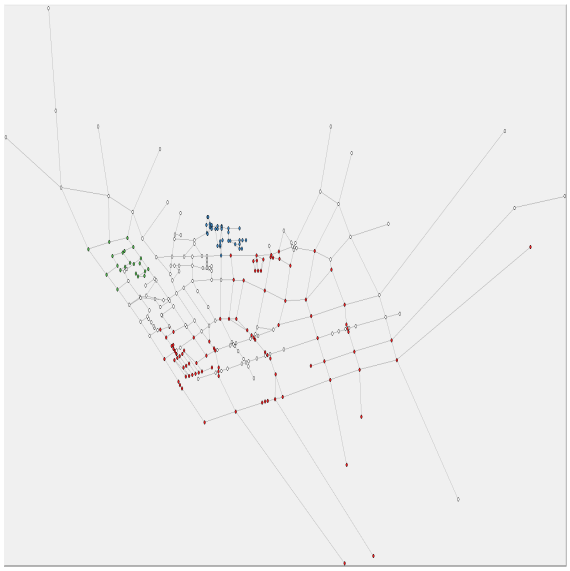
(A) $q = 0.5$



(B) $q = 0.83$



(C) $q = 0.88$



(D) $q = 0.96$

FIGURE 3.3: The network configuration for the corresponding q values at 9:15am. For simplicity, only the 3 largest clusters are shown: the largest (red), the second largest (blue), and the third largest (green). Lower velocity thresholds allow for increased connectivity of the network.

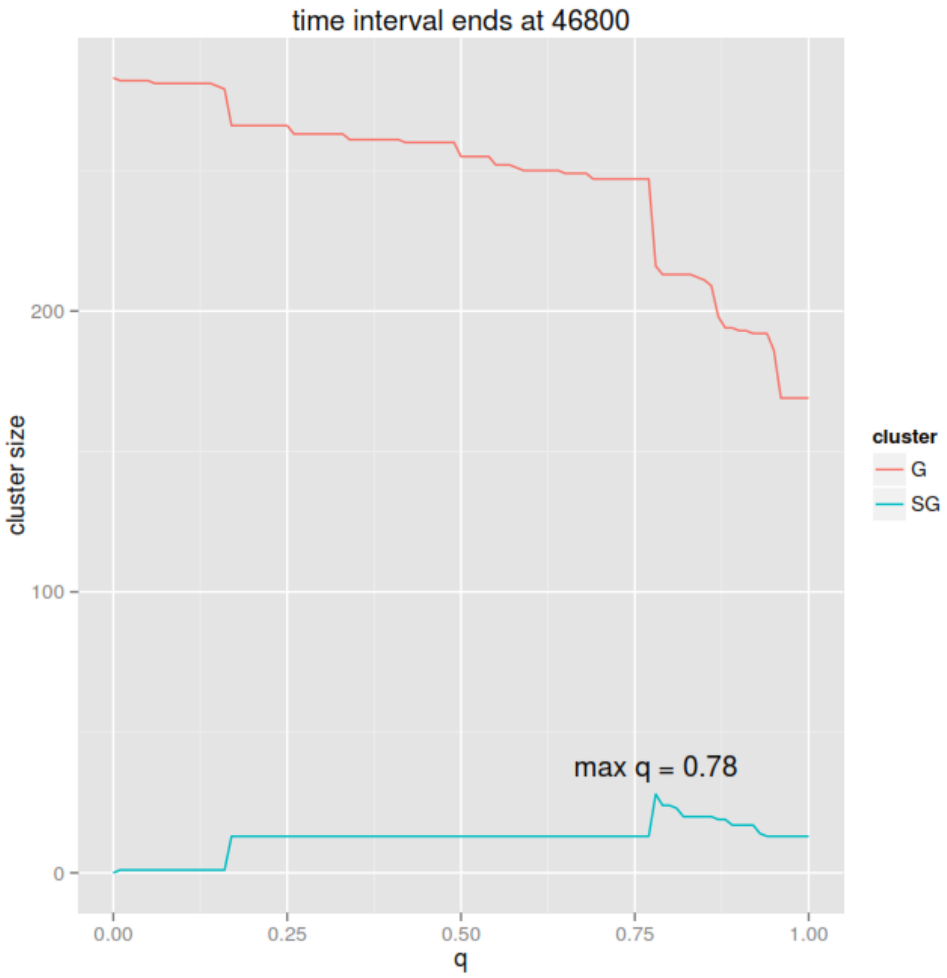
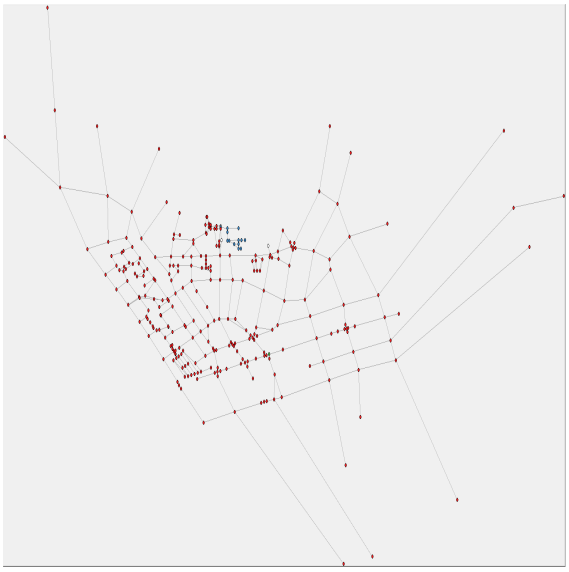
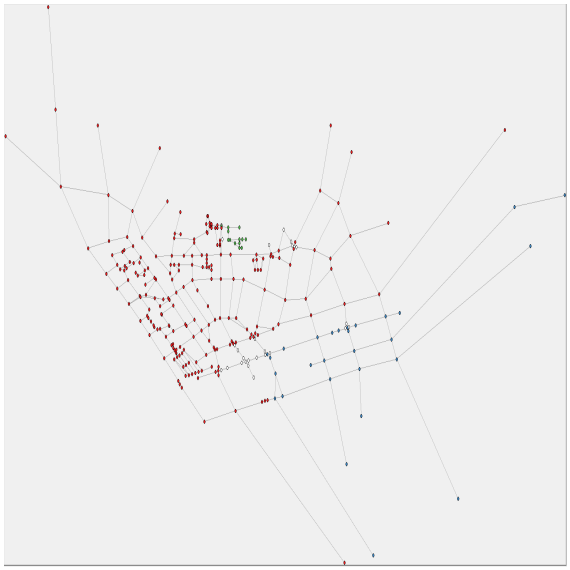


