Intersections of Our World

Paolo Fogliaroni
Vienna University of Technology, Austria
paolo.fogliaroni@geo.tuwien.ac.at

Dominik Bucher
ETH Zurich, Switzerland
dobucher@ethz.ch

Nikola Jankovic
Vienna University of Technology, Austria
nikola.jankovic@geo.tuwien.ac.at

Ioannis Giannopoulos
Vienna University of Technology, Austria
igiannopoulos@geo.tuwien.ac.at

Abstract

There are several situations where the type of a street intersections can become very important, especially in the case of navigation studies. The types of intersections affect the route complexity and this has to be accounted for, e.g., already during the experimental design phase of a navigation study. In this work we introduce a formal definition for intersection types and present a framework that allows for extracting information about the intersections of our planet. We present a case study that demonstrates the importance and necessity of being able to extract this information.

2012 ACM Subject Classification Information systems → Geographic information systems, Information systems → Data analytics

Keywords and phrases intersection types, navigation, experimental design

Digital Object Identifier 10.4230/LIPIcs.GIScience.2018.3

1 Introduction

The street network of a city is a physical artifact embedded in the natural world. Most of the times, it consists of highways (i.e., streets meant for cars only), roads (meant for cars and pedestrians) and pathways (only for pedestrians). Sometimes these networks are following strict human design guidelines and sometimes they are bounded by natural constraints. Along with historical rationales, these constraints are the primary reasons that not all parts of a city follow a gridded design structure (e.g., curvilinear). This means that beside commonly encountered 3- and 4-way intersections, also more complex ones can exist.

But what are the main implications of this diversity of streets and intersections, and why is it important to know how a city, a country or even a continent are structured? What can we learn from this information and how can this information be useful?

In the following we will exemplify our work focusing on the area of navigation studies and experimental design. Independently of the research discipline, when planning an experiment there is a certain process that is followed in order to come up with a correct design. At the very beginning, information for the various relevant variables is collected that eventually will help to make the right choices.

In the case of navigation experiments, the relevant variables concern the subjects (e.g., gender or age), the type of navigation aid [13] and the timing of instructions [12], if any, and the environment (e.g., the route). When it comes to the environment, the relevant factors that have to be considered are numerous [15] and decisions can be made by taking into account possible interactions between the relevant subjects and the environment – e.g., previous experience of the subjects with the environment. Besides factors such as architectural differentiation and environmental landmarks [30], the types of intersections are highly relevant since they contribute to the complexity of a wayfinding decision [15]. A typical question during the design process is how the decision points along the designated route should be selected in terms of number of choices. How many and what kind of crossroads should the route encompass? Of course, the number of crossroads and their shape (e.g., T- or Y-intersections) on an experimental route is strongly related to the underlying research questions.

The aim of this work is to help answering this type of questions. We computed the number and type of all intersections on Earth and developed a web application that can be easily used to extract this precomputed information for any area in the world. Of course this work is not limited to navigation and experimental design. Next to researchers of various disciplines, industries related to the areas of transportation and urban planning can use our work for their decision making processes. For instance, by comparing the intersections of a street network between two areas, interesting correlations with other phenomena could be made, allowing to draw conclusions regarding the impact of the intersection types.

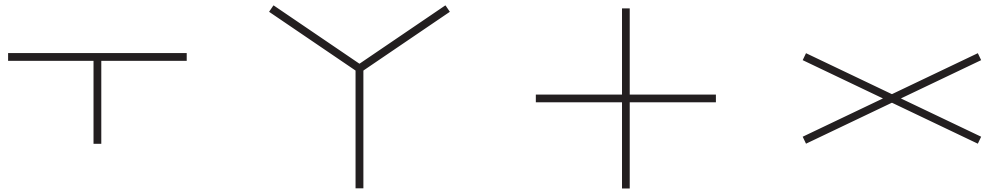
As a data source for our work we resorted to OpenStreetMap (OSM), that is one of the most commonly used source of volunteered geographical information (VGI). While approach we present does not require any particular form of road network data, the wide and free availability as well as the generally good quality of OSM [16] make it an adequate choice for intersection analysis. OSM data was analyzed in a multitude of studies before, not only in terms of quality and completeness [18], but also as a data source for answering questions about various environments [9], to determine the distribution of landmarks and points of interest [31, 28, 3], to build recommender systems [4] or as contextual enhancement for other types of data, such as Twitter posts [17].

In terms of intersection analysis, most previous work focuses around the automatic detection of roads and intersections from other sources of data, such as GPS traces [10, 7] or satellite imagery [6]. A variety of techniques exist, where intersection types are either implicitly learned using machine learning techniques (such as neural networks for satellite image analysis) [27, 32], or considered directly within the model [25]. To the best of our knowledge, in all of the automated detection methods the individual intersections are not classified in any way except based on the number of roads that lead up to them.

