Community Detection on Urban Street Networks: A Segmentation Model for Urban Logistics Policy and Planning

by

Margaret Olivia Wilson

B.S., Ohio State University (2016)

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

Signature redacted

Author .....
Department of Civil and Environmental Engineering
May 17, 2019

Signature redacted

Certified by ....
Matthias Winkenbach
Director, Megacity Logistics Lab
Thesis Supervisor

Signature redacted

Certified by .....
Yossi Sheffi
Elisha Gray II Professor of Engineering Systems
Thesis Supervisor

Signature redacted

Accepted by.....
Heidi Nepf
Donald and Martha Harleman Professor of Civil and Environmental Engineering
Chair, Graduate Program Committee

Community Detection on Urban Street Networks: A Segmentation Model for Urban Logistics Policy and Planning

by

Margaret Olivia Wilson

Submitted to the Department of Civil and Environmental Engineering
on May 17, 2019, in partial fulfillment of the
requirements for the degree of
Master of Science in Transportation

Abstract

This thesis considers the community detection methods employed by network studies in a wide variety of contexts and adapts their use to the segmentation of an urban street network. In order to form partitions of urban street networks that are manageable as delivery territories or similar units of spatial aggregation, e.g., discrete demand zones, to be used in a study of urban freight distribution, extant community detection methods are assessed and adapted. Numerical experiments demonstrate that the sub-networks formed by these partitions display travel properties that make them a useful model for logistics transportation, especially in contexts where continuum approximation methods might be employed. The ratio of simulated trip distances over the actual road network to the idealized distance between the trip endpoints is used as a metric to quantify some travel properties of these segments. This ratio describes the magnitude of detour required by network conditions, which can offer a proxy for travel efficiency due to road network variations across a city. Using this metric, network-based partitioning algorithms are shown to produce sub-networks with internal travel conditions that are on average more efficient and less variable than sub-networks produced from extant methods of urban segmentation. This result is demonstrated on a wide variety of test networks in cities worldwide. In addition, a secondary use case as a decision-support tool for policymakers is proposed. Since this algorithm creates areas with a flexible spatial resolution in which boundaries are defined by infrastructure and geography, it may constitute a useful way to delineate areas where policy interventions should be employed. Because the impact and presence of freight traffic vary with local land uses (e.g., commercial, residential, industrial), the land use characteristics of these segments are also investigated to determine if network-based segmentation models capture more variation in land use characteristics than alternative methods.

Thesis Supervisor: Matthias Winkenbach
Title: Director, Megacity Logistics Lab

Thesis Supervisor: Yossi Sheffi
Title: Elisha Gray II Professor of Engineering Systems

Acknowledgments

Many thanks to my advisor, Dr. Matthias Winkenbach, and my labmates at the MLL who have given me tremendous support, guidance, and shared knowledge over the past few years. Your attitudes of patience, perseverance, and generosity in the face of intellectual challenges are commendable and will be carried with me.

Contents

List of Figures

List of Tables

Chapter I

Urban Segmentation for Planning Logistics Distribution Networks

I.1 Introduction

In freight distribution networks, the final tier or tiers of product distribution towards a customer are referred to as the last mile. Particularly in urban environments, last-mile distribution is often the most expensive, complex, and uncertain part of the global supply chain [Gevaers et al., 2011]. The planning and operational challenges posed by this setting, motivated by the practical need to distribute goods in urban areas, have led to last-mile distribution receiving increasing attention in both practitioner-oriented and academic literature. Two major factors drive the urgency of this problem: rapid urbanization and a changing market for last-mile delivery services. Between 2018 and 2050, the UN projects that the proportion of global population residing in urban areas will rise from 55% to 68% [United Nations Population Division, 2017]. Concurrently, we observe an increasing fragmentation of demand for urban logistics services, both in business to consumer (B2C) and business to business (B2B) markets, due to an ongoing boom in e-commerce and omni- and multi-channel distribution schemes. This growth drives a trend towards a higher number of small, direct shipments to individual consumers and, consequently, rapid growth in the number of delivery destinations compared to traditional brick-and-mortar retail, where logistics providers deliver goods to central store locations. The effects of this trend are am-

plified by rising customer expectations in terms of delivery lead-time, time-windows, and shipment individualization. Meanwhile, increasing fragmentation of shipments in urban B2B logistics is largely driven by the importance of the traditional highly fragmented retail channel in emerging economies and the growth of hyper-local fulfillment (i.e., the trend of keeping inventory close to the customer to ensure short delivery times) in developed and heavily urbanized markets [Fransoo et al., 2018].

A major stream of work addressing these problems is the design of last-mile delivery strategies. This problem integrates both facility location and vehicle routing decisions, and typical approaches employ a few notable, well-established approximation methods largely focused on reducing complexity by estimating travel distances along delivery routes. Location routing problems (LRP) are well studied in literature [Prodhon and Prins, 2014, Drexl and Schneider, 2015]. It is the size and complexity of these problems, when applied to last-mile logistics operations in urban contexts, which often include over tens or hundreds of thousands of customers, that renders exact routing approaches to be intractable. Therefore, route length estimation formulae are shown to approximate operational routing decisions sufficiently well while leading to substantial reductions in computation times for the LRP [Smilowitz and Daganzo, 2007, Nagy and Salhi, 2007, Winkenbach et al., 2015]. A useful quantity in such approximations is the so-called "circuity factor", describing the directness of local travel within a certain subset of an urban area. In routing approximations such as the continuum approximation (CA)-based methods applied in the previously cited works, this quantity multiplied with the Euclidean distance between the trip's origin and destination can be used to approximate trip distances by incorporating some measurements of the properties of the network over which travel is conducted. This convention can be applied to estimate travel distances at the regional level [Ballou et al., 2002] and local trips within a city [Ballou et al., 2002, Giacomin and Levinson, 2015]. The variance of circuity factors across a contiguous urban area and the implications of this property for intra-urban travel, especially logistics-related travel, is extensively treated by Merchán [2018]. There are multiple accepted conventions to define this factor, but all compare an idealized distance between two points (e.g., rectilinear or Euclidean) with the distance of the actual path traveled over the network. The resulting ratio offers a metric of the distance of de-

tour enforced by the topology of the street network in a local area and consequently, the efficiency of local travel.

Clearly, circuity and other infrastructural characteristics are not uniform over a city. Differences in population density, road types (e.g., highway vs. boulevard, one-way vs. two-way), and road network geometry can vary considerably from one area of a city to another. Several authors, (e.g., Winkenbach et al. [2015]), address this challenge by subdividing the city into a large number of adjacent rectangular pixels and accounting for road network variations at this increased, but still discretized, resolution. Localizing the calculation of average circuity factors into small sub-networks of streets within these discrete segments provides higher-resolution information about the infrastructure of different parts of a city and the consequences that infrastructure holds for logistics operators. This is possible and efficient thanks to modern tools such as public map databases and network analysis software, but the choice of how to sub-divide a street network affects the information contained in each segment about its travel properties. As an example, the rectangular pixel-based rasterization carries several limitations. It is placed arbitrarily over the urban area, so the boundaries between pixels are arbitrarily located and not reflective of the underlying practical subdivisions often defined by the physical geography of the city (e.g., waterways or motorways across which there is no vehicular or pedestrian access). Companies often define routing or delivery territories following natural or man-made barriers such as rivers and highways since those barriers are difficult or dangerous (e.g., to pedestrians, motorbikes, bicycles, etc.) to cross and/or cause significant delays in crossing them. Figure I.1.1 illustrates one example of this - the endpoints of this trip are close together, but reaching the destination requires a detour to follow one-way streets and reach a bridge to cross the river. In this case, treating the origin and destination regions of this trip as one unit and assuming it to be served by one route is not desirable or realistic. Further, this analysis often produces results that must be manually assessed to find pixels whose characteristics offer little meaningful information. Often, some pixels may be entirely composed of large industrial compounds, natural parks, waterways, freeway interchanges, hillsides or mountains, and other unsettled or infrequently traveled areas. Properties measured on the road networks of pixels that comprise these areas contain very little useful information,

Figure I.1.1: Detour necessitated by river and one-way streets, São Paulo, Brazil

out of context, about the actual transportation conditions of the area they describe and its importance to the logistics operator. Both these highly-local travel variations and the failure of a uniform raster grid to capture deviations in settlement patterns are obstacles to the robustness and generalizability of the pixel-based rasterization method. To overcome these limitations, this work proposes the use of a graph-theoretic method known as community detection for urban street network segmentation. The aim of this network-based segmentation tool is to create territories that are directly informed by the physical geography and topology of the road transportation network.

In addition to pixelated rasterization methods, another common way of subdividing cities is by postal code. This subdivision can be of practical convenience for local delivery operators who have firsthand knowledge of the urban region, but poses some challenges analytically. Postal codes (using the U.S. terminology, ZIP codes, equivalently hereafter) are typically drawn by governmental agencies to loosely reflect the patterns of settlement within the city. As such, it is unlikely that a postal code will bisect a neighborhood with important integral qualities such as an intact small central business district, or will include large swaths of uninhabited regions such as parks or lakes. Similarly, ZIP codes may vary in size according to population density, settlement or municipality limits, or other factors deemed relevant by those who create them. These are desirable qualities that give realism and practical value to a segmentation method used in a network design problem. However, geospatial datasets describing which parts of a city are assigned to which ZIP codes can be difficult to obtain in all cities, particularly outside of the U.S. and western

Europe. Even when such datasets are available, the ZIP codes themselves actually comprise a range of addresses on various streets, so post-processing these street-based geometries into shapes that can be described with clear bounds and a physical measurement of area (strictly necessary for continuum approximation), is non-trivial.

This thesis first proposes a method to segment urban areas based on street networks using community identification algorithms that partition the urban street network into sub-areas when modeled as a mathematical graph. This method is employed on a number of test networks and numerically evaluated to assess its usefulness as a tool in continuum-approximation based transportation planning problems. The partitions identified by the algorithm represent urban sub-regions that exhibit high connectivity and highly efficient local travel within regions and less efficient, less connected transportation networks between regions. These sub-regions are proposed as a potential alternative to pixel-based rasterizations in logistics network design models.

The remainder of this thesis is structured as follows. Chapter I develops and presents the segmentation algorithm. In Section 2, we discuss the relevant extant literature. In Section 3, we introduce a segmentation methodology, including data collection and processing, and assess this against various segmentation methods according to their applicability to a logistics use case. These assessments are conducted first on the metropolitan area of São Paulo, Brazil. We identify characteristics relevant to the method's applicability in route approximation formulae, a common tool employed by urban logistics planning, and numerically evaluate the graph-driven partitions according to these characteristics. We then present an explicit numerical comparison of graph-driven segments to rectangular pixels and ZIP codes, two other segmentation modes commonly used in urban logistics planning. In Section 4, we select a number of other case study cities spanning a range of regions, geographies, and development patterns and validate the method on these. In Chapter II, we combine the graph-driven partitions with land use and development data to explore the application of this method to identifying areas for which targeted urban freight policy interventions are anticipated to have the greatest effect.

I.2 Literature Review

I.2.1 Community Detection

A core goal of this work is to identify cohesive urban areas based on road networks. Particularly, we aim to find city segments or sub-regions that are better connected and more easily traveled within segments than between segments, as indicated by measurements of circuity factors for large samples of simulated trips. We use methods of network science and graph theory that have theoretical similarity to this problem setting - notably, community detection algorithms. These methods partition a mathematical graph into sub-regions based on various network properties. The terms "community", "partition," and "segment" are used interchangeably in this paper to describe these sub-regions, although the term "segment" will be preferred when discussing applications to street networks and their real-world implications, while the former two will be reserved for theoretical discussions. This section will provide a brief introduction to the extant literature on network science and community detection techniques with a particular focus on the methods tested and implemented subsequently.

Network science is an interdisciplinary mode of analysis with applications in many fields. It can be used to model such diverse phenomena as traffic patterns between fixed origins and destinations [Guimerà et al., 2005], human behavior on social networks [Centola, 2010], relations between firms and suppliers [Lazzarini et al., 2001], or the spread of failures among financial institutions [Battiston et al., 2012]. The core feature of these models is a graph representation where some components of a system are modeled as vertices (equivalently, nodes) belonging to a set V, and the connections or interactions between these components are modeled as edges (equivalently, links) belonging to a set E. A graph $G(V,E)$ can be identified by its edge and node sets. Network models and their associated measures enjoy particularly widespread use in sociology and management science, where the edges between vertices typically represent some kind of relationship. For example, two vertices linked by an edge may be two people who are "friends" on a social network or two firms that transact with one another. Note that, depending on the context, many real-world relationships are unilateral. For example, a firm may lend money to another but not owe

any debt in return, or one could follow or befriend a social media account without reciprocation. In a network setting, this is modeled as a directed graph where the edge (i, j) may indicate a connection from node i to node j but not necessarily the reverse. This is reflected in the adjacency matrix $\mathbf{A}$ where each entry (A_{ij}) reflects the connection between nodes i and j,

$$A_{ij} = \begin{cases} 0, & \text{for } (i,j) \in E, \\ 1, & \text{for } (i,j) \notin E. \end{cases} \tag{I.1}$$

Note that for an undirected graph, but not necessarily for a directed graph, $\mathbf{A}$ will always be symmetric. Broadly, community detection can be described as the algorithmic detection of subsets of nodes that are tightly linked together displaying a high density of edges within subsets and a relatively lower density of connections between subsets (cf. Lancichinetti and Fortunato [2009], Mucha et al. [2010]). The subsets, in a sociological setting, might represent segments of a population that have frequent or close interactions with each other but far sparser interactions with people outside of their sub-group. This is precisely the example demonstrated in Blondel et al. [2008], which studies the community structure among French- and Dutch-speaking users of a Belgian mobile telephone network and finds that users dramatically segregate themselves by language, with only sparse interaction between language groups.

Louvain Multilevel Modularity Maximization

Blondel et al. [2008] also introduce, in the same paper, a community detection method called *Louvain modularity* after the university in which it was developed. This technique is also called *fast modularity optimization* in reference to the computational improvements it offers over a similar, earlier method and the sometimes *multilevel modularity maximizaiton* after its recursive, multi-stage heuristic design. This method is agglomerative, in the sense that it initially assigns all nodes to solitary communities and consolidates them until a stopping criterion is met. It is further hierarchical, in the sense that similar nodes are recursively collapsed into a single node such that communities can be recovered at varying

levels of resolution. The Louvain algorithm formalizes the relative density of internal and external connections into a single partition-quality score called modularity,

$$Q = \frac{1}{2m}\sum_{i,j}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j). \tag{I.2}$$

Here, A_{ij} represents either the value from the adjacency matrix $\mathbf{A}$, or, in the case of a network where weights are assigned to edges to describe some quality or quantity of the flow traveling over them ("weighted network"), the weight of edge (i, j). For an unweighted network, m is the number of edges in the network,

$$m = \frac{1}{2}\sum_{ij} A_{ij}. \tag{I.3}$$

Otherwise, m is the sum of edge weights. $\delta(c_i, c_j)$ denotes a Kronecker delta function [Weisstein, 2019] on the community assignments c_i and c_j of nodes i and j, respectively,

$$\delta(c_i, c_j) = \begin{cases} 1, \text{ when } c_i = c_j, \\ 0, \text{ otherwise.} \end{cases} \tag{I.4}$$

Note that Q is a measure that can be calculated for a graph that has already been partitioned. It can be used to describe the quality of a given partitioning, e.g. to compare assignments or methods, or it can be considered an objective function to be maximized by adjusting partition assignments. However, explicit modularity maximization is computationally intractable and from a complexity standpoint has been shown to be NP-complete [Brandes et al., 2006]. For this reason, a number of approximation algorithms have been developed. The Louvain algorithm improves on existing heuristics in a number ways, leading to faster computational time, better modularity scores, and more realistic community allocations.

The Louvain algorithm is divided into two phases that are repeated iteratively. In the first phase, nodes are assigned as the solitary member of their own communities. Then, for each node, its neighbors are considered and the effect of moving node i into each of the communities of its neighbors j is calculated in terms of its marginal effect on the modu-

larity score. The allocation that results in the largest gain of modularity is accepted, and the next node is considered. Blondel et al. [2008] argue that the order in which nodes are assessed in this way does not have a significant effect on the final partitioning, but can affect computation time. This process proceeds until a local maximum is reached - i.e., until no single-node relabeling can increase modularity. The second phase redraws the labeled network by collapsing all nodes with the same community assignment into a single node with a self-loop whose weight is equal to the sum of weights of internal connections in the lower-level, pre-collapse graph. It is then possible to apply the first phase to the consolidated graph (which, in the context of the collapsed graph, consolidates some neighboring communities into one another) and repeat. The authors define one "pass" of the algorithm as one complete execution of phases one (labeling and reassigning individual nodes) and two (collapsing and consolidating communities). Multiple passes are then executed until modularity Q cannot be improved. This approach is part of a family of heuristics for maximizing modularity where computations are made on local regions of a network iteratively at various scales. They have been termed "smart local moving algorithms" and are the basis of many modularity-based community detection algorithms. Subsequent work has made incremental improvements to the Louvain method [Waltman and van Eck, 2013], but the basic features are the same. Applications of this popular method are widespread across and can be found in a number of various disciplines. In addition to the mobile telephone network example of Blondel et al. [2008], the Louvain algorithm has been to explore the interlocking nature of corporate directorates whose members often sit on boards of more than one company [Heemskerk Eelke et al., 2016]. At least two papers have used the Louvain algorithm to study the United States air transportation network, where nodes are airports linked by the flows of passengers between them. Clark et al. [2018] apply Louvain to this transportation network as one tool in modeling the network's resilience to various types of security- and weather-related failures. More generally, Gegov et al. [2013] use the Louvain algorithm to explore patterns in passenger flows across regions of the United States, specific airports, and seasonal variations. In a different setting, the same algorithm is applied by Meunier et al. [2009] to examine modular structures between parts of the brain on a network constructed from the patterns of communication between brain regions.

Fluid Communities

The algorithm of Parés et al. [2017] is a more recent addition to community detection methods based on a different methodology. Qualitatively, this method is interesting for the logistics use-case because it requires a specified number of partitions to execute. While this is less insightful from an unsupervised graph clustering perspective, it offers the practical benefit of being able to control the number of segments, for example, created in an urban area. The fluid communities algorithm (hereafter, "Fluid") is based on the propagation methodology. This is based on the idea of introducing a number of fluids (alternatively, a number of labels or beliefs which may similarly propagate) to a network environment where they will expand and push against each other subject to the topology of the network, which provides them channels for flow. Formally, this is represented as a number k of communities, $0 \leq k \leq |V|$, each initialized at a different and random vertex on the network. The spread of a fluid which comprises a community c is governed by the fluid's density, measured as the inverse of the number of vertices holding that label,

$$d(c) = \frac{1}{|v \in c|} \tag{I.5}$$

In each iteration, the algorithm updates the community of each vertex in random order according to an update rule. Letting C denote the set of community labels and J denote the set of neighbors of a node i, the updated community label c_i' for a node i will be,

$$c_i' = \underset{c \in C}{argmax} \sum_{j \in J/i} d(c) \cdot \delta(c(j), c), \tag{I.6}$$

where δ is again the Kronecker delta as in Section I.2.1. For a node i, assume that it is assigned community c. Then, summing across all neighbors j, only if j is in community c, add the density of the community c as given by Equation (I.5); otherwise, add nothing. If multiple arguments c maximize this function, the algorithm chooses one among them at

random. This update rule is shown in Equation (I.7),

$$c_i'^{\text{upd}} = \begin{cases} x \sim U(c_i'), \text{ if } c_i \notin c_i', \\ c_i, \text{ if } c_i \in c_i'. \end{cases} \tag{I.7}$$

Once this updating step is executed for all nodes and a new community assignment has been calculated, the communities are collapsed into a single node as in Louvain. Note, also, that summing over the set of neighbors ensures that all nodes in the community will form a connected graph component. After updating community labels, the density of the community is re-initialized to 1 per Equation (I.5), and the process repeats until community allocations are stable.