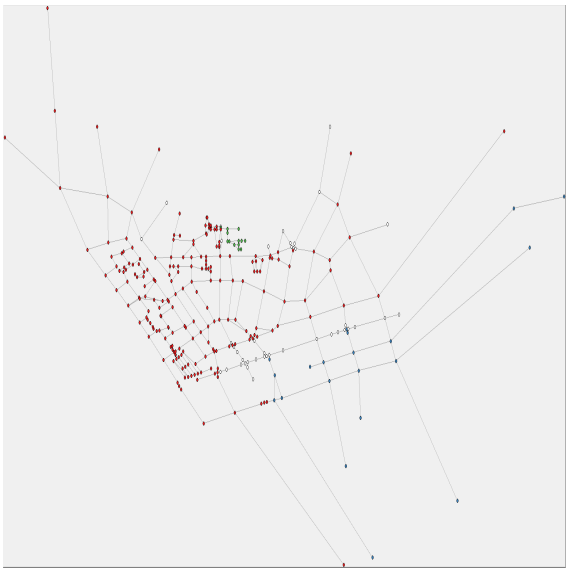
FIGURE 3.4: Size of the largest (G) and the second largest (SG) clusters as a function of q at 1:00pm during noon rush-hour.



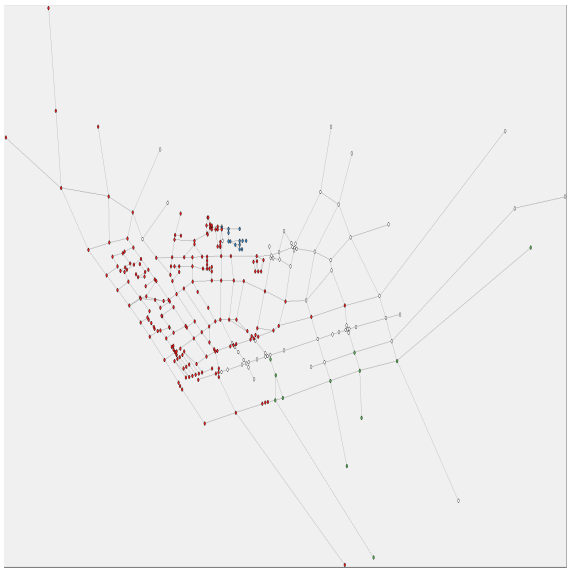
(A) $q = 0.17$



(B) $q = 0.78$



(C) $q = 0.86$



(D) $q = 0.94$

FIGURE 3.5: The network configuration for the corresponding q values at 1:00pm. For simplicity, only the 3 largest clusters are shown: the largest (red), the second largest (blue), and the third largest (green).