Intersections also play a central role in many routing applications [24]. Not only do red lights (commonly occurring at intersections) influence the driving time, behavior, and related emissions [23, 2], but even the difference between a right or left turn at an intersection incurs different penalties to route computations [20]. In addition, vehicular ad hoc networks (VANETs, which are used for inter-car communications) optimally also take intersections into consideration, as they provide data exchange points for cars driving on different routes and cars are likely to stop there [5, 1].

2 Types of Intersections

While the terms *junction* and *intersection* are commonly used interchangeably to refer street joints and crossings, they have slightly different meanings, with the term intersection referring to a specific type of junctions. According to the Oxford Dictionary, a *junction* is a place



(a) T-intersection. **(b)** Y-intersection. **(c)** Cross-intersection. **(d)** X-intersection.

Figure 1 The most common types of prototypical named intersections.

where two or more roads or railway lines meet, while an *intersection* is a point at which two or more things intersect, especially a road junction.

The term junction unambiguously relates to the mobility infrastructure domain and denotes roads coming together but does not specify the exact nature of their connection (intersect, touch, meet at a square, etc.). Conversely, the term intersection has a broader scope – as it can refer to several domains. Yet, when it comes to the mobility infrastructure domain it clearly refers to the cases where two or more roads intersect with each other.

Intersections are mostly studied in the areas of Architecture, Civil and Traffic Engineering, as well as Urban Planning. Studies in these domains are concerned with intersection design and construction to optimize traffic load, road safety, and traveling time (e.g., [29]). Intersections are typically split into two main categories: *at-grade* and *grade-separated* (see, e.g., [8]). At-grade intersections consist of roads located at the same level (grade), while the roads creating a grade-separated intersection are at different levels (grades) and pass above or below each other. Grade-separated intersections are mostly used in highways and motorways, as they allow for a faster and smoother merging of car traffic but are not well suited for pedestrian navigation.

Both categories can be more finely classified. Grade-separated intersections can be divided into *interchanges* and *grade-separations without ramps*. Subcategories of at-grade intersections include *proper intersections*, *roundabouts*, and *staggered (or offset) intersections*, among others. Proper intersections are the most prototypical type of intersection for the layman: several road segments converge to meet at the same point. Roundabouts are circular intersections that cars can enter and exit smoothly and in which road traffic flows in a single direction. In Staggered intersections several (minor) roads meet a main road at a slight distance apart such that they do not all come together at the same point.

In the scope of this work we only take into consideration proper intersections and, marginally, staggered intersections (that we regard as a composition of proper intersections). The analysis of more more complex types of intersections such as, e.g., roundabouts will be investigated in future work.

In the following we will introduce relevant terminology and discuss properties of proper intersections. The most straightforward property to classify intersections is the number n of street segments stemming out of it. We call such street segments the *branches* of the intersection. An intersection I with n branches is called an *n-way intersection* and we denote it by I^n. Obviously, we need at least two street segments to meet in order to form an intersection. In this work we focus on the intersections which call for navigational decision making: given one street segment that is used to approach an intersection, there have to be at least two more street segments that can be used to leave that intersection (i.e., $n \geq 3$).

A second discriminant that we use to classify an intersection is its shape. That is, the angular arrangement of its branches. Typically, this is done by comparing the intersections at hand to some others that are generally accepted as prototypical ones [33, 22, 26, 14]. The most common ones are reported in Figure 1: they are called T- and Y-intersections for $n = 3$; cross- and X-intersections for $n = 4$. Every intersection with more than four branches ($n > 4$) is typically referred to as a star-intersection.

There is evidence that these named intersection types are used very naturally by people when communicating route instructions verbally [33, 22] or schematically [33, 26, 14]. However, they suffer from two major drawbacks. First, these namings only exist for intersections with a small number of branches ($n \leq 4$). Second, they are often not precisely defined: for example, while most people would agree that a cross-intersection splits a revolution into four right angles, there might be a large disagreement on the skewness of an X-intersection.

For these reasons, we introduce the concept of *regular intersection*, whose branches divide a revolution into uniform parts. More formally:

▶ **Definition 1** (Regular n-way intersection). Let $b_0, \cdots, b_{n-1}$ be the branches of an n-way intersection enumerated in circular order . We define α_i as the angle formed by the pair $(b_{i-1}, b_{i \bmod n})$ for every $i \in \mathbb{N}$ such that $1 \leq i \leq n$. We say that a n-way intersection is *regular* if and only if $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 360/n$ and we denote it by R^n.