Note that if c_i, the community previously assigned to node i, is not in the set of community labels recommended at the next round, it is not assigned. This decreases the number of nodes assigned to the community, which increases the density. Each time nodes are removed from a community, it becomes more likely that that community will be chosen in the next round as a result of its higher density. This helps to ensure that the number of communities specified at the start of the routine appears in the final labeled graph, with at least some nodes allocated to that label. Examples of the Fluid algorithm are relatively rare in applied literature, which may in part be due to its recency. However, theoretical literature focusing on the network and algorithmic aspects of community detection problems commonly includes this method in comparative and benchmarking studies such as those described in Section I.2.1.

Comparative Studies

Outside of the development and design of algorithms that facilitate these analyses, a significant part of community detection literature focuses on benchmarking and assessing algorithms' performance on various types of graphs. While the suitability of a particular algorithm depends on the graph and the desired outcome, and quality likewise must be defined in context, a few notable comparative studies of algorithm performance are presented here. A few common dimensions for assessing algorithms are computation time, quali-

tative aspects of community size and distribution of community size, and benchmarking tests where algorithms are executed on a stylized graph seeded with known community structures and scored based on their normalized mutual information (NMI) measure compared to the *a priori* partition labels. NMI is a normalization of the Mutual Information (MI) score, which is borrowed from information theory and was first validated for use in community detection by Danon et al. [2005]. This score is based on a confusion matrix between the assigned community labels of nodes and the ground-truth community labels assigned manually or from context. It can be calculated from the confusion matrix N_{ij},

$$NMI = \frac{-2\sum_{i=1}^{c_A}\sum_{i=1}^{c_B} N_{ij} log(\frac{N_{ij}N}{N_i N_j})}{\sum_{i=1}^{c_A} N_i log(\frac{N_i}{N}) + \sum_{i=1}^{c_B} N_j log(\frac{N_j}{N})}, \tag{I.8}$$

where the number of real communities is c_A, the total number of nodes is N, the number of assigned communities is c_B, the sum over row i of N_{ij} is N_i and the sum over column j of N_{ij} is N_j. By this formulation from Danon et al. [2005], the NMI can then take a maximum value of 1 if the assigned communities match the real communities perfectly, and zero if there is no overlap between the assigned and actual labels.

Another useful technical term in these studies is the mixing parameter, or the average fraction of edges a vertex will have connecting it to other communities [Lancichinetti and Fortunato, 2009]. In other words, a node will have a fraction $1-\mu$ of its links connecting it to other nodes in its own community and the remaining fraction μ connecting it to nodes outside its own community, where the fraction μ is called the mixing parameter. From practical reflection, real networks and especially spatial street networks are likely to have a considerable amount of "mixing" as boundaries between communities are not defined by strictly segregated groups of nodes and many connections are likely to cross partition boundaries since nearby nodes are generally well-connected to each other. A network where groups are more clearly segregated, such as where speakers sort themselves into language groups, might not have a high mixing parameter. Additionally, as evident in the comparative studies presented below, a number of alternative partitioning methods exist in addition to the two described in Sections I.2.1 and I.2.1.

Some comparative analyses, which seek to rigorously assess the performance of partitioning methods in a number of distinct graph settings by well-defined performance criteria, are presented here. These comparative studies measure the Louvain and Fluid detection algorithms against several others, including some developed more recently. These studies find that, in a variety of settings assessed by a variety of measurement criteria, the two algorithms presented in Sections I.2.1 and I.2.1 perform well even against more recent advancements in the field.

Four comparative studies of community detection algorithms are presented in chronological order. All use similar methodologies in similar network settings. Lancichinetti and Fortunato [2009], building on their earlier work [Lancichinetti et al., 2008] which introduces one of most popular community detection benchmarking analyses, present a comparative analysis of algorithms on a large synthetic graph with heterogeneously-sized communities designed to mimic real-world systems. Especially suited to weighted and directed graphs, Louvain is outperformed only by a more recent algorithm called InfoMap [Rosvall et al., 2009] and is especially effective on "large" networks, defined here as networks with a few thousand nodes.

Orman et al. [2011] present a similar analysis on one of the benchmark graphs previously introduced by Lancichinetti and Fortunato [2009], but also challenge the supremacy of this quantitative methodology by including some qualitative aspects of network partitions as assessment criteria. In particular, they consider the number of communities and the distribution of their sizes as factors that would influence the usefulness and interpretability of community labels in real-world network settings. They find that while InfoMap again has outstanding performance on the quantitative benchmark, it tends to detect more, smaller communities than Louvain. (When tested on a sample urban street network, this holds particularly true.) While modifications of this hierarchical algorithm may be made to create or recover communities at different sizes on an urban street network, this proves to be challenging. Further, the theoretical basis of the InfoMap algorithm (based on encodings of node labels from an information-theoretic basis, employing the metric of informational entropy) has little sense in the context of an urban street network.

Yang et al. [2016] further expand the suite of community detection algorithms studied

and repeat the analysis on large numbers of realizations of large synthetic networks. In this study, Louvain was found to perform comparably to the labelings predicted by InfoMap and other more recent algorithms (in terms of NMI), especially for larger graphs. The Fluid algorithm, relatively more recent in its introduction (2017), does not appear in any of these analyses. However, in addition to introducing the Fluid algorithm, Parés et al. [2017] present benchmark tests comparing their own contribution to several existing algorithms, including Louvain and InfoMap. They find Fluid to be competitive with preexisting algorithms on large graphs with large mixing parameters. In all studies, Louvain is noted for its computational speed and scalability to large networks, which is comparable (if slightly less) than Fluid in this study. Fluid is also, notably, a strong choice in networks with large mixing parameters.

Table I.2.1 summarizes the relevant features of some algorithms and their relative strengths and deficiencies. As the top performer in many quantitative assessments, InfoMap is included. It records whether each algorithm allows a specified number of partitions, identifies communities at multiple levels of resolution ("Hierarchical"), can consider edge weights, can consider directed edges, has a bias toward small communities or large communities, performs as well on large networks as small, performs well on networks with large mixing parameters ("Mixing Tolerated"), and relatively faster than the others. Computational speed is another factor commonly used to asses these algorithms. Empirically, for this analysis, all three algorithms were found to be of reasonable and comparable computational speed on test road networks of 50,000 - 100,000 nodes and so this criterion is omitted from the table.

I.2.2 Route Length Approximation and Road Network Circuity

Section I.3 of this thesis explores the application of community detection-based segmentation methods to urban logistics applications. Specifically, application to the design of last mile distribution networks is investigated. These network design problems are often formulated as large-scale optimization models in which cities are segmented into service areas. Each service area can then be described according to its characteristic demand for logistics

	Pre-Specify Number of Communities	Hierarchical Algorithm	Weighted Network	Directed Network	Small Comm. Bias	Large Comm. Bias	Scalability	Mixing Tolerated
Louvain	No	Yes	Yes	No	No	Yes	Yes	No
Fluid	Yes	No	No	No	No	No	Yes	Yes
InfoMap	No	Yes	No	Yes	Yes	No	Yes	Yes

Table I.2.1: Summary of algorithm characteristics and relative strengths and weaknesses

services, area, and road network properties. This enables the integration of Continuum Approximation (CA) methods which are used to estimate the expected, near-optimal distance of delivery tours based on aggregate service-level characteristics at the service area or "city segment" level of aggregation. In these large-scale problems which aim to support long-term strategic decisions including facility location and fleet composition, day-to-day routing decisions are less consequential. For this reason, CA methods are often used to make approximations at this level of planning. Aggregate service area models and related routing cost estimation formulae allow sufficiently precise estimation of routing decisions while leading to substantial reductions in computational time [Nagy and Salhi, 2007], which is particularly relevant given the scale of most last-mile logistics problems. Smilowitz and Daganzo [2007] first demonstrate the application of CA methods to network design problems. To account for heterogeneity in demand density across an urban area, Winkenbach et al. [2015] impose a grid of rectangular pixels dividing a city into discrete service areas.

One key measurement of urban road networks that is commonly used to characterize these service areas is a circuity factor. Circuity measures the relative detour incurred by traveling over a road network compared to the Euclidean distance between the origin and destination. We define the circuity c as the ratio between the shortest-path network distance

d_N and the Euclidean distance d_{L_2},

$$c = \frac{d_N(p,q)}{d_{L_2}(p,q)}, \tag{I.9}$$

for any pair of path origin and destination locations (p,q). From this definition, it becomes clear that higher circuity factors indicate higher levels of network inefficiency [Barthélemy, 2011]. In route approximation settings, circuity factors are used as a multiplier to approximate actual road network travel distances. Early work on circuity factors in urban areas assumes travel over a rectilinear grid and finds a theoretical circuity factor of $c = 1.273$ [Larson and Odoni, 1981]. During the same period, empirical studies find values for the circuity factor between 1.16 and 1.28 for selected urban areas and approximately 1.35 for rural areas in the USA [Love and Morris, 1979]. A more recent empirical study by Giacomin and Levinson [2015] estimates a circuity factor of 1.34 for the fifty most populous US cities. However, these studies focus on trips of up to 60 km within the urban area. This might be misleading for last-mile delivery because last-mile trips typically constitute inter-stop distances of less than a kilometer and Levinson [2012] concludes that circuity increases with a decrease in trip distance. Most recently, Merchán and Winkenbach [2018] find that the empirically derived mean circuity factor for São Paolo is about twice as high as earlier estimates for trips of less than one kilometer in distance, $c = 2.51$. This analysis is based on the one km^2 pixelation logic introduced before.

I.2.3 Urban Network Science

Transportation networks, and particularly urban streets, are a popular topic of study in applied network science. Dating back nearly three centuries, one of the earliest seminal problems of network science — Euler's classic seven-bridge problem at Königsberg — considered travel over urban roads [Barabasi and Pósfai, 2016]. More recently, the availability of online map data describing road network geometries worldwide has fostered a number of empirical studies (e.g., Boeing [2018], Chan et al. [2011], Porta et al. [2004]) that apply network analyses as a means of describing and categorizing urban forms strictly in terms of topology and geometry. Other authors borrow methods of network science to

examine urban phenomena including city development patterns and transport flows. Road network comparison studies are mainly focused on inter-city topology differences. However, some literature exists that focuses on intra-city road network characteristics.

Applications of community detection methods to road networks are more limited and, in extant literature, not considered in relation to logistics planning. Jiang and Claramunt [2004a] use graph measures to identify areas and links that are of key importance for passenger transportation flow. Zhong et al. [2014] similarly investigate travel behavior and even study the community structure discussed elsewhere in this work, but use a network abstracted from transit card origin-destination patterns and not a spatial road network. Both Levinson [2012] and Chan et al. [2011] correlate network properties with land city size, city development patterns, and intra-urban travel behavior in several cities but do not address community structure in street networks. At the level of an intra-city investigation, Duan and Lu [2013] study the robustness of city road networks based on community structure derived from the road network. Although they detect communities in six major cities to support their analysis of robustness, the communities by themselves are of limited practical value to urban logistics decision-makers. Gudmundsson and Mohajeri [2013] analyze the geometric differences between older and more recent parts of the street networks for 41 cities in the United Kingdom. However, they manually subdivide the city into different segments, while we rely on community detection for segmentation.

Barthélemy [2011] argues that urban road networks are well approximated as two-dimensional planar spatial networks, making a wide range of network-scientific tools relevant for the application. One notable constraint of graph representations of urban street networks is the presence of directed or undirected edges (i.e., one-way streets). This is a complication in terms of implied consequences for real-world vehicular flows on the network as well as the graph-based analytical methods that may be employed. A natural representation of urban roads as networks is the "primal" representation, where streets represent links and intersections, nodes. The alternative to this representation, the "dual" graph, is the reverse of this node-edge correspondence: streets are nodes, and two nodes are connected if the corresponding streets intersect. Various network measures can then be re-interpreted in the context of the dual graph. Node degree, for example, in a primal graph

represents the number of streets joined at one intersection, but in a dual graph represents the number of streets that intersect a given road. Both primal and dual graph representations are frequently employed in the study of urban street networks. In this work, we too examine both.

A relevant literature stream surrounding the study of dual street graphs is known as *space syntax* theory. Space syntax constitutes not only a notable use of dual street graphs in network-scientific methods but also a major intersection of network science and urban design. Drawn from Hillier and Hanson [1984], space syntax analyzes urban areas and their street networks in terms of axial lines, continuous streets, and lines-of-sight as perceived by human inhabitants. It uses measurements of connectivity to interpret human interactions with urban spaces and correlates measures of connectivity and accessibility (easily conceived and measured in network representations) with urban phenomena such as traffic flows, human navigation, safety, and livability. It has been criticized for its subjectivity with respect to the methods used to account for continuity of space and spatial experience, the basis of graph construction in space syntax, and for the use of dual graphs which discard meaningful metric information and street geometries [Ratti, 2004]. Methods for constructing these generalized dual graphs while selectively retaining relevant information are a popular topic in the field [Porta et al., 2006, Jiang and Claramunt, 2004b]. However, as this study focuses solely on network topology and wishes to later draw correlations between network properties and network phenomena, all dual graphs used here are derived only from the adjacency matrices of the readily available primal graphs.

I.3 Urban Segmentation Model Implications for Urban Logistics

The two community detection algorithms surveyed in Section I.2.1, Louvain and Fluid, are implemented on a test network centered on the metropolitan area of São Paulo, Brazil. Louvain is chosen for its favorable performance in a number of benchmarking assessments detailed in Section I.2.1 and the close analogy of modularity to in-group travel efficiency

relevant for logistics use cases. Fluid is chosen for the convenience of a pre-specified number of partitions, quick computation time, and the sensible interpretability of flow propagation in the context of a road network.

A set of experiments is designed to assess each algorithm's relevance to forming districts for a logistics planning problem in which route approximation formulae will be applied. Recall that three important criteria for this application are (1) low value of road network circuity within partitions relative to between partitions, (2) low variance of road network circuity within partitions, and (3) relatively convex territory shapes. Numerical experiments are designed and presented to consider all three of these criteria, and the results of both algorithms are compared. In addition, two possible transformations of the urban road network graph, the primal and dual network representations, are explored.

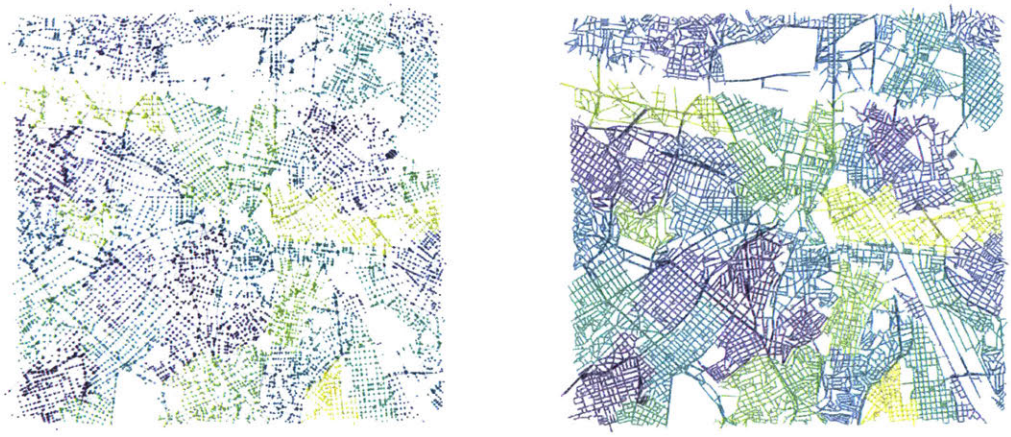
The Python package OSMnx [Boeing, 2017] is used to query road network data from Open Street Maps (OSM). This package integrates network information with contextual street data from OSM such as road classification, number of lanes, link length, and one-way/restricted access streets. Similarly, the Python package Networkx [Hagberg et al., 2008] version 2.0 is used for subsequent analyses on these street networks in Python. Other non-original code, notably the Python implementation of the Louvain algorithm, is taken from an online repository maintained by one of the original academic authors[1].

I.3.1 Primal and Dual Representations of Street Networks

Section I.2.3 introduces the terminology of primal and dual graphs and their relevance to urban street network analysis. Recall that in the primal graph representation, streets are edges and intersections are nodes with the graph preserving regular network geometry. In the dual graph representation, streets are nodes and two nodes (streets) are only connected by an edge if they intersect. The dual graph preserves the connectivity properties of the primal, geographic street graph but is less intuitive to visualize and conceptualize. The application to logistics distribution motivates the need to label streets, not intersections, with community membership. In a network design or route-districting model, these partitions are used as a unit of aggregation for customer demand, which occurs at customer locations

[1] https://github.com/taynaud/python-louvain

along city streets. Thus, each point of customer demand should be labeled with a community identifier so that all demand occurring within a partition can be described by aggregate community-level properties. However, the major disadvantage of using the dual graph as input is the confusion of relevant properties with which to assign edge weights. Many community detection algorithms, especially those using modularity maximization, can be parameterized with unequally weighted edges, typically describing some network quantity such as traffic or flow. In the case of a street network, travel time, vehicular capacity, or traffic might all be useful quantities to consider. However, these are clearly all linked to street (primal edge, dual node) properties and the selection of a method to assign a relevant weight to street intersections (dual edges, primal nodes) is far less intuitive.



(a) Output with primal graph as input

(b) Output with dual graph as input

Figure I.3.1: Executing a community detection algorithm on a primal street graph labels nodes, which must then be translated to edge labels, while executing on a dual street graph labels edges directly.

Because community detection algorithms are formulated to label nodes with community membership according to the presence or absence of connecting edges, when executed on a primal network representation they produce a graph with labeled nodes such as that shown in Figure I.3.1a. Inverting the graph to its dual representation and executing the community detection algorithm on the dual input produces a graph with labeled edges such as the one shown in I.3.1b. While creating the dual representation of an urban street network necessitates discarding edge direction information and is harder to visualize, it does

not require a method for translating labels from nodes to edges. However, for some purposes (such as detecting communities on a graph with weighted edges) the use of a primal graph may be preferred. A simple method is adopted to translate node labels to edge labels. Given that each edge has exactly two endpoints in the graph, if both endpoints are labeled with the same segment index, the edge adopts that segment index. If the endpoints are differently labeled, a random choice between the two endpoint indices is made. Extensions to this simple heuristic are certainly possible - for example, the neighboring nodes of each endpoint could be considered and the majority label among these votes for the label applied to the edge. However, this more informed method provides little marginal gain in the numerical assessments presented below, and in the interest of saving computation time on large networks, the simpler approach is employed here.

I.3.2 Comparative Study of Partitioning Algorithms on Street Networks

Descriptive Segment Properties

For each of the candidate algorithms, the method is applied to a test network consisting of a rectangular 100 km^2 sample of the urban core of São Paulo, Brazil centered at 23.563346 S, 46.654165 W, which is the area highlighted in Figure I.3.2. From this network, local roads and streets are extracted by filtering the graph according to OSM's road classifications.