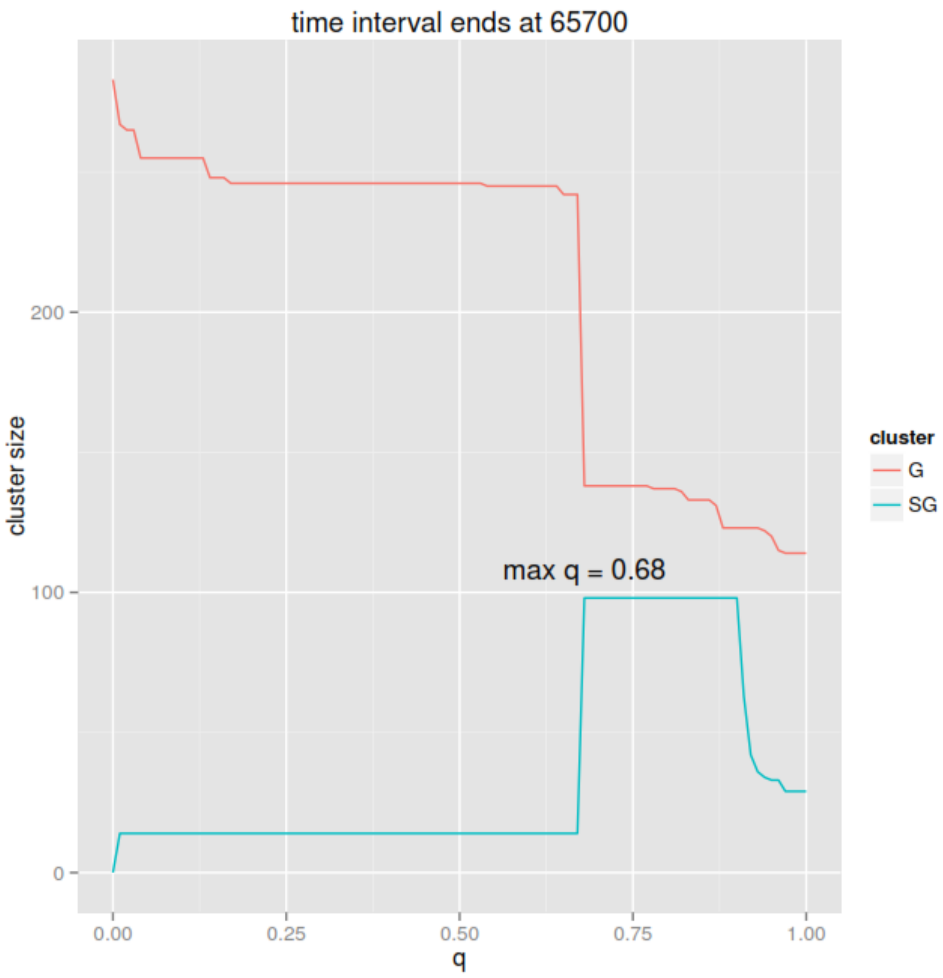
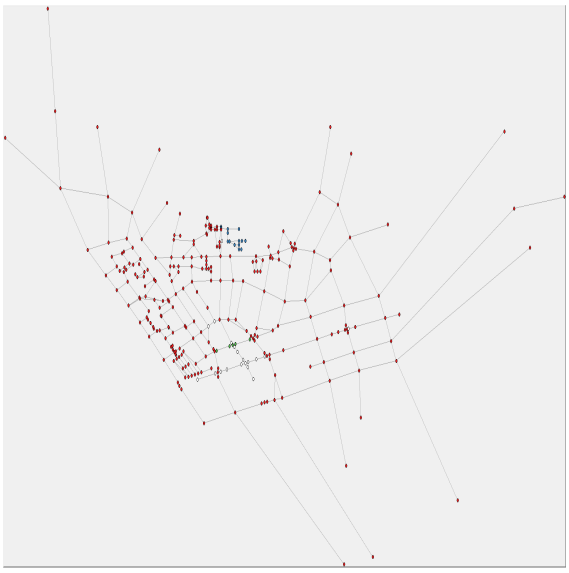
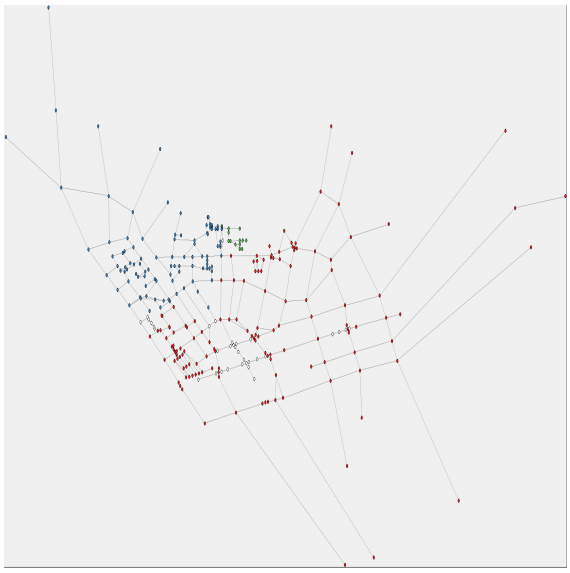


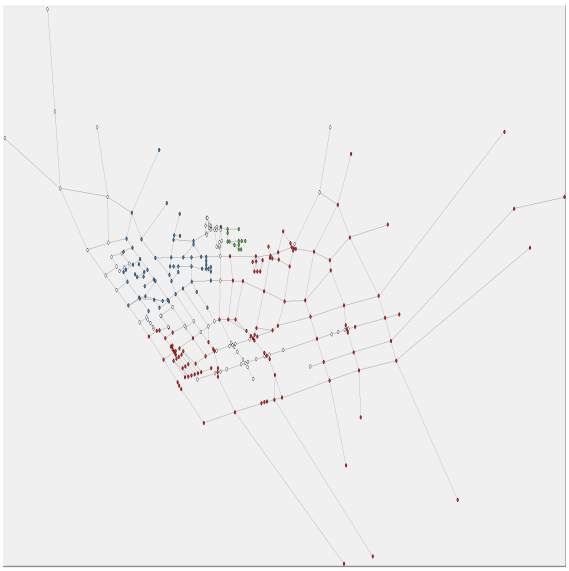
FIGURE 3.6: Size of the largest (G) and the second largest (SG) clusters as a function of q at 6:15pm during evening rush-hour.