In general, to further characterize an n-way intersection we compare it to its regular counterpart, rather than to the aforementioned named intersection types. However, it has to be noted that regular 3- and 4-way intersections can be interpreted as exact definitions for Y- and cross-intersections, respectively. The arbitrary skewness of X-intersections makes them unsuitable to be taken as an objective reference for comparison. T-intersections, on the other hand, are well defined. For this reason, for 3-way intersections we also perform a comparison to T- intersections.

Finally, we define the angular distance $\Delta(I^n, R^n)$ among a generic n-way intersection I^n and its regular counterpart R^n as the minimum sum of angles that we have to rotate the branches of I^n to perfectly match R^n, while preserving the circular order of I^n's branches. Note that there are $n - 1$ possible rotations that can be performed to match I^n to R^n (see Sec 3.2 for more details).

3 Intersections Framework

In the following, we present our framework that was implemented for the classification and analysis of intersections. As one of the goals was to make worldwide intersection data available, the presented framework is based on OpenStreetMap data and is publicly available[1]. The framework is able to periodically process this data and writes the resulting intersection measures into a database, where they can be accessed through a web application.

3.1 Data Source

OpenStreetMap (OSM) is arguably one of the largest and most important volunteered geographic information (VGI) projects. As VGI is often not only the cheapest source of geographic information, but even the only one available in certain regions [16], it is an agreeable data source for a global intersection analysis. It needs to be noted that even though OSM data quality can be considered adequate for many purposes, its spatial distribution is not uniform, but depends on factors such as the information of interest or social events (e.g., an upcoming Football World Cup) in a region [18, 19, 11]. However, these quality

[1] See http://intersection.geo.tuwien.ac.at.

Analysis Class	Highway Tag Values	Description
Road	*living_street*, *primary*, *secondary*, *tertiary*, *unclassified*, *residential*, *service*, *primary_link*, *secondary_link*, *tertiary_link*	All ways that can be traversed by *both cars and pedestrians*, namely all normal roads.
Path	Road highway tag values **plus** *path*, *steps*, *bridleway*, *footway*, *track*, *pedestrian*	All ways that can be used by *pedestrians*. Including smaller tracks, hiking routes, etc. where cars cannot drive.
Car	Road highway tag values **plus** *motorway*, *motorway_link*, *trunk*, *trunk_link*	All ways that can be traversed by *car*. This additionally includes highways and motorways, where pedestrian access is usually forbidden.

Table 1 Different highway tag values used within the intersection analysis framework.

issues often concern single newly built roads or geographical information unrelated to the road network, which make up for a negligible amount of data with respect to a regional intersection analysis.

The three primary data structures of OSM are *nodes*, *ways* and *relations*. Nodes represent single points in space (i.e., they have a *longitude* and *latitude*), such as points of interest or individual objects. Ways are ordered lists of nodes, and encode linear features (like roads or rivers) and boundaries of areas (when the first and last node are equal). Finally, relations describe relationships between multiple elements, e.g., a collection of ways which form a scenic *route*, or turn restrictions, which state that you cannot cross from one way into another at a certain intersection.

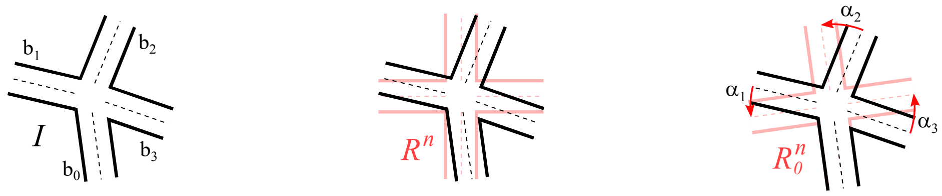
All the node, way and relation objects can have an arbitrary number of *tags*, which have a simple *key* → *value* form (both *key* and *value* are arbitrary strings). The tags themselves are not formally specified, but are chosen based on a consensus in the OSM community. For example, the very common tag *highway* is assigned to way objects which can somehow be used for travel, e.g., for walking or driving. It can take the values described in Table 1[2]. Note that we distinguish between three analysis classes, one with ways solely accessible to pedestrians, and another two with ways accessible to cars (including resp. excluding motorways). To find intersections in the OSM data, it suffices to look at ways that carry a *highway* tag, and to determine which nodes are shared among several of these ways.

OSM data is available in different formats. As the whole uncompressed xml planet file is around 850 GB at the time of this writing, we opted for the protocol buffer binary format (PBF) instead, which is available as a 40 GB gzipped file[3] and consists of around 4.3 billion nodes and 470 million ways.

3.2 Data Processing

After uncompressing the PBF file, we first search for nodes that should be considered intersections. As stated above, this corresponds to nodes which have more than two branches ($n \geq 3$). For each way in the OSM dataset that has one of the appropriate *highway* tag values

[2] For a detailed description of the individual values, and also additional ones that are not used in this framework, please consult the OSM documentation under wiki.openstreetmap.org/wiki/Key:highway.
[3] For details see wiki.openstreetmap.org/wiki/PBF_Format.