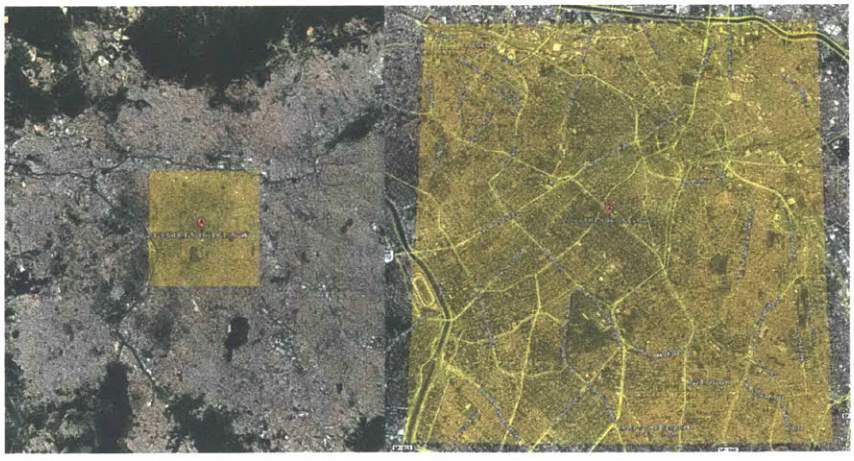


Figure I.3.2: Area of 100 km^2 test network in São Paulo, Brazil.

OSMnx assigns to all road segments various metadata such as infrastructure types (e.g., cycle path, power line, or vehicular highway), length, number of lanes and roadway classification. However, these metadata are not entirely consistent between cities. Entries are frequently absent, particularly in network queries outside of the U.S. and western Europe, and some descriptive classifications (e.g., "motorway" vs. "trunk" roadway) are only qualitatively defined and may be altered by users (Open Street Maps is an open-source platform), so their assignments may not be consistent across different networks.

Roadway classifications provided by OSM range from motorways to local streets. The largest-volume roadways are categorized as a motorway, motorway link (e.g., access ramp), trunk, and trunk link. Links with classification below trunk link are used and higher (larger-volume) road classifications are discarded. The rationale for this is that last-mile deliveries will mostly be conducted using local streets, whereas large-volume road types will be used for line-haul transportation (a separate problem from last-mile distribution). However, due to the potentially imprecise nature of this metadata, it should be used with caution in non-urban or remote regions where data collection efforts may be sparse. Any small disconnected subgraphs, which occasionally occur with or without this filtering step due to local inconsistencies in street network data and geometry, are discarded and only the largest connected component is considered for testing. In the more complex multi-stage algorithm presented in I.3.4 and subsequently, exceptions are designed to ensure that all viable sub-networks of a query area, not just the largest connected component, are processed. The resulting test network, in this case, has 10,962 nodes and 23,035 edges. For each community detection algorithm, a brief general summary of relevant output characteristics is described in Table I.3.1. For the Fluid community detection algorithm, the number of partitions is specified *a priori* and is selected to target an average segment area of 1 km^2. Because the partitions are irregularly shaped, the area of each is approximated by the α-shape method [Edelsbrunner et al., 1983], which is further explained in Section I.3.2.

Note in Table I.3.1 that the main advantage of the supervised Fluid method is the ability to fix the number of communities assigned. For this method, the average area of segments produced was slightly lower than the intended 1 km^2. It is also observed that implementing the same algorithm with the same parameters on the primal vs. dual representations of

the same graph can produce significantly different results in terms of the number, size, and roadway length of communities identified. While from a practical perspective all the values in this example are within reasonable ranges, this indicates that choosing a primal vs. dual graph representation as input to the algorithm could have meaningful consequences and should be further considered (as in Section I.3.2, for example).

		Number of Segments	Avg. Segment Area [km^2]	Avg. Nodes per Segment	Avg. Edges per Segment	Avg. Street Length per Segment [km]
Method	Louvain, Dual	55	1.273	228	330	33.825
	Louvain, Primal	49	1.397	204	295	29.114
	Fluid, Dual	100	0.711	118	162	16.139
	Fluid, Primal	100	0.886	120	162	16.267

Table I.3.1: Summary of community detection algorithm results on 100 km^2 sample network.

Road Network Circuity

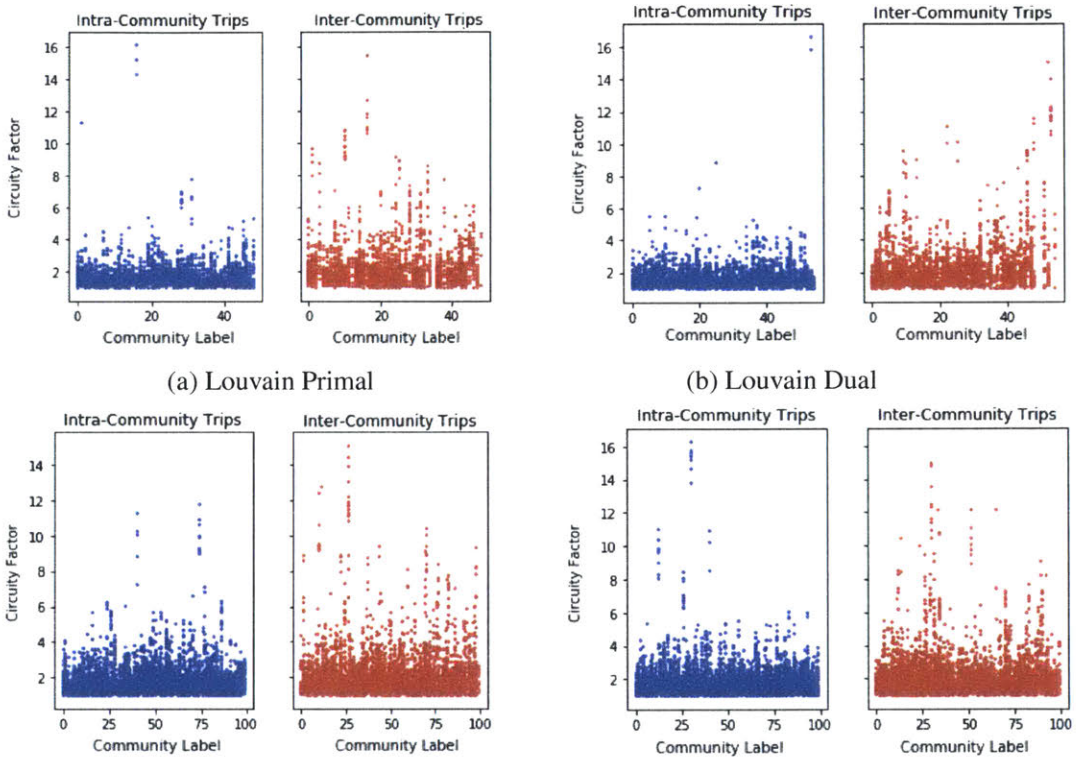
Road network circuity, as defined in Section I.2.2, is used to assess both the efficacy of community detection methods in identifying areas of highly efficient local travel and to compare the performance of candidate algorithms. Once a method (algorithm) and implementation (primal vs. dual) is chosen, we compare the mean and variance of simulated circuity factors on a community-partitioned network to networks partitioned with pixels and ZIP codes in Section I.3.4. To assess the efficacy of the candidate algorithms and graph transformations in identifying local regions of low circuity, we compare the average circuity factor for trips within partitions (internal trips) to the average circuity factor

		Mean, Internal Trips	Mean, External Trips	% Difference in Mean	CV, Internal trips	CV, External Trips	% Difference in CV	p
Method	Louvain, Dual	1.724	2.283	24.5	0.443	0.627	41.5	$p < 0.01$
	Louvain, Primal	1.763	2.314	31.3	0.502	0.618	23.1	$p < 0.01$
	Fluid, Dual	1.829	2.044	11.7	0.535	0.584	9.2	$p < 0.01$
	Fluid, Primal	1.818	2.069	13.8	0.481	0.608	26.4	$p < 0.01$

Table I.3.2: Summary of circuity results on 100 km^2 sample network

for trips that cross partition boundaries (external trips). Node-to-node trips are simulated with Euclidean origin-destination distances in a specified range, 500 to 600 meters, chosen to approximate a realistic inter-stop distance that might be encountered during last-mile delivery operations. The Networkx *shortest_path* function is used to calculate the length of each node-to-node trip in meters on the directed graph that models the street network. This function can only calculate trip distances between nodes, hence the choice of only node locations as trip endpoints. Table I.3.2 presents mean values for node-to-node trips between 500 and 600 meters in Euclidean distance. 50 internal trips are simulated within each segment and an equal number of external trips, with no constraints on the origin or destination segments other than that they must be different, is also created. Coefficients of variation (CV) are shown for each subset to indicate the mean-normalized standard deviation within each class. A Welch's t-test (not enforcing the assumption of equal variance between populations, but assuming a normally distributed population) is used to assess the similarity of circuity factors between the internal and external groups, and p-values are reported in the table. For each test, p-values indicate the rejection of the null hypothesis that the internal and external trip circuity factors have equal means to high levels of confidence. Recall that internal trips are conducted entirely within a partition while external trips cross partition boundaries, and all trips are within the same distance range. As Table I.3.2 and Figure I.3.3 show, for all community detection methods, there are meaningful differences in the properties of local travel within segments compared to between segments. These

results indicate that a partition-based territory might reduce routing inefficiencies in local travel, and that while results for all methods investigated are comparable, the Louvain modularity computed on a dual street graph reduces the mean circuity factor within segments comparably to other methods, and shows the greatest potential to produce segments that reduce variance in inter-stop travel conditions relative to other segmentation methods. The Louvain Primal implementation performs comparably.



(a) Louvain Primal

(b) Louvain Dual

(c) Fluid Primal

(d) Fluid Dual

Figure I.3.3: Circuity factors for four detection methods, compared between intra-community (internal) and inter-community (external) trip classes.

Validity of Circuity Factors in Non-Convex Segments: Measuring Irregularity of Partition Shapes

The convexity of the shapes defined by partitions is another metric for assessing these methods in the context of route length estimation formulae and, in particular, the accuracy of the circuity factor in describing the road network over which local trips are conducted.

In this section, the first of two measures of partition convexity is introduced. This first metric considers only the geometric regularity of the partition boundary, and the second incorporates network information and simulated travel patterns to determine the frequency with which a road user leaves the partition to conduct a trip with origin and destination in the partition. Having shown in Section I.3.2 that these partitions describe areas of more efficient local travel, it may be assumed that a knowledgeable driver or routing program would lead drivers along routes within this less-circuitous region, and this is the assumption tested by the second convexity metric. The convexity of partitions from a strictly geometric perspective implies that a continuum approximation would be an appropriate solution for routing problems on the partition area based on the accuracy of a theoretical abstraction of the real properties of delivery territory. Using travel information to model the "practical" convexity, i.e., the frequency with which internal trips are realistically expected to be routed outside the partition based on network topology, describes the validity of a given circuity factor in characterizing real inter-stop travel behavior within the partition. Aside from these considerations, roughly convex and regular shapes are also easier to visualize and interpret.

While the mathematical definition of a convex polygon may be simple, quantifying the non-convexity of shapes described by groups of points or lines (as here) is not necessarily so. To address this, we first construct an approximate polygon from the endpoint nodes of all edges comprising each partition. This polygon should roughly cover edge endpoints as closely as possible, i.e., with a minimum area such that polygon boundaries do not collapse into one another. This shape will represent the compact, non-convex shape of the point set and is constructed with the use of α-shapes [Edelsbrunner et al., 1983], which parametrically collapse a convex hull while maintaining coverage over a set of points. As the parameter increases, the shape boundary collapses towards the interior as shown in Figure I.3.4. This approach relies on the Delaunay triangulation of a point set, a non-unique set of triangles drawn such that i) all triangle vertices coincide with points in the point set, and ii) no points lie inside the circum-circle of a triangle. By specifying a parameter α, the value of which is arbitrary but relative to the coordinate system and scale of the point set, the maximum allowable radius of these circum-circles is changed, which has the effect of removing triangulation lines around the boundaries of the point set in areas where point

density is low and contracting the α-shape boundary towards the denser middle of the point set. The "minimum viable" α-shape is defined as the shape produced for the largest value of α (i.e., most restrictive with respect to circum-circle radius) for which the triangulation describes one contiguous shape. "Islands" of only one point bridging between two other point clusters are allowed. In subsequent analyses, this α-shape is used to approximate the polygon defined by sets of nodes comprising a partition subgraph. All measurements describing the true area of a segment are calculated from the area of this shape.

The α-shape method of polygon approximation is combined with the approach of Brinkhoff [1995] who creates an index to quantify the irregularity of polygons based on measurements of the shape's area and perimeter. This index, which is termed complexity, is calculated from three measures of an irregularly shaped polygon's perimeter, area, and boundary when compared to the convex hull of the same polygon. The three components of the index are termed "frequency of vibration", "amplitude of vibration" (where vibration is used to describe the jagged deviations from straight-line polygon boundaries), and "deviation from convex hull". They are calculated as follows.

Frequency of vibration: A measure of the frequency with which a polygon edge deviates from the convex equivalent edge. Measured by the number of points on the boundary of the non-convex polygon compared to the number of points on the boundary of the convex polygon. For a polygon *pol*, where *notches* are boundary points of the non-convex shape not coinciding with the convex boundary,

$$notches(pol) \leq vertices(pol) - 3. \tag{I.10}$$

Normalized to [0,1], this yields,

$$notches_{norm} = \frac{notches(pol)}{vertices(pol) - 3}. \tag{I.11}$$

Amplitude of vibration: A measure of the increase in perimeter added by deviations from the convex boundary,

$$ampl = \frac{perimeter(pol) - perimeter(convexhull(pol))}{perimeter(pol)}. \tag{I.12}$$

Deviation from the convex hull: A measure of the change in area caused by deviations from the convex hull,

$$conv = \frac{area(convexhull(pol)) - area(pol)}{area(convexhull(pol))}. \tag{I.13}$$

Complexity Index: The complexity index is then calculated, per Brinkhoff [1995],

$$compl(pol) = 0.8 \times ampl(pol) \times freq(pol) + 0.2 \times conv(pol). \tag{I.14}$$

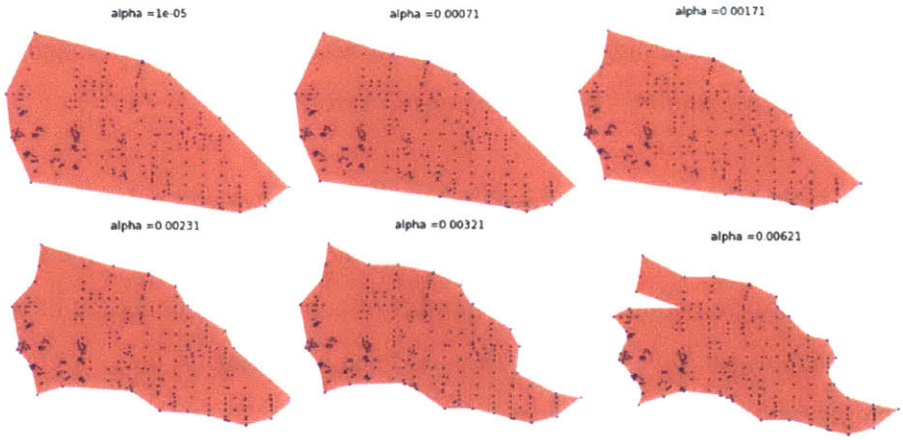


Figure I.3.4: Evolution of α-shapes as parameter α changes.

Table I.3.3: Average convexity measures of community maps produced by each algorithm on the test network

Method	Frequency	Amplitude	Deviation	Complexity
Louvain (dual)	0.781	0.133	0.257	0.137
Fluid (dual)	0.770	0.156	0.300	0.159
Louvain (primal)	0.768	0.143	0.249	0.140
Fluid (primal)	0.759	0.139	0.285	0.144

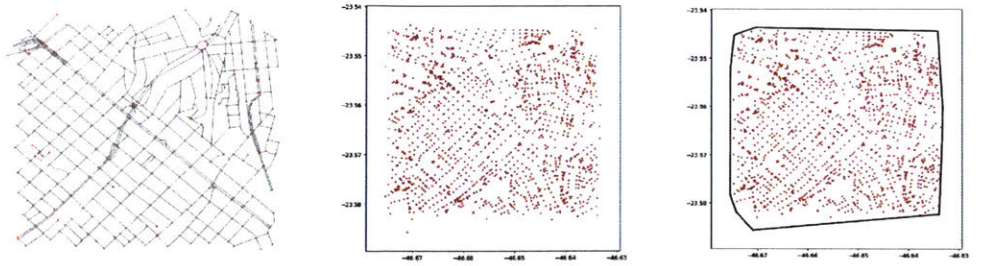
Table I.3.3 presents the results of this analysis performed on the polygons that ap-

proximate the node sets of community-level subgraphs produced by the four investigated community detection algorithms. Results between algorithm implementations are not dramatically different. However, it is noticed that the shape complexity appears to be lower for both Louvain implementations than both Fluid implementations. This could potentially be the result of the Fluid algorithm's "density" rule, which restricts the growth of communities according to the density rule (as described in I.2) and may distort community assignments by this metric which does not consider realistic local variations in edge density or community size. Louvain has no such constraint and operates only on the connectivity patterns of the network, and so may end up producing communities that are spatially more well-defined than those of the Fluid algorithm.

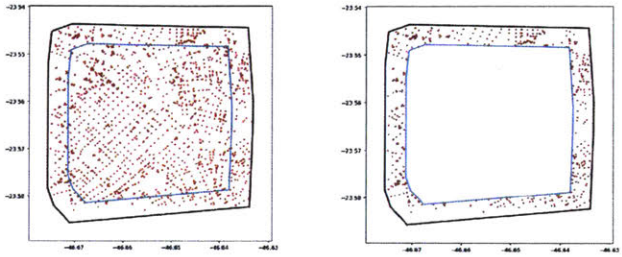
Validity of Circuity Factors in Non-Convex Segments: Measuring Fraction of Simulated Trips Departing from Partition Boundary

An additional measure of partition non-convexity is obtained by simulating trips around the boundary of the shape defined by the partition subgraph and counting the fraction of shortest path trips, simulated across the entire graph, that utilize links outside the subgraph. The goal of this metric is to assess the likelihood that, in a real travel setting, local trips conducted between points inside the partition would depart from the road network described by the partition label. In the network design context, local trips utilizing these external links would violate the assumption that the circuity factor used in the route length estimation formula accurately describes the road network over which local trips are actually conducted. This assumes that local trips are actually conducted using shortest network paths between subsequent stops, which is a reasonable assumption in light of navigation aids such as GPS maps and proprietary routing software used by delivery operators.

To develop this metric, a scaling operation is performed on the partition geometry in an attempt to limit the nodes that may be candidates for trip endpoints. We observe that some sampling bias that would occur if any nodes on the subgraph were eligible to be trip endpoints. In larger partitions with more nodes comprising the interior, random selection of endpoints would be more likely to include these interior points, which have no possibility of producing trips that depart the subgraph. Since this metric does not seek to penalize or



(a) Original network (b) Extracted nodes (c) Nodes with convex hull



(d) Scaled convex hull (e) Outer nodes are O/D candidates

Figure I.3.5: Schematic representation of sampling process for origins and destinations in trip-based convexity metric

advantage segment size and only to examine the characteristics of trips that occur around the boundary of the partition, an unconstrained sample would be undesirable. A schematic representation of this process is shown in Figure I.3.5. Numerical comparisons of this metric between algorithms are shown in Table I.3.4.