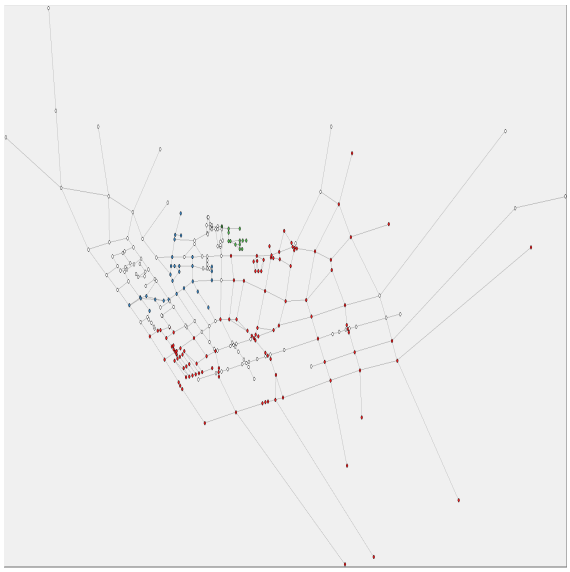
(A) $q = 0.6$



(B) $q = 0.68$



(C) $q = 0.91$



(D) $q = 0.97$

FIGURE 3.7: The network configuration for the corresponding q values at 6:15pm. For simplicity, only the 3 largest clusters are shown: the largest (red), the second largest (blue), and the third largest (green).

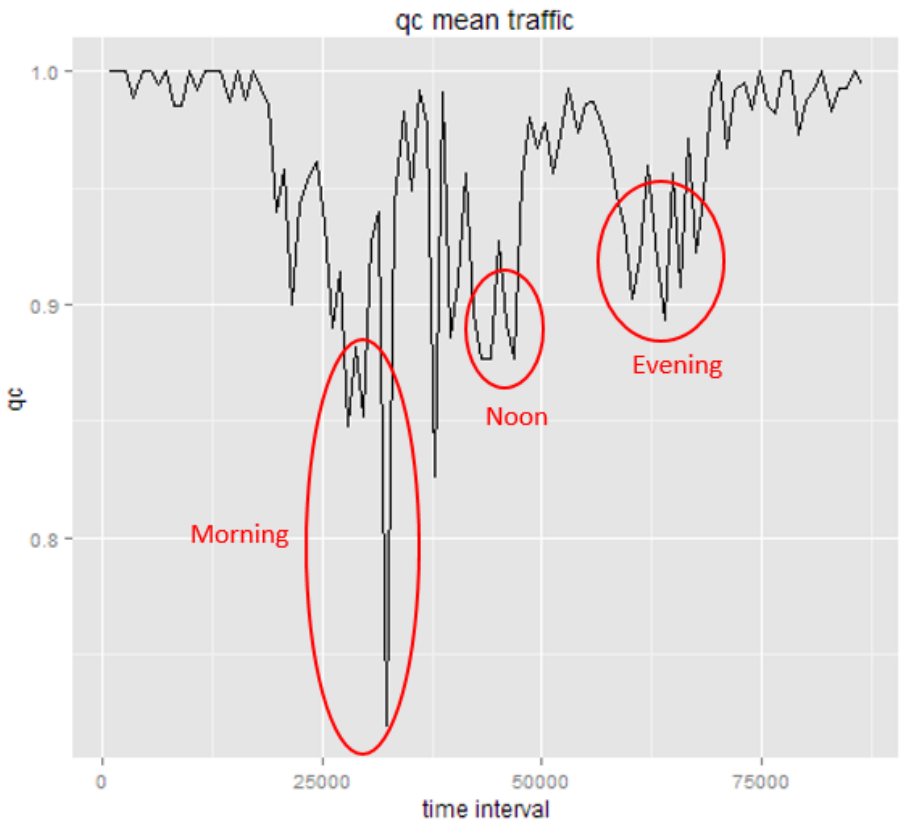


FIGURE 3.8: q_c as a function of time (sec) in the 24-hr rush-hour scenario featuring morning (7am-9am), noon (11:30am-1:30pm), and evening (4pm-7pm) rush-hour periods. The local minima approximately correspond to the time intervals for the rush-hour periods.

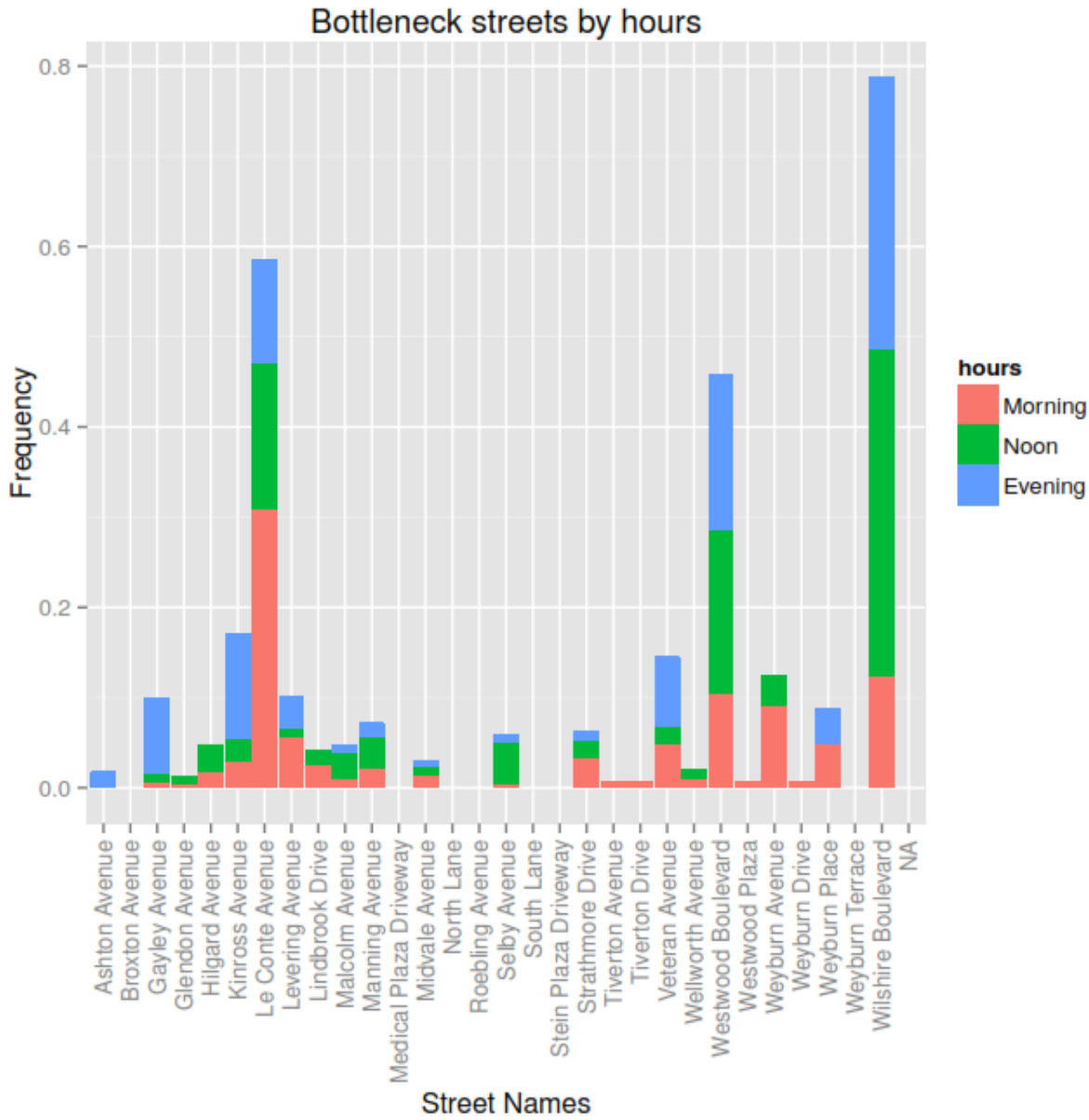
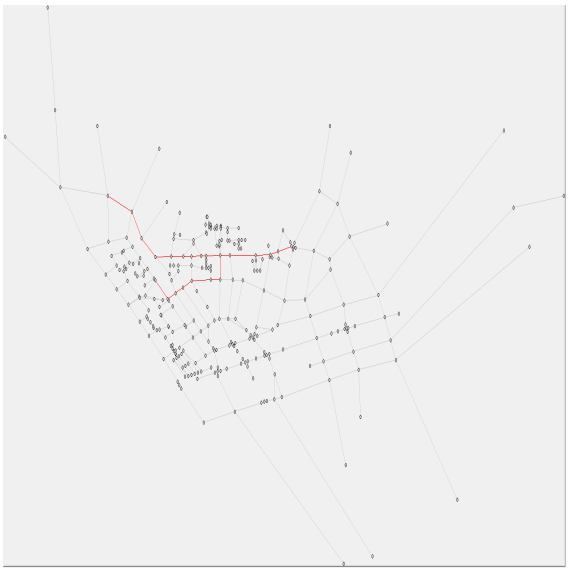
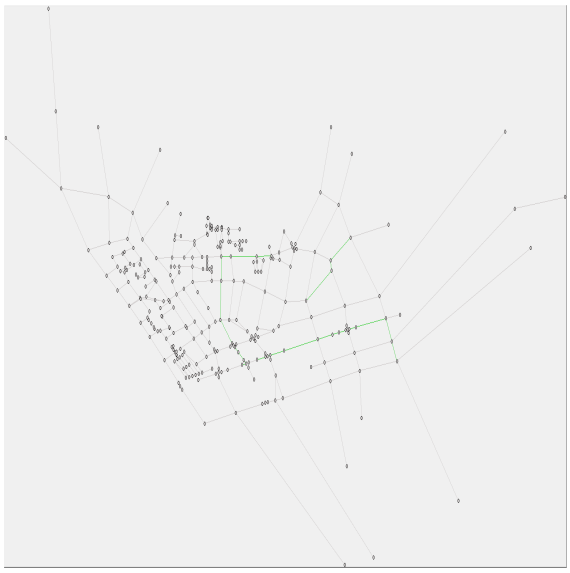