To develop the endpoint candidates, the subgraph defined by the segment label is extracted from the larger road network graph (Figure I.3.5a). Nodes are extracted from the subgraph (Figure I.3.5b) and the convex hull of the node set is calculated (Figure I.3.5c). The convex hull is then scaled by some factor (a chosen parameter; in this case, 0.8 is used) along both dimensions in the plane, centered at the center of the bounding box defined by the hull shape (Figure I.3.5d). This can be accomplished with the Python Shapely library's *affine* module which applies various types of affine transformations to 2D and 3D geometric objects in Python. Nodes belonging to the partition subgraph that intersect with the original convex hull but not the scaled convex hull (i.e., are near the boundary) are eligible to be origins or destinations of simulated trips (Figure I.3.5e). Between each origin and destination, the shortest path over the larger directed street network graph is computed. The edges

comprising this path are then compared to the edge set of the partition subgraph. If any of the used edges do not also belong to the subgraph, the trip is counted as utilizing an external edge. The higher the fraction of trips that utilize external edges, the less reasonable it is to assume that the calculated circuity factor for a given segment sub-network accurately describes the real-world travel that would actually occur between delivery locations located in the segment.

Table I.3.4: Fraction of peripheral trips departing subgraph, average across all subgraphs produced by each algorithm

Method	**Fraction**
Louvain (dual)	0.190
Fluid (dual)	0.216
Louvain (primal)	0.177
Fluid (primal)	0.228

The results of this experimental metric appear to reveal a significant difference in the travel properties of Louvain-based and Fluid-based road network communities. The trip convexity measure for both implementations of the Louvain algorithm produces an average value that is considerably lower than the Fluid algorithm. In other words, on average, simulated trips between points in Louvain communities depart the designated community-labeled road network less often than simulated trips in Fluid communities, when trip paths are calculated by a shortest network distance algorithm that considers real street lengths and real street directional constraints. This indicates that Louvain communities may be more compact and relatively convex, which is a desirable property for the theoretical validity of CA methods that also supports manageable visualization and interpretation.

I.3.3 Recursive Algorithm for Controlling Segment Size

To facilitate the use of these segments in a last-mile distribution network design problem, the segmentation method must be able to guarantee that segments emerge at a reasonable size and scale. A two-stage, recursive algorithm is developed that applies the Fluid and Louvain partitioning methods sequentially until a convergence criterion relating to maximum segment size is met. Unlike a single-stage, standalone implementation of either

algorithm, this method also retains all connected components of a regional street network graph and operates on each of them sequentially, so that no sub-sections of the network are lost. Algorithm 1 describes this routine in detail.

I.3.4 Comparison to Alternative Segmentation Methods

As noted previously, several conventions exist for dividing an urban area into segments for network design studies and other logistics applications. One useful segmentation strategy is the pixelation approach presented elsewhere in this thesis. In practice, ZIP codes are commonly used to delineate delivery territories and serve as a practical level of aggregation for planning last-mile delivery services that mirrors what is used in practice. ZIP codes in São Paulo are 8-digit numeric strings organized hierarchically, where the left digit(s) represent a broad area and the right digit(s) describe successively smaller regions. 3-digit ZIP codes are superimposed on the test network. This level of resolution is used due to the difficulty of recovering polygons from higher-resolution ZIP code shapes.[2] In this section, a grid of 1 km^2 pixels is imposed on a 400 km^2 test network of downtown São Paulo. The graph-based multistage urban segmentation of Section I.3.3 is implemented on the same test network. The result of this execution is shown in Figure I.3.10. Measurements comparing circuity between segmentation methods are summarized in Table I.3.5. Simulated trips are again constrained to distances between 500 and 600 meters, as measured by the Euclidean distance between origin and destination.

The key result of Table I.3.5 is the trend in circuity factor properties from pixels, to ZIP codes, to the Urban Segmentation Model (USM) result. Pixels are located without regard to the underlying network and its geography, and should not contain any meaningful information about travel behaviors. ZIP codes, in theory, are drawn around geographically or socially cohesive areas like neighborhoods and administrative zones. (It is a reasonable observation that areas with clustered occupancy and activity patterns would develop well-connected road networks as a reflection of inhabitants' activities and habits, but this

[2]ZIP codes are assigned as address ranges across groups of streets (essentially, small graphs), not polygons. Significant spatial processing is required to transform the geometry from collections of streets to contiguous and non-overlapping polygons.

Algorithm 1 Multi-Stage Urban Segmentation Routine

procedure MULTISTAGE USM (G)

User specifies:

minimum_edges_to_detect: the minimum number of edges required on a graph component to execute a partition algorithm (prevents execution of algorithm on anomalous, disconnected geometries consisting of only a few nodes or edges)

minimum_area_to_detect: the minimum area of a graph component required to execute a partition algorithm, intended to prevent partition areas that are spatially compact even if many edges are present

maximum_segment_size: maximum allowed size (in real spatial area measurement, e.g. km^2 as here), of segments

▷ G is a graph of an urban street network; possibly disconnected, possibly directed, possibly weighted

$G \leftarrow (V, E)$

$E' = (i', j')$ if (i, j) or $(j, i) \in G$

▷ Convert graph to undirected equivalent

$G_u \leftarrow (V, E')$

for $c = (V^c, E^c)$ in connected components of G_u **do**

$\alpha_A \leftarrow Area(\alpha(c))$

if $|E^c| \geq$ *minimum_edges_to_detect* **and** $\alpha_A \geq$ *minimum_area_to_detect* **then**

Execute Louvain modularity on c

while maximum α-area for any partition in c is above threshold **do**

Let E^P denote the edge set of the partition subgraph

Let α^P denote the minimum viable α-shape of the partition subgraph and α_A^P its area

if $|E^P| \geq$ *minimum_edges_to_detect* and $\alpha_A^P \geq$ *minimum_area_to_detect* **then**

$$k \leftarrow \left\lfloor \frac{\alpha_A^P}{(minimum\_segment\_size + maximum\_segment\_size) \cdot 0.5} \right\rfloor$$

▷ Determine k, the number of partitions to create in the subgraph, by dividing the average desired segment size into the subgraph area

Execute Fluid with k partitions

end if

end while

end if

end for

end procedure

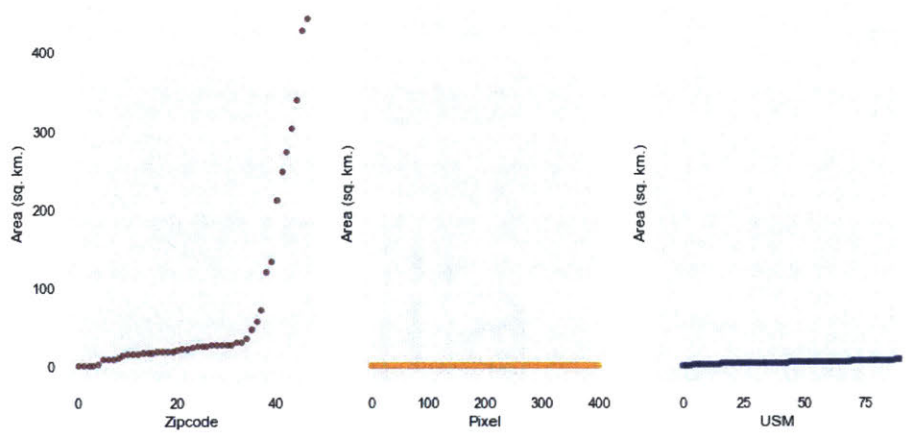
will be further explored in Chapter II.) The hypothesis is that ZIP codes will display an intermediate level of efficiency with respect to measurements of local travel because they are partly formed on the basis of network information, while pixels will be less efficient and the USM will be more efficient. The USM is entirely grounded in network properties and should produce segments where local trips are served by local street networks that are dense and well-connected, only incurring small detours with a high degree of reliability. Trips within these subnetworks should be characterized by low circuity factors, while trips between subnetworks are more likely to be constrained by natural and infrastructural boundaries like roads, motorways, and one-way streets; these trips are hypothesized to have a higher average circuity factor. Dispersion of circuity factors is important for CA-based network design problems where a single circuity factor is used to describe an area's road network. For this factor to be reasonable, internal trips (analogous to inter-stop trips in the continuum approximation) should display a lower coefficient of variation than arbitrarily located external trips.

	Mean, Internal Trips	Mean, External Trips	% Difference in Mean	CV, Internal trips	CV, External Trips	% Difference in CV	p	Trip Fraction
USM	1.44	2.04	42%	0.28	0.87	209%	$p < 0.01$	0.23
ZIP	1.52	1.70	12%	0.40	0.55	38%	$p < 0.01$	0.56
Pixels	1.60	1.51	-5%	0.48	0.38	-26%	$p < 0.05$	0.49

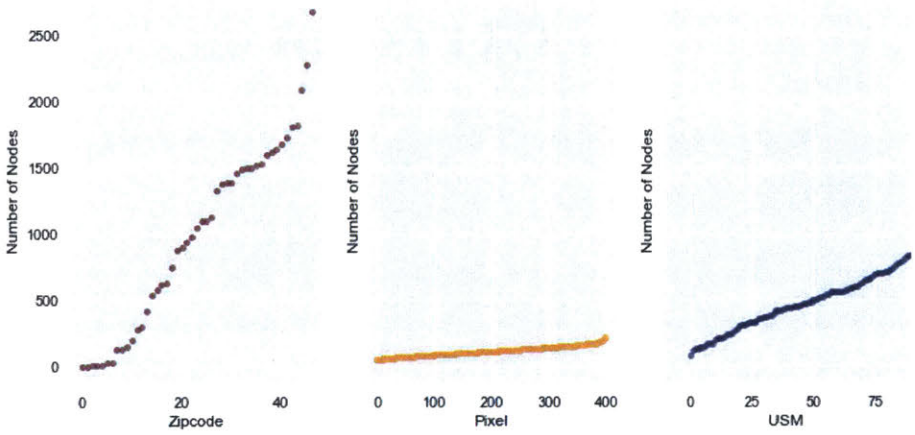
Table I.3.5: Summary of circuity and trip-based convexity results on 400 km^2 sample network

Considering each segmentation method in this order, note that the mean and CV of circuity factors measured for internal trips (origin and destination in the same segment) decrease. Pixels contain the least network information and subsequently produce segments with internal travel properties that are the least desirable of all three segmentation methods; trips inside pixels incur the highest average detours and the magnitude of those detours is

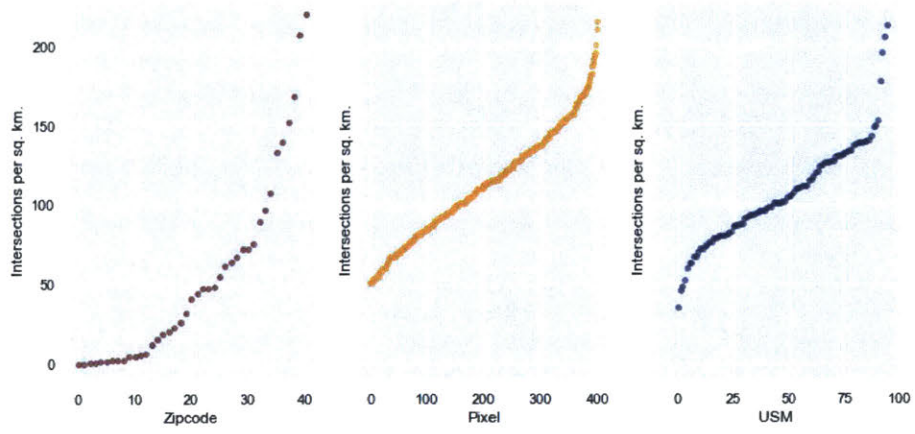
the most variable among all segmentation methods. Similarly, the circuity factors measured on external trips of the same length increase from pixels to ZIP codes to the multistage USM. In fact, for pixels, the external trips display a lower circuity factor on average than the internal trips, indicating considerable randomness and lack of travel efficiency when segments are delineated as pixels. ZIP codes and the USM, both of which capture some level of information about the geographic and travel properties of the urban street network, do not have this property and instead produce segments where internal trips are, on average, more direct with a high degree of confidence. Finally, the dispersion behaves as hypothesized: circuity of internal trips on USM segments are considerably less variable than for either of the other two methods. For both ZIP code and USM segments a lower variance in circuity factor is observed for internal trips than for external trips.



(a) Segment Area (km^2)

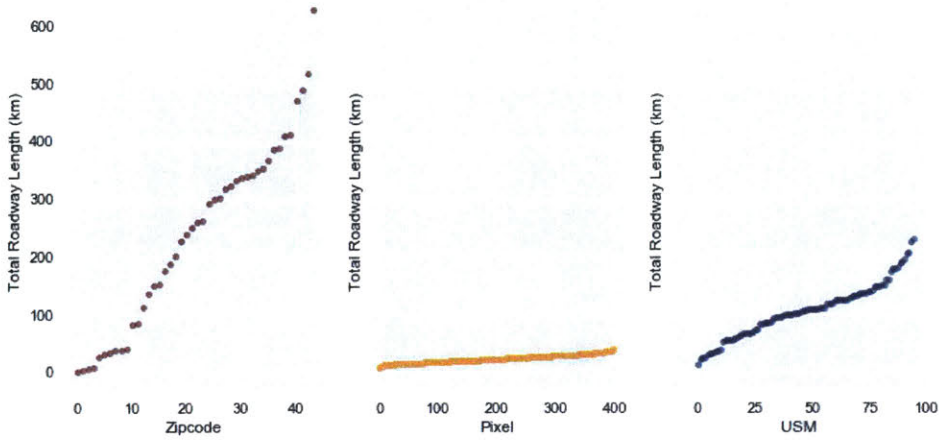


(b) Number of Nodes per Segment

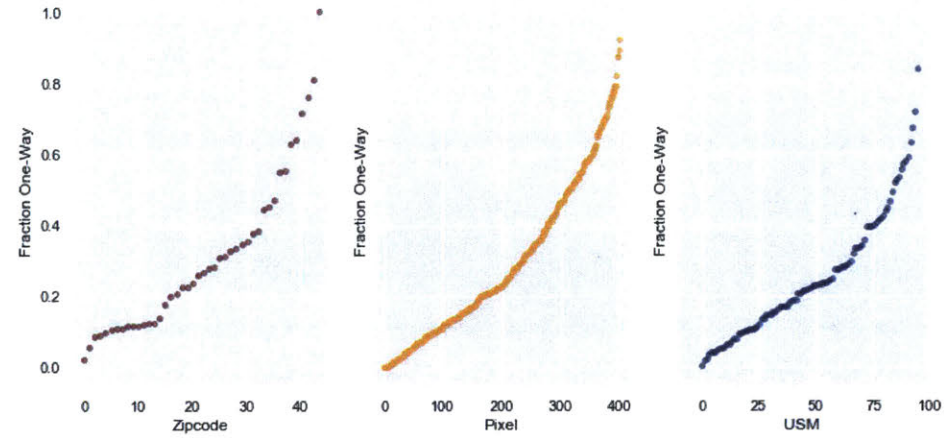


(c) Intersections per km^2

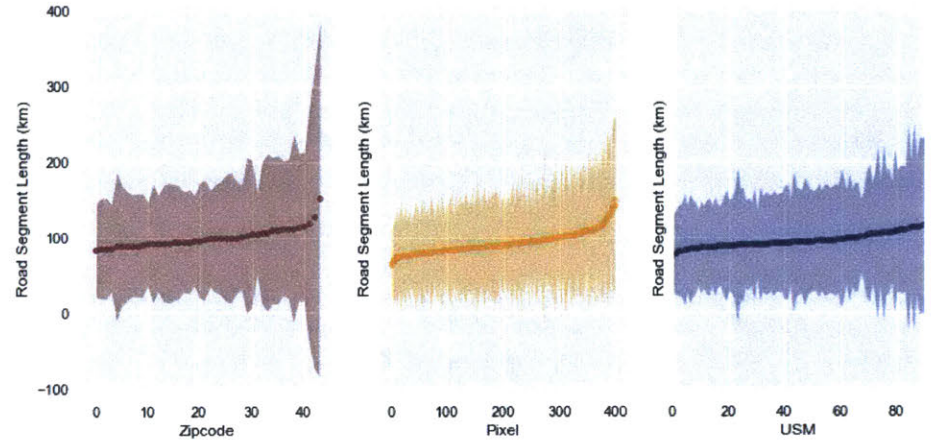
Figure I.3.6: Dimensional characteristics describing roadways of subnetworks in city segments produced by various segmentation methods (1 of 2)



(a) Total Roadway Length (km)

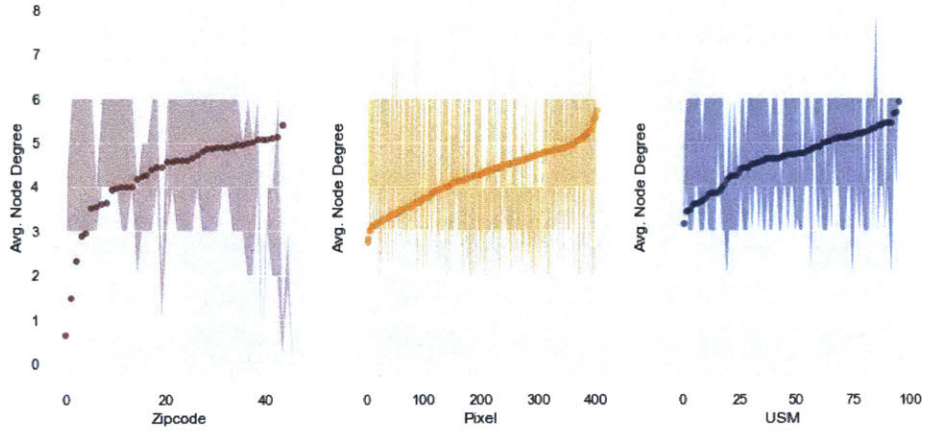


(b) Fraction of Streets Restricted One-Way

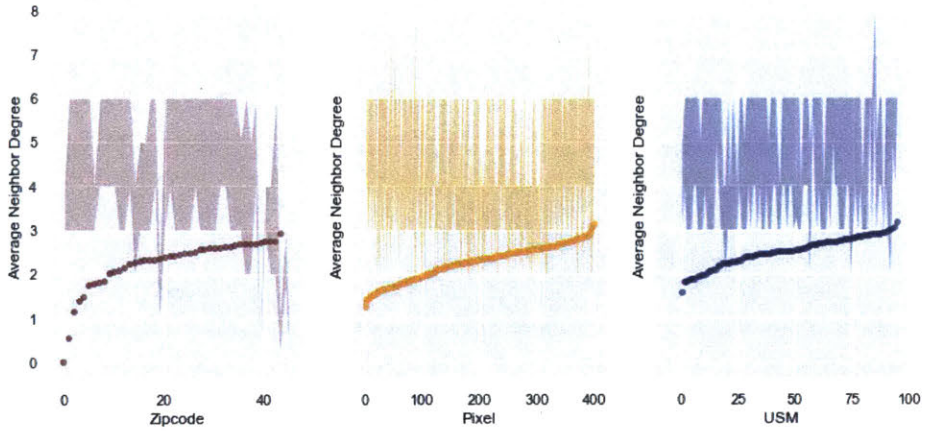


(c) Road Segment Length (km)