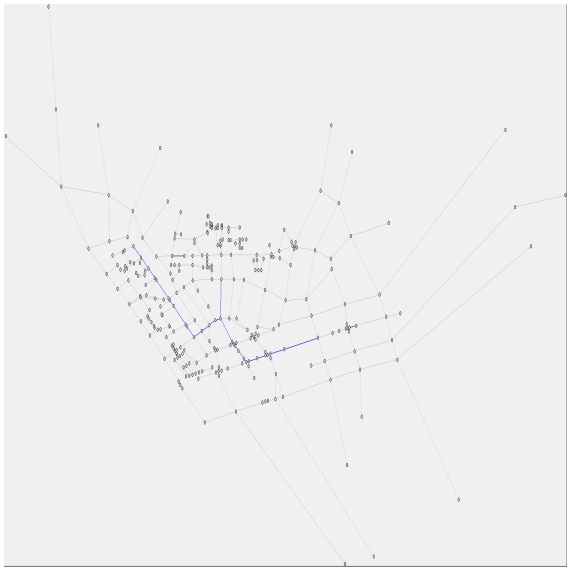
FIGURE 3.9: The evolving bottlenecks identified by percolation transition during morning, noon, and evening rush-hour periods. Multiple street segments that belong to the same street are aggregated for visualization purposes.



(A) 9:15am



(B) 1:00pm



(C) 6:15pm

FIGURE 3.10: Bottleneck links with high occurrence in each rush-hour period. These links are the critical links that cause the fragmentation of the network during the beginning of traffic formation. Only links with the top 3 highest frequency counts are shown for each rush-hour period.

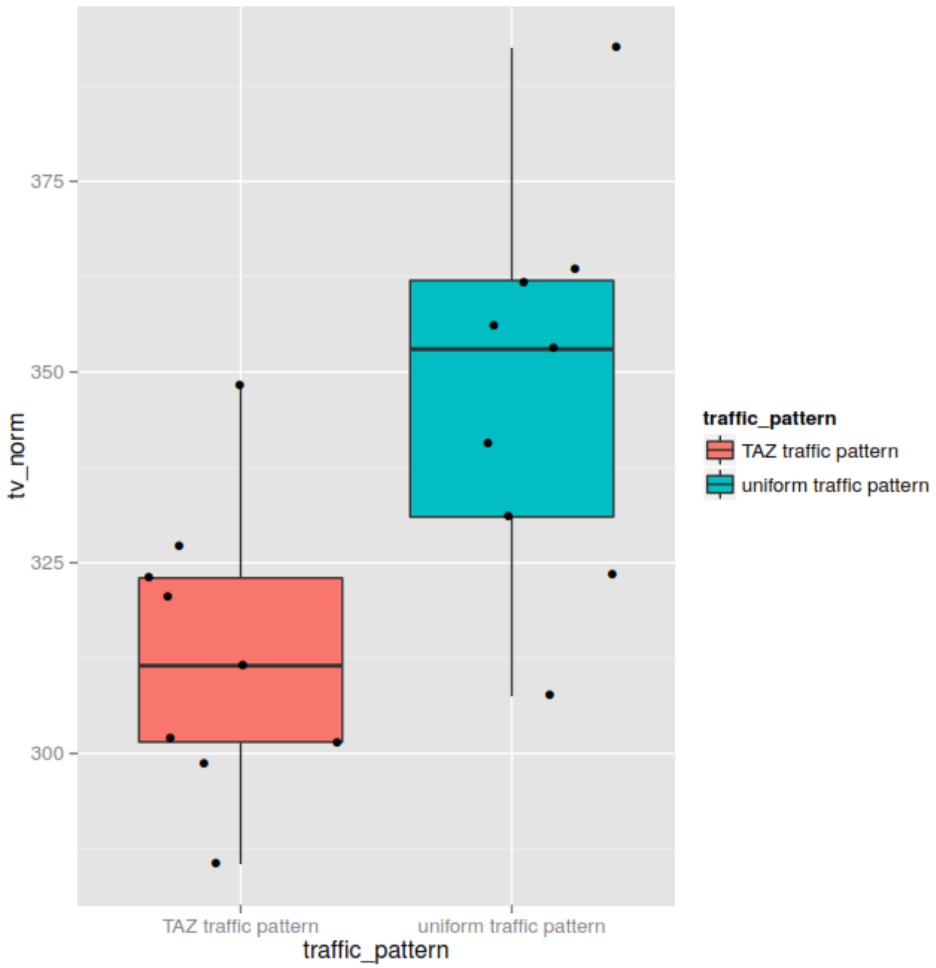


FIGURE 3.11: The TV-norm for the 10 realizations of rush-hour scenario and the 10 realizations of the uniform traffic flow scenario. Within-group difference is smaller for the rush-hour scenario, and the uniform traffic flow scenario generally has larger variance and higher mean for the TV-norm. This result is consistent with the construction of the two scenario types.

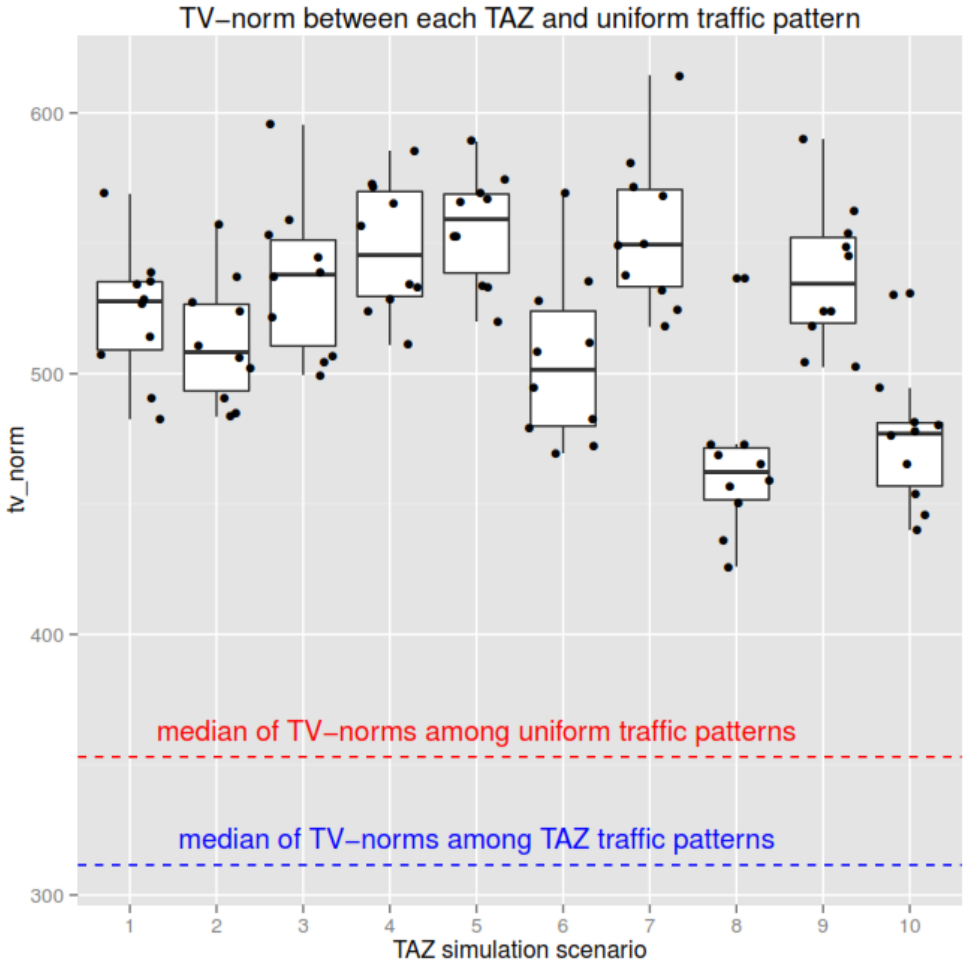


FIGURE 3.12: The TV-norm value between each realization of the rush-hour scenario and those of the uniform traffic flow scenario is much larger than the median values of the within-group TV-norm for both groups (replotted from Figure 3.11 as the red and blue dotted lines).