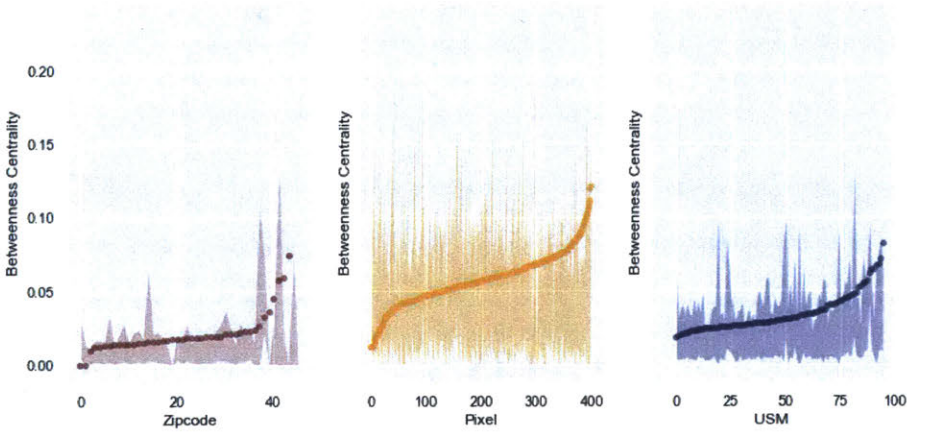
Figure I.3.7: Dimensional characteristics describing roadways of subnetworks in city segments produced by various segmentation methods (2 of 2)



(a) Node Degree

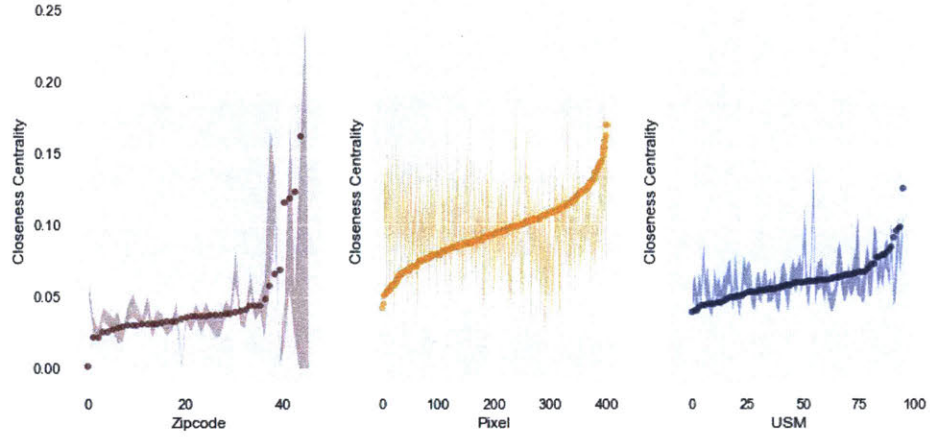


(b) Neighbor Degree

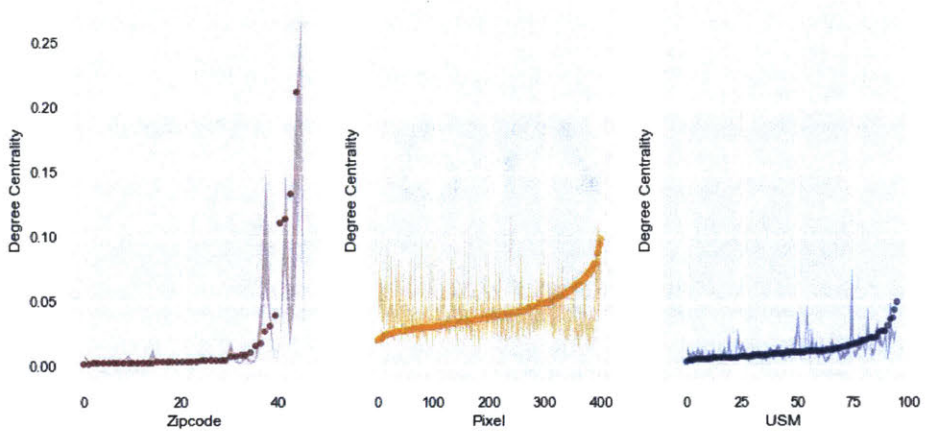


(c) Betweenness Centrality

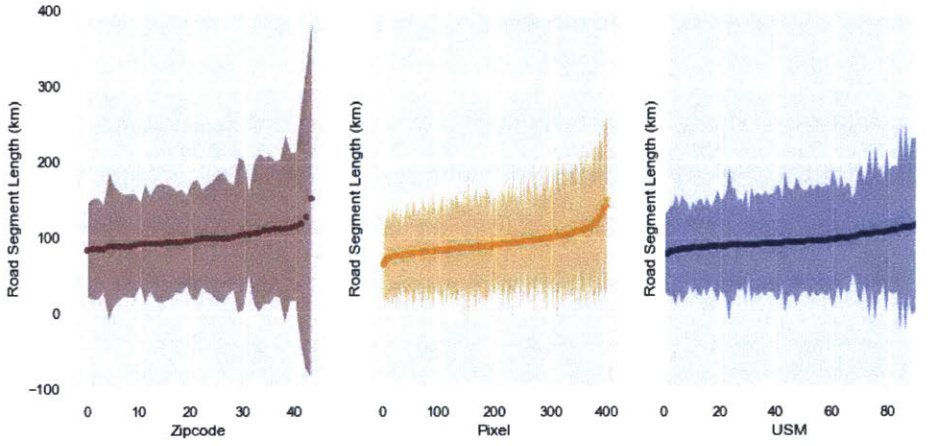
Figure I.3.8: Topological characteristics describing roadways of subnetworks in city segments produced by various segmentation methods (1 of 2). Shaded areas represent plus/minus one standard deviation of all values in each segment. Note that for all metrics presented here, values below zero are not possible.



(a) Closeness Centrality



(b) Degree Centrality



(c) Road Segment Length (km)

Figure I.3.9: Topological characteristics describing roadways of subnetworks in city segments produced by various segmentation methods (2 of 2). Shaded areas represent plus/minus one standard deviation of all values in each segment. Note that for all metrics presented here, values below zero are not possible.

In addition to the circuity experiment presented in Table I.3.5, a number of metrics describing the topological network properties and dimensional (descriptive) characteristics of the partitions are compared. Figures I.3.6, I.3.7, I.3.8, and I.3.9 display these results. Figure I.3.6 shows various measurements on the segment networks based on map database information describing the real-world network properties. Horizontal axes in these figures do not refer to fixed segment indices, but rather, for legibility, rank segments by the metric value for each plot (i.e., the n^{th} segment in one subplot does not correspond to the n^{th} segment in another). For metrics closely related to absolute segment size (e.g., Figures I.3.6 (a), (b), I.3.7 (a)), pixels show low variation between segments and ZIP codes display wide variation. This is to be expected since ZIP codes are not constrained by area while segments in the other two methods are. While the area of USM segments is closely controlled (see subplot (a)) in a range similar to pixels, note that for other size-driven measurements such as the number of nodes (b) and total roadway length (a), the USM produces dimensional metrics very different from pixels. For example, USM segments are similar in area to pixels and dissimilar to ZIP codes. The number of nodes in a USM segment is roughly similar to ZIP codes, while the node density (relative to area) roughly similar for all three segmentation methods. A possible explanation for this is that the USM segmentation is flexible enough to detect small, internally connected, externally isolated sub-networks and assign these groups of small numbers of nodes with poor accessibility and link density to their own segments while at the same time assigning large, better-connected, more easily accessed groups of nodes to their own larger communities. In other words, for the relationships observed in (b) and (c) to co-exist, segment size must be flexible enough to capture these local variations in network connectivity. The surprising result, and a supportive one for the potential of these segments to be used in practical territory-division applications, is that this flexibility is possible while still constraining segment size to a reasonable range (cf. Figure I.3.6 (a)).

Topological metrics describing the network properties, outside of the street network context, are presented for each segment subnetwork in Figure I.3.7. These measurements are calculated for nodes in each of the segment subnetworks created by each segmentation method. Shaded areas represent the interquartile range (twenty-fifth to seventy-fifth

percentiles) of measurements for each segment. Overall, there is less obvious variation between segmentation methods than seen among the dimensional metrics in Figure I.3.6. This may be at least in part due to the fact that nodes in these networks represent intersections, which, unlike edge-derived metrics or dimensional metrics, are less meaningful in the context of street networks than in other empirical networks. In other empirical networks, nodes represent objects or actors and node connectivity is a function of those objects' properties and actions, while in street networks, nodes are of secondary importance and only exist as a result of edge geometries. However, from the dispersion, it appears likely that subnetworks derived from the USM segments are less variable in terms of these topological metrics, with narrower confidence intervals, than metrics derived from other segmentation methods.

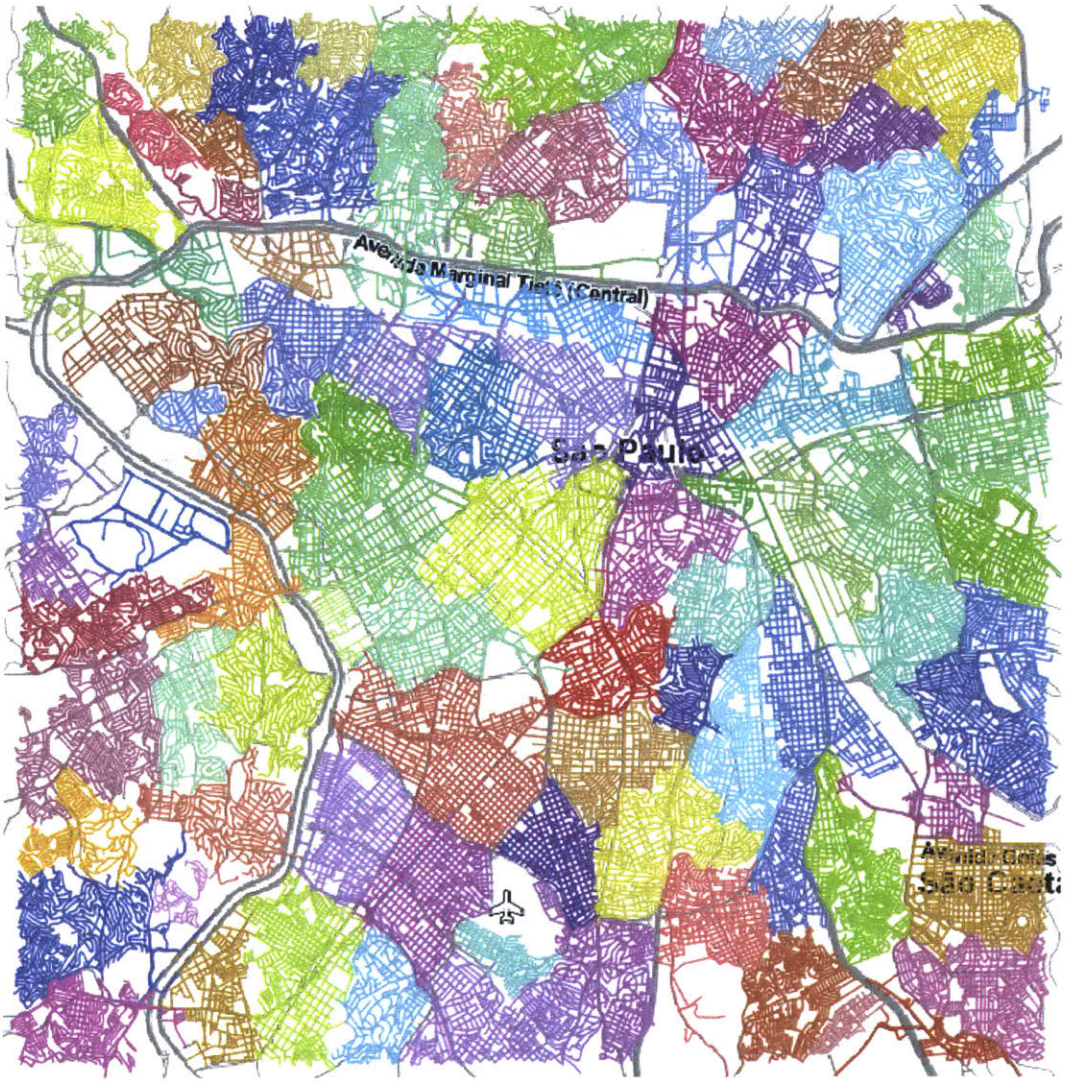


Figure I.3.10: Result of recursive USM algorithm on 400 km^2 of São Paulo local streets

I.3.5 Validation On Other Case Study Cities

To further explore the transferability of this method and the validity of results in a wider variety of settings, it is executed on test networks from a number of other cities around the world. These cities are chosen to exemplify a diversity of road network characteristics and last-mile logistics needs. The classification proposed in Winkenbach et al. [2017] is used to select candidates for these case studies. This work classifies 36 major cities worldwide into last-mile delivery market archetypes. These archetypes are developed from a set of variables chosen to characterize relevant aspects of the urban freight landscape in each location, encompassing demographics (e.g., household size, employment statistics, GDP), political and social conditions (e.g., industrial pollution levels, Economic Freedom Index, Gini Index), and indicators of wealth and industrialization (e.g., internet access, mobile phone penetration, internet retail adoption). This classification is adapted here under the hypothesis that the variables used to inform the cluster analysis also describe relevant facets of urban development and urban inhabitants. From 2015 data, the 36 cities are classified into 10 archetypes. The archetype containing São Paulo, which is demonstrated elsewhere here, is excluded and representative cities of the nine remaining clusters are chosen. While the analysis of Winkenbach et al. [2017] is a useful quantitative support for the selection of these cities, they are also chosen to qualitatively exemplify a range of street network typologies as shaped by urban geography and history. For example, Paris, Lagos, and Bangkok are all large megacities defined by major rivers and bodies of water; Amsterdam and Manhattan are both small, high-density regions with highly regular, patterned street networks (that is, Manhattan's streets form a strict grid while Amsterdam's are defined by concentric canals); and Los Angeles is a sprawling city with a fairly regular grid pattern but not constrained by geographic features such as mountains or coastline until well outside the urban core. Figures I.3.5, I.3.5, I.3.5, I.3.5, and I.3.5 display the results of executing the multi-stage algorithm on a 100 km^2 test network in each of nine example cities. Table I.3.6 repeats the circuity experiment for each of these test networks.

	Mean, Internal Trips	Mean, External Trips	% Difference in Mean	CV, Internal trips	CV, External Trips	% Difference in CV	p
Lagos, Nigeria	1.520	3.380	122%	0.3332	0.9619	189%	p <0.01
Bangkok, Thailand	1.734	2.559	48%	0.4058	0.6803	68%	p <0.01
Mexico City, Mexico	1.319	1.601	21%	0.1810	0.3692	104%	p <0.01
Paris, France	1.257	1.362	8%	0.1582	0.2423	53%	p <0.01
Los Angeles, CA, USA	1.338	1.500	12%	0.2365	0.4970	110%	p <0.01
Amsterdam, NL	1.379	1.789	30%	0.2345	0.4878	108%	p <0.01
Beijing, China	1.377	1.733	26%	0.2661	0.4718	77%	p <0.01
Bangalore, India	1.345	1.523	13%	0.1620	0.3220	99%	p <0.01
Manhattan, NY, USA	1.255	1.329	6%	0.1781	0.2699	52%	p <0.01
Brooklyn, NY, USA[3]	1.273	1.448	14 %	0.2182	0.5042	131%	p <0.01

Table I.3.6: Summary of circuity results on 100 km^2 road network samples in various cities

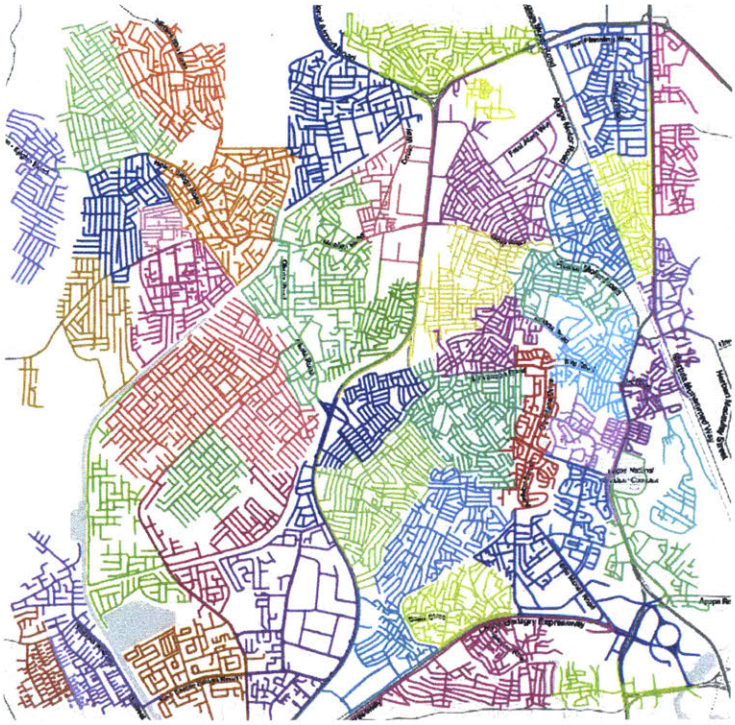
As noted, the test networks constitute a group of cities whose urban road networks are qualitatively very different in terms of their regularity, age, traffic patterns, density, and geographic constraints. Despite these differences, Table I.3.6 shows that a USM-based segmentation method successfully identifies areas where local travel is significantly more efficient within segments than between segments. This result is consistent across every test case, although the magnitude of circuity reduction varies considerably between cities.

The largest differences between internal and external mean circuity factors are observed in Lagos, Bangkok, and Amsterdam. In terms of population, Lagos is almost three times as large as Bangkok which is ten times larger than Amsterdam (21 mm vs. 8.2 mm vs. 821,000)[4]. A number of qualitative differences are apparent in the size, density, and economic characteristics of each city. However, one commonality they share is the prominence of waterways in the city center. All three are transected by large rivers, and both Bangkok and Amsterdam are largely built on land reclaimed from waterways by the historic networks of canals that cover the urban region. One hypothesis could be that the presence of these geographic features, which are a definitive obstacle to road transport, form reg-

[4]Source: http://worldpopulationreview.com/world-cities/

imented patterns in the urban road networks that the USM is able to account for. The presence of these unambiguous geographic boundaries within the city center incurs more severe detours when a trip necessitates that one of them be crossed. In the circuity experiment, crossing one of these obstacles likely constitutes an inter-segment trip since the USM segments are likely to be formed by the obstacles themselves. In cities with other network typologies where segments are not likely to be defined by "hard" obstacles, the efficiency gains may be less pronounced as better alternative routes between segments may be available. By segmenting these cities in a way that respects these obstacles, the circuity characteristics of local trips can be improved to realize gains in travel efficiency that are very high compared to a non-network-aware segmentation. A segmentation that discretizes an urban area into parts not informed by the network or geography - for example, the grid rasterization used previously - is more likely to produce local trips between end customers that cross these boundaries than a network-informed segmentation. These local trips between customers on a route would likely have similar characteristics to external trips under the USM model, where trip endpoints are chosen without respect to segment boundaries and obstacles, in the USM segmentation.

In the context of urban logistics planning, the success of this experiment across a wide range of locales and network typologies is encouraging. The segmentation model proposed here is robust enough to effectively and consistently produce gains in local travel efficiency in a variety of urban settings. This model could potentially reduce transportation costs in the last mile of delivery operations by constraining local (inter-stop) travel to these city segments by the assignment of routes to territories defined by these segments. Travel between territories could then be accomplished by the use of more efficient highways or arterial roads, or local deliveries in these areas could be assigned to different routes. Either accommodation would avoid the need for delivery operators to conduct consecutive deliveries between customers located in a sub-region of the urban transportation network with challenging transportation characteristics.



(a) Lagos, Nigeria

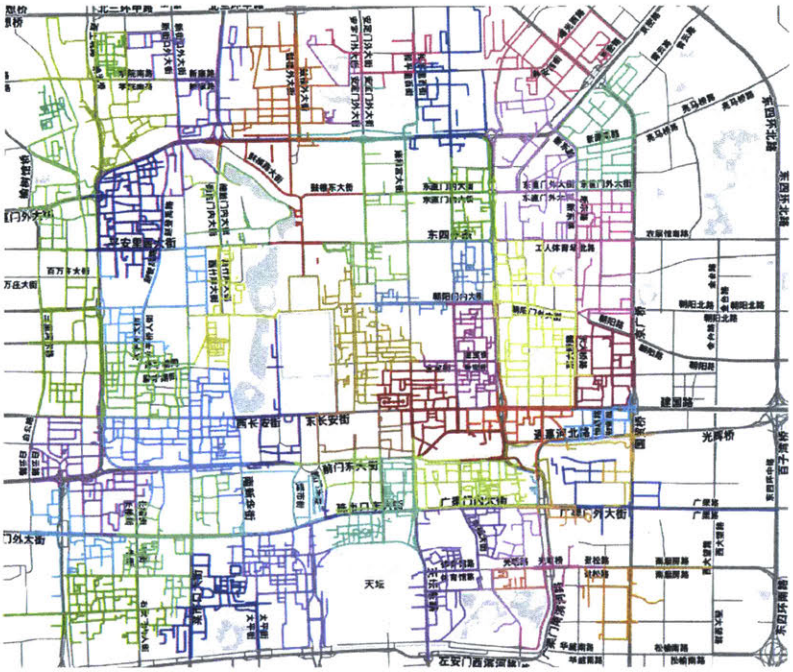
(b) Bangkok, Thailand

Figure I.3.11: Comparison of segment maps produced for test networks in global cities (1 of 5)

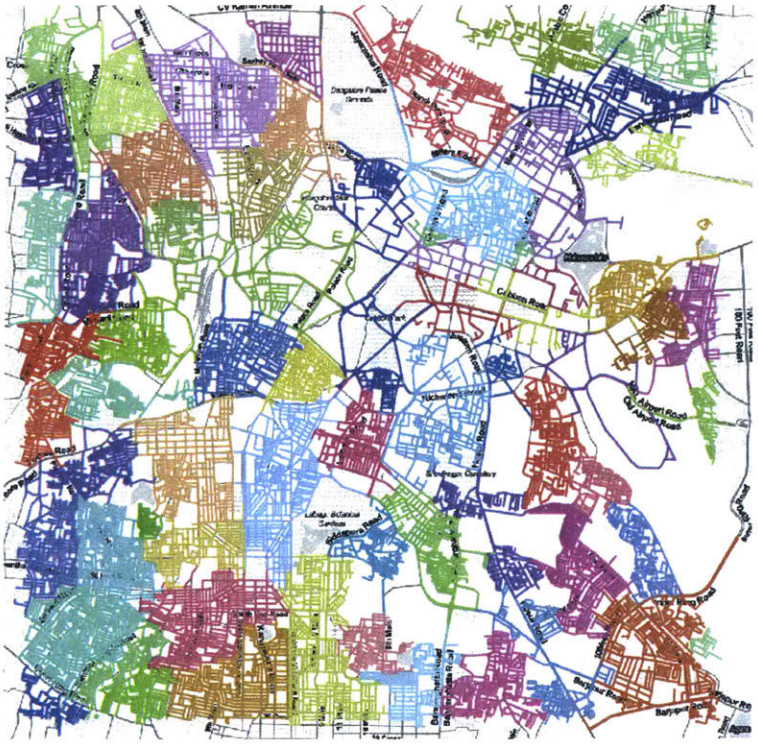
(a) Mexico City, Mexico

(b) Paris, France

Figure I.3.12: Comparison of segment maps produced for test networks in global cities (2 of 5)

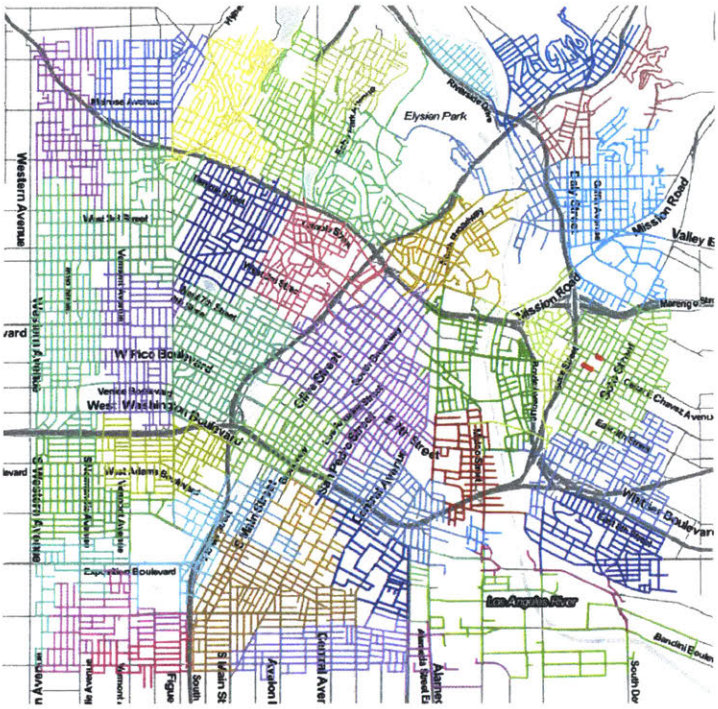


(a) Beijing, China



(b) Bangalore, India

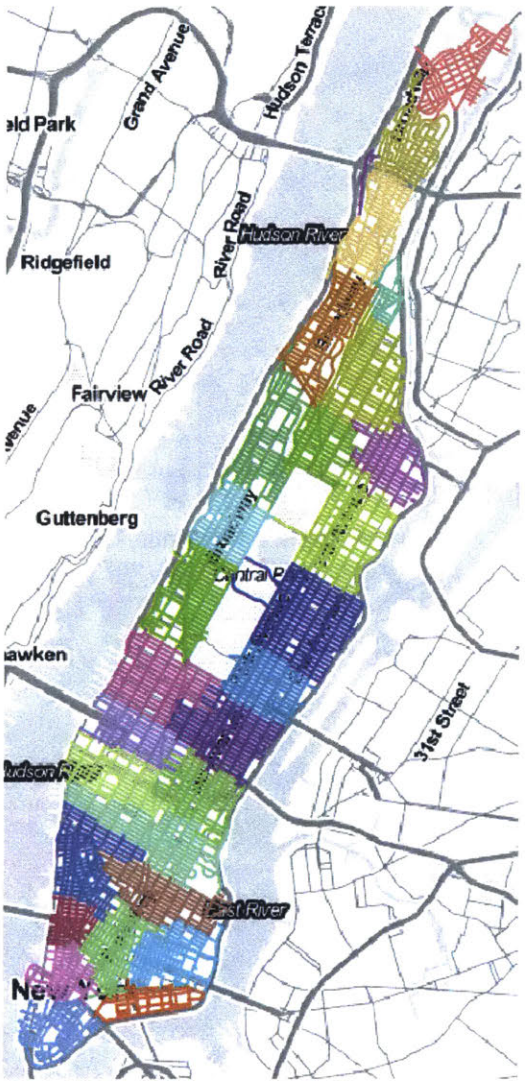
Figure I.3.13: Comparison of segment maps produced for test networks in global cities (3 of 5)



(a) Los Angeles, California, USA

(b) Amsterdam, Netherlands

Figure I.3.14: Comparison of segment maps produced for test networks in global cities (4 of 5)



(a) Manhattan, NY, USA

Figure I.3.15: Comparison of segment maps produced for test networks in global cities (5 of 5)

Chapter II

Case Study: Urban Segmentation and Freight Policy Development

This section explores the relevance of the USM's network-based zone generation to policymakers interested in mitigating certain negative outcomes associated with urban freight transportation. By overlaying the results of a USM segmentation with data on freight generation and economic activity, USM zones can become a way for policymakers to target specific address ranges of streets for policy intervention. The size of USM segments makes them feasible for targeted policy measures since they can be configured to cover areas much smaller than administrative zones such as ZIP codes. Additionally, since they consist of groups of street links that start and stop at intersections (i.e., no street changes zone mid-block), they are a natural fit for block-level, small-scale policy measures such as parking restrictions. The intuition of this approach is that USM segments reflect two pressing needs of policymakers seeking to intervene in freight-related transportation externalities: identifying clearly-defined areas where intervention is realistic and feasible, and identifying areas where the potential for nuisance is potentially greatest based on land use characteristics (e.g., those with high residential populations, small streets, or large populations of vulnerable residents). USM segments, by design, consist of subsets of local streets on which travel is not conducted via avenues or large-capacity roadways. Navigating small-capacity local streets is particularly challenging for large or heavy freight vehicles. The infrastructure of these areas poses the most constraints on the freight vehicles' operation,

leading them potentially to conduct maneuvers that are more impactful to other road users, e.g., blocking a narrow lane or double parking. Further, considering that street network typologies often reflect changes in development patterns and land uses across an urban area, the network-informed segmentation method may be able to capture patterns in land use and development better than grid-based rasterization. The aim of this analysis is to investigate if the USM segments correlated with the spatial distribution of land uses across an urban region that is more realistic than randomly-divided segments. It relies on the assumption that land use is not uniformly distributed across a city, and so the "ground truth" spatial land use distribution should not expect the same fraction of commercial, residential, office, etc. uses in every part of the city. To confirm this assumption, on which the rest of the analysis relies, point estimates of land uses mixes in the case study region are compared using a public map database.

II.1 Motivation

II.1.1 Urban Freight Policy

Especially in light of growing demand for urban freight transportation (cf. Section I.1), the environmental and human impact of this travel is of significant and compelling interest to public policymakers and sustainability advocates. Freight transportation may account for as much as 15 percent of vehicle miles traveled on urban streets and is particularly impactful in terms of fuel consumption due to the large vehicles, short trips, and frequent stops and re-starts associated with this type of travel [Dablanc, 2011]. The design of urban transportation policy to mitigate these outcomes is a complex challenge in that it must reconcile the needs of multiple stakeholders while providing an efficient flow of goods to urban destinations. On urban streets with limited parking availability and vehicular capacity, even without regulatory action, freight vehicles may be constrained in such a way that limits the efficiency and quality of logistics services. Surveys and multi-criteria studies are typical ways of studying shippers' policy resistance and attempting to reconcile the needs of multiple stakeholders present in this setting. Stathopoulos et al. [2012] offer

one such large survey detailing shippers' attitudes towards various regulatory schemes and propose an empirical multi-criteria analysis towards reconciling the needs of shippers with policy goals. Indeed, J-L. Routhier [2002] find that, particularly in urban centers, a majority of freight vehicles parked illegally while conducting deliveries, not only causing nuisance but further contributing to congestion. This underscores the need for freight policy to be targeted to specific areas where it can be most effective and most feasibly enforced.

Extant literature presents a number of policy interventions that are typically imposed at the level of a street (e.g., parking facilities) or zone (e.g., travel restrictions). Among the most common provisions is dedicated parking infrastructure for freight vehicles. This can be in the form of on-street or off-street loading zones or parking spaces, managed by business owners or public entities. In addition to the classic single-vehicle zones, sometimes pooled freight parking locations may be provided [Allen et al., 2008] where numerous vehicles can park and conduct deliveries to nearby customers. Naturally, these measures require identifying and dedicating space on a particular street segment and are made more effective by incorporating some knowledge of nearby demand for freight services into the location decision. Stathopoulos et al. [2012] describe policies proposed to mitigate freight problems in six categories: market-based measures such as congestion pricing and tolling; regulatory measures, e.g. an enforced delivery time-window system; land use and zoning efforts restricting the interference between commercial and non-commercial uses where possible; infrastructural measures such as consolidating shipments and encouraging alternative modes of transport; technologies that share information between relevant parties; and management measures aimed at promoting cooperation between stakeholders at a high level. With the exception of the latter two policy archetypes, each of these classes of policies can be targeted to a specific zone or neighborhood. It is noted that the case study below is conducted on the borough of Brooklyn, New York City.

While relatively little news has been generated about freight restriction policies in Brooklyn (although they exist in various forms at various levels of stringency), a 2018 recommendation made by New York State government task force Fix NYC[1] to begin an ag-

[1]Fix NYC Report (January 2018): http://hntb.com/HNTB/media/HNTBMediaLibrary/Home/Fix-NYC-Panel-Report.pdf

gressive truck tolling regimen in lower Manhattan generated considerable discussion about the benefits and drawbacks of such a regulation.[2] The NYC DOT has a dedicated Office of Freight Mobility, and a number of other initiatives focused on active freight management have been sponsored by various City offices. For example, a three-year pilot study[3] of off-peak deliveries was conducted between 2007-2010 at over 400 business locations. Thus, collectively, and perhaps not surprisingly given the city's scale and density, New York City transportation engineering and policy agencies have an established record of active management of freight traffic. This, alongside the rich mix of land uses and ongoing history of development and re-development in Brooklyn particularly, motivate the selection of Brooklyn as the setting for this case study.

II.1.2 Factors Impacting Demand for Freight Transportation

Targeted application to well-defined zones is one way to make urban policy implementations more tractable and effective. A complementary land-use dataset could further identify which urban segments display certain relevant land-use characteristics and help shape decisions on whether to implement or forgo certain policy measures. A number of important drivers of freight activity could be described under the umbrella of "land use characteristics" - for example, how many shops exist in a certain zone, their sizes, and what type of business they conduct. Cherrett et al. [2012] summarize a few of the characteristics that drive freight activity and would be of interest for policymakers seeking to understand spatial and temporal patterns in freight. Namely, this study notes business category (e.g., clothing retailer vs. food services), physical store area, and supply chain system as important determinants of the number of deliveries made to a store. Sánchez-Díaz et al. [2016] propose a model of freight traffic attraction and similarly take each business' category and size as inputs to the traffic generation model. Data for this model are taken from a survey conducted by the authors, where establishment category is described as wholesale trade, retail, or accommodation and food and establishment size is described in terms of

[2]For example, a collection of NY Times articles on the subject can be found at https://www.nytimes.com/topic/subject/congestion-pricing

[3]https://www1.nyc.gov/html/dot/downloads/pdf/truck-deliveries-ll189.pdf

a number of employees (as opposed to building size). In their analysis of their estimation results, the authors note the importance of having onsite inventory storage, more common among some business types than others, as a noteworthy determinant of traffic generation. In practitioner-oriented activity, a 2017 report jointly funded by the NY State Research and Development Commission and the NY State Department of Transportation [Lubinsky et al., 2017] focuses primarily on loading/unloading bays and presents a comprehensive study of current NYC loading/unloading regulations and their relationship to the buildings and zoning denominations they govern. For example, in NYC as in many peer cities, the study notes that buildings are required to provide a number of freight loading areas relative to the floor area dedicated to a particular zoning designation. The study targets three neighborhood-level segments of Manhattan for detailed analysis in which zoning designations are divided into residential, retail, and office space (as done in this work) for determination of freight facility requirements in each district. One reason for the selection of NYC as a case study setting is the public availability of city-maintained data at the building level, which can provide the square meters of building area in categories at this level of detail, although not at the level of specific retail category as might be ideal.

II.2 Case Study: Borough of Brooklyn, NY, USA

A case study is conducted on Kings County, NY (coincident to the borough of Brooklyn) to determine if the network-based urban segmentation model introduced here is capable of producing segments with distinct land use characteristics relevant to freight activity and policy. The motivation of this is that if the segmentation algorithm does identify network segments where land uses are distinct from other segments, these segments can then be used as a tool for policymakers to target areas where policy interventions would be most appropriate. Targeting policy interventions to small districts where freight activity patterns are most salient may be one element in helping these policies be successful and politically tenable.

II.2.1 Data Sources

Brooklyn is chosen as the location for this case study in part based on the availability of public data provided by various New York City offices. With a population density of 14,716 persons per km^2 and a land area of 180 km^2, it also provides a sufficient population density to make traffic policy intervention relevant and an area large enough that targeted sub-districts would have to be developed to effectively manage any policy intervention.

To describe land uses in this region, the MapPLUTO shapefile dataset[4] of land use and geographic data is obtained from the NYC Department of City Planning. This dataset is free, publicly available, and provides zoning, building area, and land use data at the tax lot (typically, equivalent to building) level. Over 250,000 buildings are mapped in Brooklyn alone. Fields used from this dataset are summarized in Table II.2.1. With the exception of the field "Overlay1," which comes from NYC Department of City Planning GIS, all fields are calculated from NYC Department of Finance property appraisal data. More detailed descriptions of these fields, their original sources, and collection methods are available in the PLUTO project documentation.[5] As in previous studies, OSMnx via Python is used to query the road network and highways are removed. Unlike previous studies presented here, the query is performed on the administrative region of Brooklyn (Kings County, NY) and not with a rectangular bounding box. (Both query types are supported by the OSMnx library).

II.2.2 Analysis

The multistage urban segmentation algorithm is executed on the road network of Brooklyn with high-volume road classifications removed. The maximum segment size allowed is 10 km^2 and the ideal segment size is 1 km^2. Figure II.2.1 shows the street labeling output by this execution. For each segment in this result, a polygon is approximated by calculating the minimum non-intersecting α-shape as described in Section I.3.2 and the area of this shape

[4]PLUTO home page: https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page

[5]PLUTO Data Dictionary: https://www1.nyc.gov/assets/planning/download/pdf/data-maps/open-data/pluto_datadictionary.pdf?v=18v2beta

Field Name	Description
ComArea	An estimate of the exterior dimensions of the portion of the structure(s) allocated for commercial use
ResArea	An estimate of the exterior dimensions of the portion of the structure(s) allocated for residential use
OfficeArea	An estimate of the exterior dimensions of the portion of the structure(s) allocated for office use
RetailArea	An estimate of the exterior dimensions of the portion of the structure(s) allocated for retail use
GarageArea	An estimate of the exterior dimensions of the portion of the structure(s) allocated for garage use
StorageArea	An estimate of the exterior dimensions of the portion of the structure(s) allocated for storage or loft purposes
FactoryArea	An estimate of the exterior dimensions of the portion of the structure(s) allocated for factory, warehouse, of loft purposes

Table II.2.1: Field descriptions from NYC Department of Planning PLUTO Dataset used in Brooklyn case study

is measured. Using QGIS, an open source GIS platform, the USM segments are mapped against the PLUTO building dataset and a spatial join is performed to aggregate the total building area in each of the seven categories described in Table II.2.1 that is recorded for buildings which overlap the segment polygon. Buildings that partially overlap a segment are included in this aggregation, and any building that intersects two polygons would be counted towards the total area of both. The average area of segments produced from this method is calculated (approximately 3 km^2) and a grid of square pixels with the average segment area is laid over the study region. The spatial join is then repeated for the pixel layer so that each pixel is described by a value that represents the total building area in square feet dedicated towards each land use in Table II.2.1 within that pixel. Square feet are then converted to square meters to maintain a consistent measurement system. The results of these two spatial joins are saved as shapefiles and further analysis is conducted in Python. Figure II.2.2 shows the superposition of pixels and segment-based polygons on the Brooklyn study area.