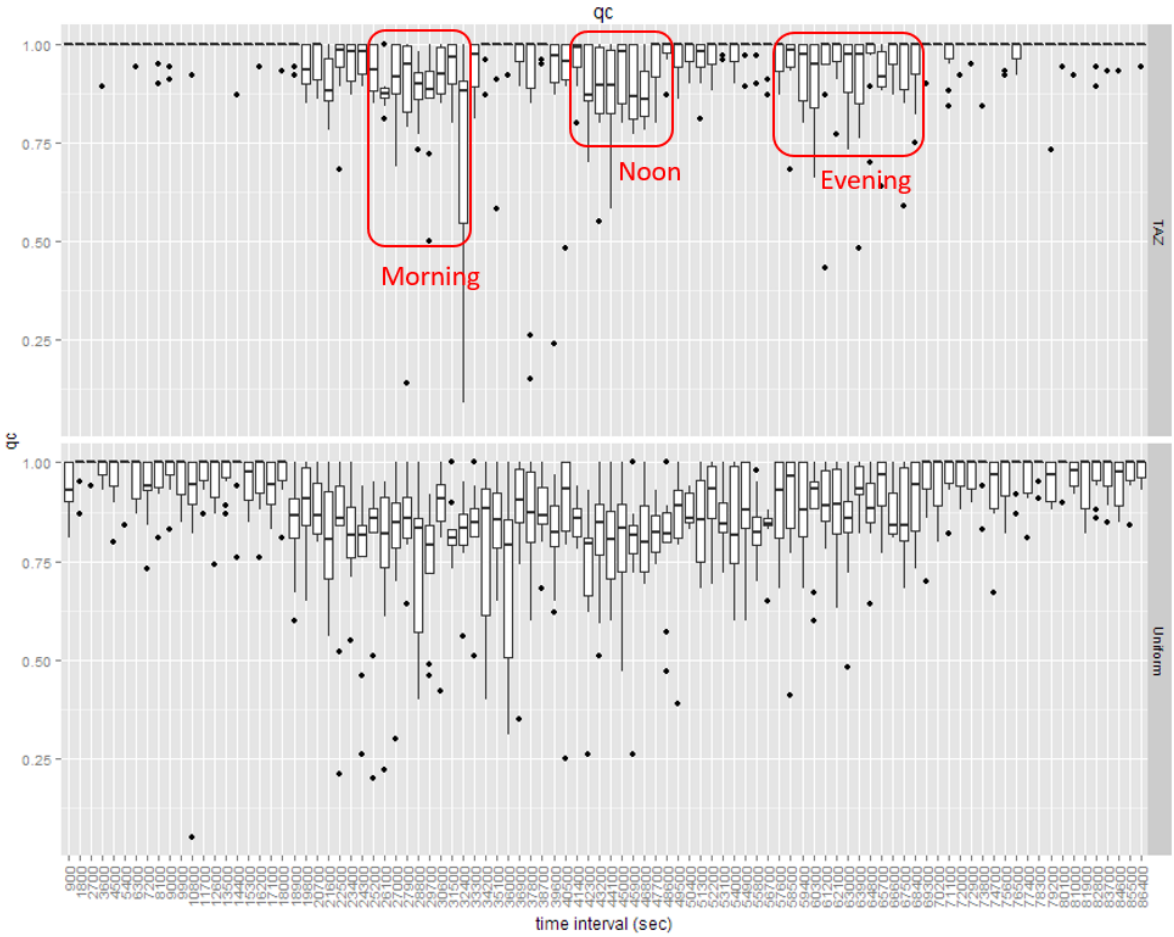


FIGURE 3.13: The boxplot comparison of critical threshold (q_c) values from the 10 realizations of rush-hour scenario and the 10 realizations of the uniform scenario by 15-min time intervals. In comparison with the uniform scenario (bottom), the rush-hour scenario (top) exhibits a clustering pattern around the morning, noon, and evening rush-hour periods. This supports the idea of using critical threshold to signify phase transition from a non-congested to a congested state and vice versa in modeling traffic dynamics.

CHAPTER 4

Conclusion

4.1 Summary

The experiment conducted in this chapter shows promising results for modeling traffic dynamics as a percolation process. The road segments that serve as the critical links that bridge the different functional clusters of traffic are identified as the bottlenecks through the percolation process. Our experiment using simulated rush-hour traffic data gives supportive evidence that these bottlenecks identified are not caused by randomness, but rather by the traffic demand generated by the commuters in the region during traffic rush-hour periods. As such, this method is suited for defining the normal traffic pattern in terms of the time and location of the bottleneck formation. Therefore, the result can be useful in prioritizing the junction traffic to improve the global traffic efficiency.

4.2 Shortcomings of the Percolation Theory Approach

Our results from Sections 3.4 and 3.5 lend supportive evidence for using the percolation theory to model traffic dynamics and identify road segments that become the bottlenecks when the traffic network first changes from a non-congested state to a congested state. This approach, however, can only reveal the traffic dynamics through the use of historical data. As a result, it fails to adapt to any changes in traffic demand in real time. Furthermore, [LFW15] showed in their study that if the relative velocity of the bottlenecks is increased, q_c would increase, which signifies an improvement of the global traffic efficiency. Nevertheless, the percolation theory approach alone, can only provide a static observation; it does not address

how one can increase the relative velocity of a road segment since increasing the number of lanes is usually infeasible in an urban setting. Increasing the speed limit on a road segment will not have the desired effect unless the traffic light signal timing is synchronized in such a way to increase the overall throughput. Moreover, if we lengthen the green light on a road segment, the cross traffic would incur a longer wait time at the red light. Consequently, this perturbation may change the traffic dynamics and cause a congestion elsewhere. Perhaps one can take advantage of traffic simulations such as SUMO to compare the performance of different traffic light synchronization schemes to find the configuration with the highest throughput in terms of vehicles entering and exiting a region. Results from the percolation theory may be used as an initial starting configuration for an iterative algorithm to establish junction priorities under normal traffic demand in a region. One can perturb the traffic dynamics by first giving priorities to the bottlenecks and then finding the best traffic light configuration iteratively.

CHAPTER 5

Future Work

In order to improve the global traffic efficiency, more experiments need to be carried out in order to study the effects of perturbing the traffic dynamics by ways of traffic light synchronization, since it is usually infeasible to increase the road capacity in an urban setting where traffic congestion occurs frequently. The effectiveness of any proposed traffic flow management algorithm may be measured by the travel time of each individual vehicle and the collective global travel time, which is defined as the combined travel time of all vehicles in the scenario. It would be interesting to investigate the characteristics of any cases where there is a substantial increase in individual travel time when the collective global travel time decreases significantly. Other measures of effectiveness such as regional throughput would also be appropriate since its purpose is to assess the productivity of the transportation system in moving vehicles out of the region.

Moreover, the current literature does not seem to address how to prioritize the road segments to break the emerging traffic congestion in real time adaptively when abnormal traffic flow is caused by unexpected events such as accidents, lane closure, crowd dispersing from a stadium, etc. Therefore, the future work of this thesis would focus on extending existing methods that may be useful for such application. We would also anticipate developing new mathematical techniques to predict the collective behavior of the individual drivers in terms of traffic congestion forecast, and identify interventions to break this behavior in a network setting. Techniques stemming from self-organizing behavior, information diffusion, network synchrony, and reinforcement learning in general seem to be promising in pursuing this goal.

REFERENCES

[LFW15] 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.

[SS10] Matt Selinger and Luke Schmidt. "Adaptive Traffic Control Systems in the United States: Updated Summary and Comparison." *HDR*, 2010.