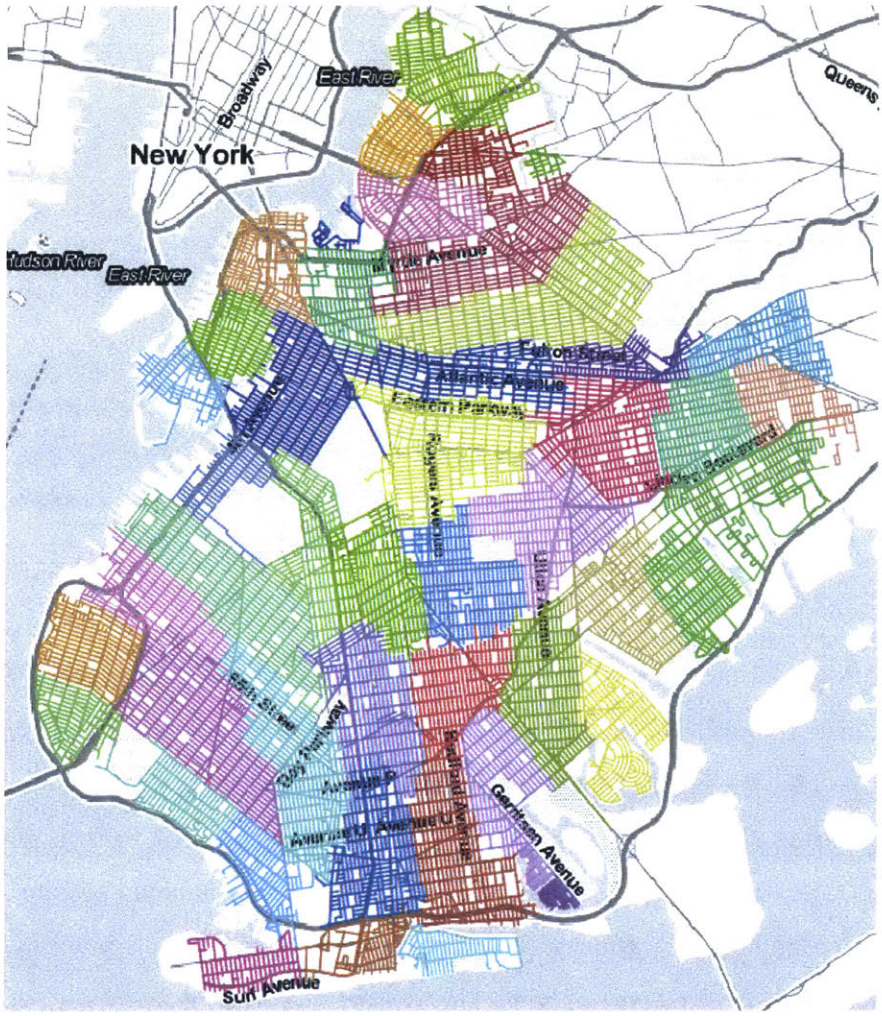


Figure II.2.1: Map of segments identified by USM algorithm in Brooklyn, NY

II.2.3 Results

The distribution of land uses across city segments is compared between the USM, pixel, and ZIP code segmentation models. ZIP code data is obtained as a shapefile from the United States Census Bureau[6]. For each segment, a measurement is made of the total square meters of building area dedicated towards certain freight-relevant land uses. This is normalized by the area of each segment so that measurements of square-meters-per-square-kilometer dedicated towards a specific land use category can be compared between segments as a measure of intensity rather than absolute magnitude. For each category, the distribution of

[6]Source: https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html

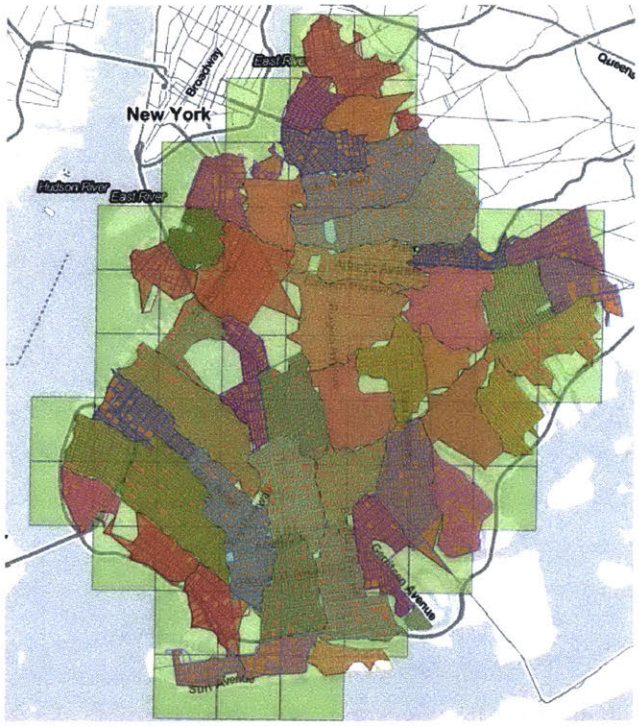


Figure II.2.2: Pixels and USM-based polygons in Brooklyn study area

this metric is compared between pixel and USM segments. Under the hypotheses that true land use patterns are not uniformly distributed across the city, and that spatial variations in land use are correlated with spatial variations in road network topology, a rasterization method that is uncorrelated with network variations should also depict a relatively uniform distribution of land uses in each segment.

By the same assumptions, the more heavily the segmentation draws on street network information, the more accurately it should capture local variations in streets and neighborhoods where a certain land use is more or less intense. In other words, each of the land use intensity metrics should display the greatest variation for USM segments, an intermediate variance for ZIP codes, and the lowest variance for pixels. Figure II.2.3 and Table II.2.2 summarize the distributions of land use intensity in each of the categories provided by the PLUTO dataset. From Figure II.2.3, it seems that ZIP codes often produce a distribution similar in shape to pixels, despite the former having been formed at least partially on the basis of road network information. In these figures, segments are sorted within their segmentation method by land use intensity in decreasing order and plotted. Pixels, which

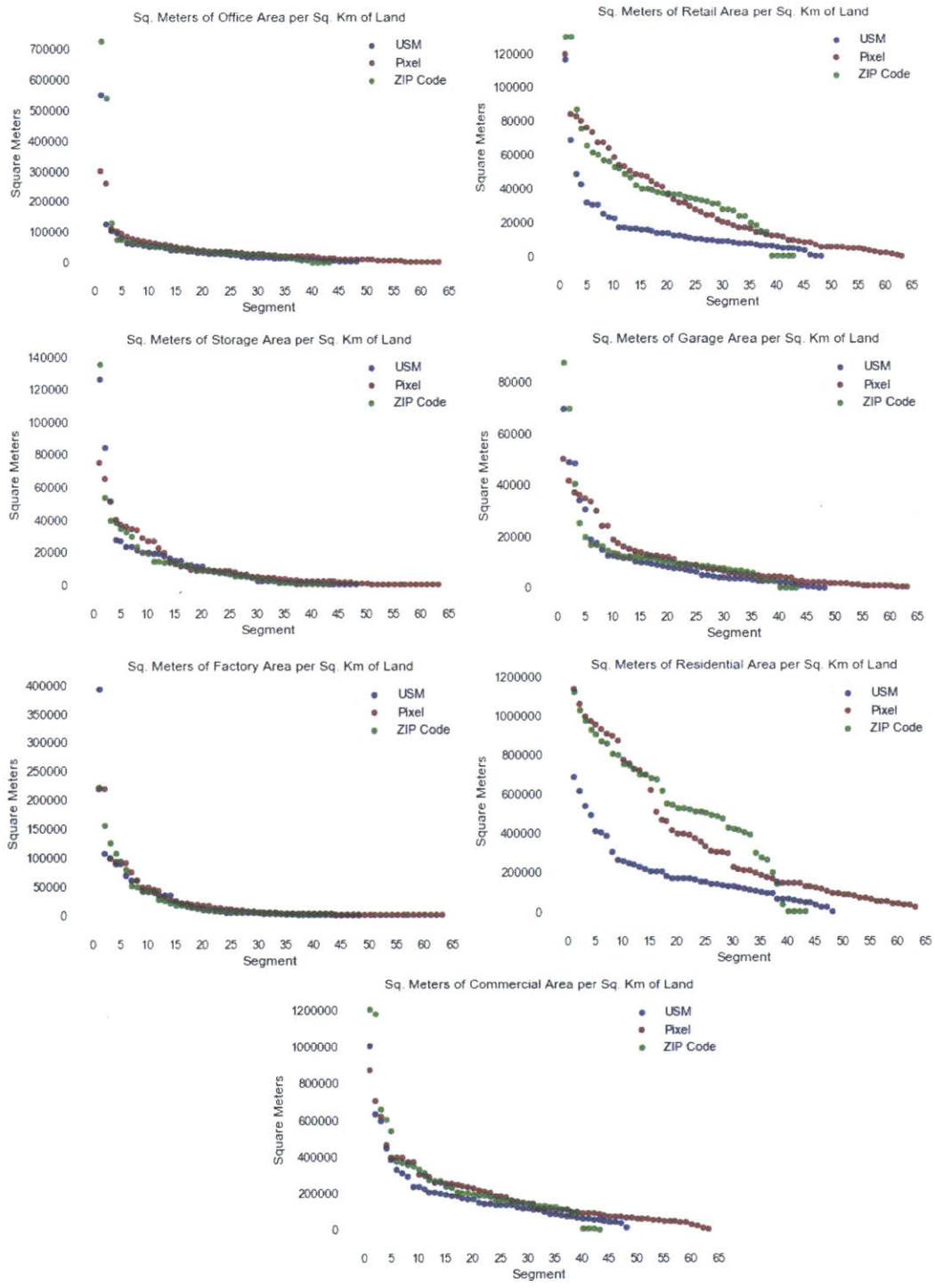


Figure II.2.3: Distributions of dedicated square meters of floor area per square kilometer of land, separated by land use category

divide the city arbitrarily relative to the grid origin and pixel size, fail to group blocks or streets with high intensities of certain land uses together. As a result, the land uses within pixels are mixed and the area measured in a given land use tends to decrease from the maximum more gradually than it does for USM segments. Surprisingly, a similar shape is observed in ZIP codes, which might be expected to behave more similarly to USM segments since they are supposedly formed partially on the basis of natural and manmade boundaries. Particularly for categories particularly relevant to freight traffic generation and externality, the three segmentation models produce distributions in which ZIP codes and pixels behave similarly to each other, in distributions with notably more gradual decline than the USM segments (see Retail, Residential, and, to a lesser extent, Commercial subplots in Figure II.2.3).

Table II.2.2 compares the standard deviations and coefficients of variation for the USM and pixel land use distributions. Measures of variation are used to characterize the shape of the distributions observed in Figure II.2.3. For more uniform distributions, more segments will have a value closer to the mean and thus, the distribution will have a lower variance. For distributions in which more intensities are observed near the extremes, fewer observations will be near the mean and a higher variance will be observed. Standard deviation is chosen as the metric of variation to preserve the effects of outlier high or lower intensity segments which might be smoothed by other variance measures. While the building areas per category are normalized by segment area, it is noted that there are still significant differences in the mean land use intensity of many categories between segmentation models. For this reason, the CV, or mean-normalized standard deviation, is also reported in Table II.2.2.

USM segments have the largest standard deviation for only one category out of seven, while ZIP codes have the largest standard deviation for five categories of seven. However, for five of seven land uses, USM segments have a larger standard deviation than pixels and the largest CV in four of seven categories. Likewise, pixels have the smallest standard deviation for five of seven land use categories and the smallest CV in four of seven. While these results are mixed, they do not rule out the possibility that ZIP codes and USM segments behave similarly in terms of their sensitivity to local changes in land use intensity

		Std. Dev.	CV	Max. SD	Max. CV	Min. SD	Min. CV
Commercial	USM	177,022	0.963415	ZIP	ZIP	Pixel	Pixel
	Pixel	166,803	0.906238				
	ZIP	251,709	0.989025				
Residential	USM	152,460	0.796335	Pixel	Pixel	USM	ZIP
	Pixel	318,602	0.898279				
	ZIP	288,868	0.544579				
Office	USM	78,698	1.83949	ZIP	ZIP	Pixel	Pixel
	Pixel	50,992	1.29963				
	ZIP	129,077	2.03257				
Retail	USM	19,244.5	1.17065	ZIP	USM	USM	ZIP
	Pixel	26,503	0.935816				
	ZIP	27,857	0.70899				
Garage	USM	13,766.8	1.32626	ZIP	USM	Pixel	Pixel
	Pixel	11,418.2	1.12851				
	ZIP	16,164.5	1.2391				
Storage	USM	22,026.4	1.61268	ZIP	USM	Pixel	Pixel
	Pixel	16,008.1	1.4367				
	ZIP	22,452.9	1.55927				
Factory	USM	60,515.2	2.14227	USM	USM	Pixel	ZIP
	Pixel	43,893	1.89973				
	ZIP	46,468.2	1.6518				

Table II.2.2: Summary statistics of land use distributions for USM, ZIP code, and pixel-derived segments

by category, and that a network-informed segmentation method is better at characterizing these patterns than pixels. This could potentially be useful in guiding a policy decision which would need to, for example, identify a group of streets in which commercial activity is most dense, or in which dense commercial activity coincides with dense residential zoning.

As an additional investigation and also a ground-truth test of the density measurements derived from the PLUTO dataset, the Google Places API[7] is used to examine some high-intensity segments in more detail. The USM segment in which Commercial zoning density is highest also, incidentally, contains the highest density of Office area and represents an area located near the Brooklyn and Manhattan Bridges. Google-defined neighborhoods comprising this densely settled area include Brooklyn Heights and DUMBO. The USM segment, shown in Figure II.2.4, also coincides closely with ZIP code 11201.

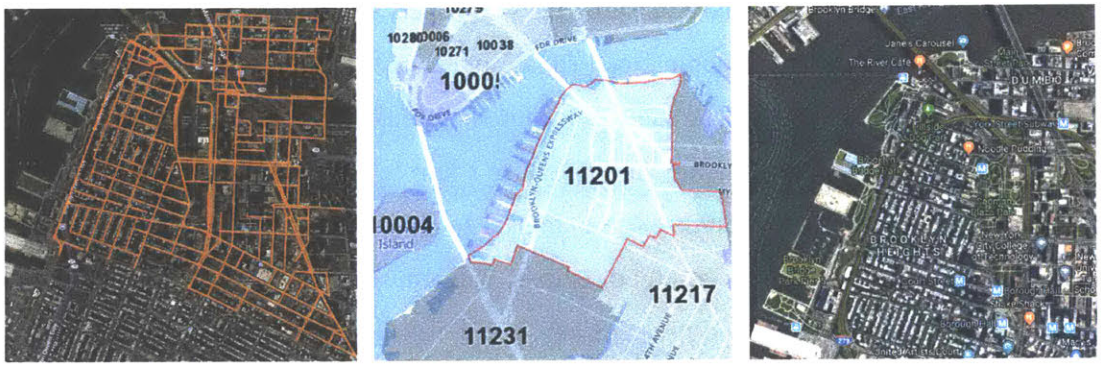


Figure II.2.4: Maximum-intensity office/commercial segment colored on road map, next to its coinciding ZIP code 11201, and with Google neighborhoods labeled

To more closely investigate the density of Commercial activity and compare it to an additional measurement of commercial establishment density, a simple experiment is conducted with the Google Places API via Python. For a circle originating at a point located roughly in the center of a city segment, the API returns all mapped establishments and points of interest, which may include businesses, offices, government agencies, and locations such as parks and administrative designations. To ensure that only commercial establishments are returned, the query is narrowed by a list of types[8] (as specified in the

[7]Documentation: https://developers.google.com/maps/documentation/javascript/places

[8]The following place type designations are queried: store (general), cafe, ATM, bakery, bank, bar, beauty

API documentation). These types are chosen to describe a broad range of commercial establishments while still filtering the query for geographic designations such as "Brooklyn, NY" or points of interest such as parks, which are also considered landmarks in Google Places. From a point selected roughly in the center of the maximum-intensity commercial USM Segment (i.e., the segment with the highest square-meters-per-square-kilometer of commercial floor area), a query is made for every establishment within a 500-meter radius belonging to the specified categories. There are 542 establishments returned in this radius. For comparison, the median-intensity segment (i.e., that which has the median value of square-meters-per-square-kilometer) is also selected and the same procedure is repeated. Because the number of segments generated in Brooklyn is even, the segment with the next highest commercial intensity above the true median is chosen to provide a conservative estimate. For the 500-meter radius around a point at the center of this segment, the same query returns 267 results. Repeating the same query for the maximum- and medium-intensity pixels (in terms of commercial square-meters-per-square-kilometer) returns 222 and 241 shops, respectively. The same procedure is repeated for zip codes, but in this case the maximum and median among only zip codes 3 km^2 in area or larger is considered (on inspection, the Census Bureau dataset contains a few ZIP codes comprised only of highway interchanges and largely unsettled regions, with unusually small areas that contribute to unrealistic estimates of land use intensity based overlap of surrounding buildings). A query originating at the center of the maximum-intensity ZIP code returns 668 points and the medium-intensity ZIP code returns 263 points.

The three maximum-intensity segments overlap and represent roughly the same area, as shown in Figure II.2.5. However, it is observed that for pixels the median-intensity pixel actually returns more commercial establishments in the Places query than this maximum-intensity pixel. This supports the hypothesis that randomly located raster zones are largely uncorrelated with land uses. In fact, observing Figure II.2.5, a more favorable result might be achieved by shifting the centroid of the maximum-intensity pixel to more closely match

salon, bicycle store, book store, clothing store, convenience store, department store, electronics store, florist, furniture store, gas station, gym, hair care, hardware store, home goods store, jewelry, laundry, liquor store, night club, meal delivery, meal takeaway, pet store, pharmacy, real estate agency, restaurant, shoe store, shopping mall

the ZIP and USM segments pictured, but this is precisely the weakness of the pixel grid generated from an arbitrarily-located grid origin which does not permit this translation.

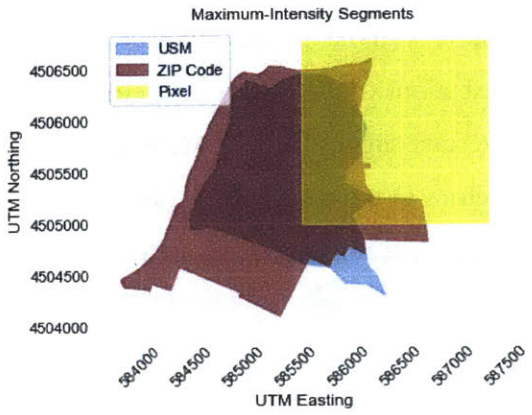


Figure II.2.5: Maximum-intensity office/commercial segments under each segmentation model

Together, these results suggest that network-based partition methods create partitions that concisely identify regions of coherent land use mixes in a way that may be useful in urban planning applications. While the proper metrics for describing this outcome are difficult to formulate, a variety of comparative analyses presented here suggest that USM segments and ZIP codes may produce zones of similar shape, size, and land-use characteristics that confer various advantages in the urban planning and policy setting relative to pixels. When the need for reproducibility and flexible scale is considered, USM segments may be favorable to ZIP codes. This case selects a high-density region of a major city and performs the algorithmic segmentation of the local street network. The algorithmically generated city segments are explored for characteristics relevant to policy interventions regulating freight activity. Unlike manually selected streets or neighborhoods, these segments do not rely on existing knowledge of the urban landscape. Unlike ZIP codes, which may be hierarchically sized but variable in terms of resolution and often quite large even in densely settled areas, these segments can be sized appropriately to the specific circumstances of a planned intervention. By layering these divisions with a publicly available database of building-level land uses that describes the area in each building dedicated towards residential, office, retail, storage, and other categories, a proxy metric for the amount of freight

activity anticipated in a given zone is calculated. The patterns of this metric across all zones are then compared across the network-based segments, similarly-sized square pixels, and ZIP codes. Square pixels, formed without any correlation to the road network, are shown to have very little correlation with the interrelated patterns of transportation and land use. By investigating the distributions of the calculated metric and comparing rough estimates of commercial activity, results suggest that USM segments result in distributions of the co-vary with land-use metrics in a manner that is distinct from and more accurate than a segmentation model that does not consider street network properties.

Chapter III

Conclusions

This thesis examines the methods of community detection used in the study of a wide variety of relational networks and adapts them to spatial networks of urban streets. The goal of this adaptation is to produce an algorithm that can reliably and efficiently segment urban areas into regions of reasonable size, the bounds of which are determined by a user relative to the problem setting. Further, this segmentation should be based on road network topology such that the resulting segments are areas of the city within which local travel has improved efficiency and reliability compared to extant methods of urban segmentation. Tailored to the context of adapting these segments to a model of logistics transportation, particularly one in which a continuum approximation might be employed, the efficiency and reliability of inter-stop transportation is quantified by the mean and variance, respectively, of the circuity factor metric which describes the detour required to make local trips inside segments and between segments. In addition to consistently producing segments in which local travel has properties which are very desirable in terms of circuity factor, this method's basis in the street network ensures that areas of the city such as parks, waterways, and uninhabited regions are not considered for segmentation, making this method readily adaptable to diverse urban geographies. In addition to exploring the feasibility of these segments' use in an operational model of freight distribution, the segmentation model is also explored as a decision support tool for policies regulating urban freight. In a case study of Brooklyn, NY, the segmentation model is compared to land use datasets describing the intensity of various activity types in each segment of the study region. Demand profiles

and externalities derived from freight services vary depending on the types of customers and stakeholders present in a zone, and so these activity profiles are expected to have some relevance to policy makers aiming to understand and regulate urban freight activities.

Several opportunities for the extension of this method are suggested. Instead of abstracting groups of street links into jagged, approximated α-shapes, an alternative method of "blending" the edges of USM segments into regular polygons using the geometry of the street network can be developed. This could be attained by, for example, moving each edge of each convex hull to the closest matching street link and then iteratively adjusting all edges of all polygons until none overlap and no spaces are left out. This is non-trivial and requires significant spatial post-processing that may jeopardize the topological properties of the partition sub-networks, but may also result in segment polygons that are visually and practically more tractable than the α-shapes used here. The algorithm could also be extended to account for both an upper and lower bound where presently it imposes only an upper bound on segment size. Further, due to abnormal street geometries, some isolated road links become disconnected as the network is filtered and segmented. These are omitted from the visualizations created here and are not relevant in the macroscopic scale. A pre- or post-processing method can be developed to re-combine these isolated links with adjacent segments as appropriate while ensuring that integrating these links does not affect the size or circuity attributes of the resulting segment. Depending on the reliability of street metadata such as link lengths, edge weights could also be used in the segmentation model and possibly subjected to some numerical transformations that might improve the resulting segment properties. In addition to these incremental improvements, an extended segmentation algorithm could be developed that would explicitly consider external information such as the intensity of demand for deliveries or the number of commercial establishments in a segment, for example. This could be accomplished by assigning a weight to network links and establishing a criterion for the total weight of a community. Land-use data similar to that used in Chapter II may be useful here. This would result in a segmentation method that not only considers network topology but also measurable characteristics of human activity and predictors of freight traffic generation in an urban zone. A method that combines these two elements might be of interest for both freight planning and policy applications.

Bibliography

J. Allen, M. Browne, T. Cherrett, and F. McLeod. Review of UK Urban Freight Studies. Technical report, University of Westminster and University of Southampton, London, U.K., 2008. URL https://westminsterresearch.westminster.ac.uk/item/91114/review-of-uk-urban-freight-studies.

Ronald H. Ballou, Handoko Rahardja, and Noriaki Sakai. Selected Country Circuity Factors for Road Travel Distance Estimation. *Transportation Research Part A: Policy and Practice*, 36(9):843–848, 11 2002. ISSN 0965-8564. doi: 10.1016/S0965-8564(01)00044-1. URL https://www.sciencedirect.com/science/article/pii/S0965856401000441.

Albert-László Barabasi and Márton Pósfai. *Network Science*. Cambridge University Press, Cambridge, 2016. ISBN 978-1-107-07626-6 1-107-07626-9. URL http://barabasi.com/networksciencebook/.

Marc Barthélemy. Spatial Networks. *Physics Reports*, 499(1):1–101, 2 2011. ISSN 0370-1573. doi: 10.1016/j.physrep.2010.11.002. URL http://www.sciencedirect.com/science/article/pii/S037015731000308X.

Stefano Battiston, Michelangelo Puliga, Rahul Kaushik, Paolo Tasca, and Guido Caldarelli. DebtRank: Too Central to Fail? Financial Networks, the FED and Systemic Risk. *Scientific Reports*, 2(1):541, 12 2012. ISSN 2045-2322. doi: 10.1038/srep00541. URL http://www.nature.com/articles/srep00541.

Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast Unfolding of Communities in Large Networks. *J. Stat. Mech.*, P10008, 2008. doi: 10.1088/1742-5468/2008/10/P10008.

Geoff Boeing. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65:126–139, 9 2017. doi: 10.1016/j.compenvurbsys.2017.05.004. URL https://doi.org/10.1016/j.compenvurbsys.2017.05.004.

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 239980831878459, 8 2018. ISSN 2399-8083, 2399-8091. doi: 10.1177/2399808318784595. URL http://arxiv.org/abs/1705.02198.

Ulrik Brandes, Daniel Delling, Marco Gaertler, Robert Görke, Martin Hoefer, Zoran Nikoloski, and Dorothea Wagner. Maximizing Modularity is Hard. Technical report, University of Konstanz, Karlsruhe University, Charles University, 2006. URL https://arxiv.org/pdf/physics/0608255v2.pdf.

Thomas Brinkhoff. Measuring the Complexity of Polygonal Objects. Simulation Framework for Continuous Phenomena View project. Technical report, University of Münich, 1995. URL https://www.researchgate.net/publication/221589831.

Damon Centola. The Spread of Behavior in an Online Social Network Experiment. *Science (New York, N.Y.)*, 329(5996):1194–7, 9 2010. ISSN 1095-9203. doi: 10.1126/science.1185231. URL http://www.ncbi.nlm.nih.gov/pubmed/20813952.

S. H. Y. Chan, R. V. Donner, and S. Lämmer. Urban road networks — Spatial Networks with Universal Geometric Features? *The European Physical Journal B*, 84(4):563–577, 12 2011. ISSN 1434-6036. doi: 10.1140/epjb/e2011-10889-3. URL https://doi.org/10.1140/epjb/e2011-10889-3.

Tom Cherrett, Julian Allen, Fraser Mcleod, Sarah Maynard, Adrian Hickford, and Mike Browne. Understanding Urban Freight Activity: Key Issues for Freight Planning. *Journal of Transport Geography*, 24:22–32, 2012. doi: 10.1016/j.jtrangeo.2012.05.008. URL http://dx.doi.org/10.1016/j.jtrangeo.2012.05.008.

Kevin L. Clark, Udit Bhatia, Evan A. Kodra, and Auroop R. Ganguly. Resilience of the U.S. National Airspace System Airport Network. *IEEE Transactions on Intelligent Transportation Systems*, 19(12):3785–3794, 12 2018. ISSN 1524-9050. doi: 10.1109/TITS.2017.2784391. URL https://ieeexplore.ieee.org/document/8306449/.

Laetitia Dablanc. City Distribution, a Key Element of the Urban Economy: Guidelines for Practitioners. In *City Distribution and Urban Freight Transport*. Edward Elgar Publishing, 2011.

Leon Danon, Albert D. Iaz-Guilera, Jordi Duch, and Alex Arenas. Comparing community structure identification. Technical report, Universitat de Barcelona, Universitat Rovira i Virgili, 2005. URL https://arxiv.org/pdf/cond-mat/0505245.pdf.

Michael Drexl and Michael Schneider. A Survey of Variants and Extensions of the Location-Routing Problem. *European Journal of Operational Research*, 241(2):283–308, 2015. URL https://ideas.repec.org/a/eee/ejores/v241y2015i2p283-308.html.

Yingying Duan and Feng Lu. Structural Robustness of City Road Networks Based on Community. *Computers, Environment and Urban Systems*, 41:75–87, 2013.

H. Edelsbrunner, D. Kirkpatrick, and R. Seidel. On the Shape of a Set of Points in the Plane. *IEEE Transactions on Information Theory*, 29(4):551–559, 7 1983. ISSN 0018-9448. doi: 10.1109/TIT.1983.1056714. URL http://ieeexplore.ieee.org/document/1056714/.

Jan C. Fransoo, Edgar E. Blanco, and Christopher Mejía-Argueta. *Reaching 50 Million Nanostores: Retail Distribution in Emerging Megacities.* CreateSpace, 2018.

Emil Gegov, M. Nadia Postorino, Mark Atherton, and Fernand Gobet. Community Structure Detection in the Evolution of the United States Airport Network. *Advances in Complex Systems*, 16(01):1350003, 3 2013. ISSN 0219-5259. doi: 10.1142/S0219525913500033. URL http://www.worldscientific.com/doi/abs/10.1142/S0219525913500033.

Roel Gevaers, Eddy Van de Voorde, and Thierry Vanelslander. Characteristics and Typology of Last Mile Logistics from an Innovation Perspective in an Urban Area. *TPR: Department of transport and regional economics (University of Antwerp)*, pages 1–14, 2011.

David J. Giacomin and David M. Levinson. Road Network Circuity in Metropolitan Areas. *Environment and Planning B: Planning and Design*, 42(6):1040–1053, 11 2015. ISSN 0265-8135. doi: 10.1068/b130131p. URL http://journals.sagepub.com/doi/10.1068/b130131p.

Agust Gudmundsson and Nahid Mohajeri. Entropy and Order in Urban Street Networks. *Scientific reports*, 3:3324, 2013.

R. Guimerà, S. Mossa, A. Turtschi, and L. a. N. Amaral. The Worldwide Air Transportation Network: Anomalous Centrality, Community Structure, and Cities' Global Roles. *Proceedings of the National Academy of Sciences*, 102(22):7794–7799, 5 2005. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0407994102. URL http://www.pnas.org/content/102/22/7794.

Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring Network Structure, Dynamics, and Function using NetworkX. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11–15, Pasadena, CA USA, 2008.

M. Heemskerk Eelke, W. Takes Frank, Javier Garcia-Bernardo, and Huijzer Jouke M. Where is the Global Corporate Elite? A Large-scale Network Study of Local and Non-local Interlocking Directorates. *Sociologica*, 2/2016:0, 2016. ISSN 1971-8853. doi: 10.2383/85292. URL https://www.rivisteweb.it/doi/10.2383/85292https://www.rivisteweb.it/download/article/10.2383/85292.

Bill Hillier and Julienne Hanson. The Social Logic of Space. *Cambridge: Press syndicate of the University of Cambridge*, 1984.

J-L. Routhier. Du Transport de Marchandises en Ville à la Logistique Urbaine. *Centre de Prospective et de Veille Scientifique*, pages 1–66, 2002. URL https://halshs.archives-ouvertes.fr/halshs-00079005.

B. Jiang and C. Claramunt. A Structural Approach to the Model Generalization of an Urban Street Network. *GeoInformatica*, 8(2):157–171, 6 2004a. ISSN 1573-7624. doi:

10.1023/B:GEIN.0000017746.44824.70. URL https://doi.org/10.1023/B:GEIN.0000017746.44824.70.

Bin Jiang and Christophe Claramunt. Topological Analysis of Urban Street Networks. *Environment and Planning B: Planning and Design*, 31(1):151–162, 2004b. doi: 10.1068/b306. URL https://doi.org/10.1068/b306.

Andrea Lancichinetti and Santo Fortunato. Community Detection Algorithms: a Comparative Analysis. *Physical Review E*, 80(5), 11 2009. ISSN 1539-3755, 1550-2376. doi: 10.1103/PhysRevE.80.056117. URL http://arxiv.org/abs/0908.1062.

Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark Graphs for Testing Community Detection Algorithms. *Physical Review E*, 78(4):046110, 10 2008. ISSN 1539-3755. doi: 10.1103/PhysRevE.78.046110. URL https://link.aps.org/doi/10.1103/PhysRevE.78.046110.

Richard C. Larson and Amedeo R. Odoni. *Urban Operations Research*. Prentice-Hall, 1981. ISBN 0139394478. URL https://trid.trb.org/view.aspx?id=205030.

Sergio Lazzarini, Fabio Chaddad, and Michael Cook. Integrating Supply Chain and Network Analyses: The Study of Netchains. *Journal on Chain and Network Science*, 1(1):7–22, 6 2001. ISSN 1569-1829. doi: 10.3920/JCNS2001.x002. URL https://www.wageningenacademic.com/doi/10.3920/JCNS2001.x002.

David Levinson. Network Structure and City Size. *PLOS ONE*, 7(1):1–11, 2012. doi: 10.1371/journal.pone.0029721. URL https://doi.org/10.1371/journal.pone.0029721.

Robert F. Love and James G. Morris. Mathematical Models of Road Travel Distances. *Management Science*, 25(2):130–139, 2 1979. ISSN 0025-1909. doi: 10.1287/mnsc.25.2.130. URL http://pubsonline.informs.org/doi/abs/10.1287/mnsc.25.2.130.

Adam Lubinsky, Kushan Dave, and WXY architecture + urban Design. Integrated Strategies to Address Emerging Freight and Deliery Challenges in New York City : Final Report, 1 2017. URL https://rosap.ntl.bts.gov/view/dot/31861.

Daniel Merchán. *Effects of Road-Network Circuity on Strategic Decisions in Urban Logistics*. PhD thesis, Massachusetts Institute of Technology, 2018. URL https://dspace.mit.edu/handle/1721.1/119911.

Daniel Merchán and Matthias Winkenbach. An Emprical Validation and Data-Driven Extension of Continuum Approximation Approaches for Urban Route Distances (In Press). *Networks*, 2018.

David Meunier, Renaud Lambiotte, Alex Fornito, Karen Ersche, and Edward T Bullmore. Hierarchical Modularity in Human Brain Functional Networks. *Frontiers in Neuroinformatics*, 3:37, 10 2009. ISSN 16625196. doi: 10.3389/neuro.11.037.2009. URL http://journal.frontiersin.org/article/10.3389/neuro.11.037.2009/abstract.

Peter J. Mucha, Thomas Richardson, Kevin Macon, Mason A. Porter, and Jukka-Pekka Onnela. Community Structure in Time-Dependent, Multicale, and Multiplex Networks. *Science*, 328(5980):876–878, 5 2010. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1184819. URL http://science.sciencemag.org/content/328/5980/876.

Gábor Nagy and Saïd Salhi. Location-Routing: Issues, Models and Methods, 2007. URL /paper/Location-routing%3A-Issues%2C-models-and-methods-Nagy-Salhi/11a63eca3295fb0ad7f1dea2afd17906b5d02dee.

Günce Orman, Vincent Labatut, and Hocine Cherifi. Qualitative Comparison of Community Detection Algorithms. *arXiv:1207.3603 [physics]*, 167:265–279, 2011. doi: 10.1007/978-3-642-22027-2{_}23. URL http://arxiv.org/abs/1207.3603.

Ferran Parés, Dario Garcia-Gasulla, Armand Vilalta, Jonatan Moreno, Eduard Ayguadé, Jesús Labarta, Ulises Cortés, and Toyotaro Suzumura. Fluid Communities: A Competitive, Scalable, and Diverse Community Detection Algorithm. *arXiv:1703.09307 [physics]*, 3 2017. URL http://arxiv.org/abs/1703.09307.

Sergio Porta, Paolo Crucitti, and Vito Latora. The Network Analysis of Urban Streets: A Dual Approach. *arXiv:cond-mat/0411241*, 11 2004. URL http://arxiv.org/abs/cond-mat/0411241.

Sergio Porta, Paolo Crucitti, and Vito Latora. The Network Analysis of Urban Streets: A Primal Approach. *Environment and Planning B: Planning and Design*, 33(5):705–725, 2006. doi: 10.1068/b32045. URL https://doi.org/10.1068/b32045.

Caroline Prodhon and Christian Prins. A Survey of Recent Research on Location-Routing Problems. *European Journal of Operational Research*, 238(1):1–17, 2014. ISSN 0377-2217. URL https://econpapers.repec.org/article/eeeejores/v_3a238_3ay_3a2014_3ai_3a1_3ap_3a1-17.htm.

Carlo Ratti. Space Syntax: Some Inconsistencies. *Environment and Planning B: Planning and Design*, 31(4):487–499, 2004.

M. Rosvall, D. Axelsson, and C. T. Bergstrom. The Map Equation. *The European Physical Journal Special Topics*, 178(1):13–23, 11 2009. doi: 10.1140/epjst/e2010-01179-1. URL https://doi.org/10.1140/epjst/e2010-01179-1.

Iván Sánchez-Díaz, José Holguín-Veras, and Xiaokun Wang. An Exploratory Analysis of Spatial Effects on Freight Trip Attraction. *Transportation*, 43(1):177–196, 1 2016. ISSN 0049-4488. doi: 10.1007/s11116-014-9570-1. URL http://link.springer.com/10.1007/s11116-014-9570-1.

Karen R. Smilowitz and Carlos F. Daganzo. Continuum Approximation Techniques for the Design of Integrated Package Distribution Systems. *Networks*, 50(3):183–196, 10 2007. ISSN 1097-0037. doi: 10.1002/net.20189. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/net.20189.

Amanda Stathopoulos, Eva Valeri, and Edoardo Marcucci. Stakeholder Reactions to Urban Freight Policy Innovation. *Journal of Transport Geography*, 22:34–45, 5 2012. ISSN 0966-6923. doi: 10.1016/J.JTRANGEO.2011.11.017. URL https://www.sciencedirect.com/science/article/pii/S0966692311002195.

United Nations Population Division. World Population Prospects: the 2017 Revision. Technical report, United Nations, 2017. URL https://www.un.org/development/desa/publications/world-population-prospects-the-2017-revision.html.

Ludo Waltman and Nees Jan van Eck. A Smart Local Moving Algorithm for Large-Scale Modularity-Based Commmunity Detection. *The European Physical Journal B*, 86(11), 11 2013. ISSN 1434-6028, 1434-6036. doi: 10.1140/epjb/e2013-40829-0. URL http://arxiv.org/abs/1308.6604.

Eric W. Weisstein. Kronecker Delta. From MathWorld - A Wolfram Web Resource, 2019. URL http://mathworld.wolfram.com/KroneckerDelta.html.

Matthias Winkenbach, Paul R Kleindorfer, and Stefan Spinler. Enabling Urban Logistics Services at La Poste Through Multi-Echelon Routing. *Transportation Science*, 50(2): 520–540, 10 2015. ISSN 0041-1655. doi: 10.1287/trsc.2015.0624. URL https://pubsonline.informs.org/doi/abs/10.1287/trsc.2015.0624.

Matthias Winkenbach, Esteban Mascarino, and André Snoeck. Vision 2030: The Future of Urban Last-Mile Delivery. *CTL Working Paper*, page 277, 2017.

Zhao Yang, René Algesheimer, and Claudio Juan Tessone. A Comparative Analysis of Community Detection Algorithms on Artificial Networks. *Scientific Reports*, 6(30750), 2016. doi: 10.1038/srep30750.

Chen Zhong, Stefan Müller Arisona, Xianfeng Huang, Michael Batty, and Gerhard Schmitt. Detecting the Dynamics of Urban Structure Through Spatial Network Analysis. *International Journal of Geographical Information Science*, 28(11):2178–2199, 2014. ISSN 13623087. doi: 10.1080/13658816.2014.914521.