(a) Original 4-way intersection I with branches b_0-b_3.

(b) Overlay of perfect 4-way intersection R^4.

(c) Angles between original and perfect intersection.

Figure 2 Computation of $\Delta(I^n, R^n)$, the sum of all angles that each branch b_i has to be rotated in order to produce a regular n-way intersection. Note that it suffices to align the regular intersection with each branch (as is done for b_0 in **(c)**), and take the minimal Δ of all possible alignments.

(cf. Table 1), we iterate through all the nodes making up this way, and build a mapping that stores all neighboring nodes of each node. To be able to distinguish the different analysis classes later on, the highway tag value is additionally stored for each neighboring node. In essence, we define intersections as a function mapping a center node p to a number n of adjacent nodes $p_{p,i}$, where for each $p_{p,i}$ in addition the highway tag value $t_{h,i}$ of the connecting way is stored:

$$I^n : p \mapsto \{(p_{p,i}, t_{h,i}) \mid 0 \leq i < n\} \tag{1}$$

As this is done for all nodes in the OSM dataset (irrespectively of n), in a second iteration, a final set of intersections $\{I_0, ..., I_k\}$ has to be built by removing all nodes that dissatisfy the minimal number of branches condition (i.e., $|I(p)| < 3$). This set of intersections contains all the relevant OSM nodes for the purposes of the here presented framework. To compare each intersection to its regular counterpart (in the case of a 3-way intersection additionally to a perfect T-intersection), it is required to compute all angles between the different roads in a next step.

Thus, for the remaining intersections, a second pass through the OSM data collects the coordinates of the center node p itself, as well as the coordinates of all the neighboring nodes $p_{p,i}$ that can be reached by traversing its branches b_i. Using these coordinates, it is possible to compute all angles between the branches and the meridian passing through the center node. Figure 2 depicts a hypothetical 4-way intersection in black and, beneath it, the regular 4-way intersection, where the angles between branches are always 90°. In order to compute the angular distance $\Delta(I^n, R^n)$ to the regular intersection, we rotate the regular intersection n times, so that it always aligns perfectly with one of the branches b_i. Figure 2c shows one of the four possible alignments for a 4-way intersection. For each non-aligned branch, α_i denotes the required rotation to reach an alignment with the next “free” branch of the regular intersection (in this respect, “free” simply means that no two branches of the original intersection may be rotated to the same branch of the regular intersection). For any alignment with a branch of the regular intersection, a Δ' is computed as the sum of all α_i. The final Δ takes the value of the minimal Δ' over all n possible alignments. Note that this is a globally minimal Δ, even if arbitrary rotations of the regular intersection were allowed (and not just “snapping” to branches of the original intersection), as rotating the regular intersection monotonically increases or decreases Δ, until another alignment is reached. As such, all minima and maxima of Δ must occur at an alignment with the regular intersection.

All the intersections with their coordinates, the number of branches, as well as the computed $\Delta(I^n, R^n)$ are finally written to a PostGIS database[4]. Since it is required to know the analysis class of each intersection, an additional database field denotes if an intersection is valid for *road*, *path*, and *car*, or only any subset thereof.

3.3 Data Service

We provide public access to the intersection data computed with our framework through a web application that is accessible at `intersection.geo.tuwien.ac.at`.

The interface provides a map canvas with OSM as a basemap that can be used to freely browse the whole globe. With the current release of the application, the user is provided with a selection menu from where she can specify the type of intersections of interest (column "Analysis Class" in Table 1). We plan to extend this in future releases to allow the selection of combinations of the base intersection types.

We offer three possibilities to specify the region of interest: polygon drawing, viewport, and name search. In the first case, the user can specify a region by drawing a polygon on the map. With the canvas selection, the viewport currently shown on the map canvas is used to perform the database query. Finally, it is possible to look for named entities via a search box that provides a live interface to an OSM Nominatim[5] server. After typing in the name of the searched feature, the user can ask the interface to draw the corresponding polygon on the map. Given the huge amount of intersection data available, we decided to limit the area of the search region to not overload the server. In future releases this limitation might be removed. Also, in order to promote interoperability, we plan to include the possibility of specifying custom geometries expressed in different type formats (e.g., KML, geoJSON, etc.).

The intersection type and the region specified are used to submit a query that returns a statistical summary for intersections of the given type in the provided region. This summary contains the number of occurrences for each n-way intersection, the average Δ from the corresponding regular intersection – for 3-ways, also the average Δ from the regular T intersection. Besides the statistical summary the user is also provided with a link to download the whole intersection data set for the specified region and type as a CSV file.

At the time of writing the CSV file only contains information about the intersection points that were computed from OSM nodes. Beside the geometric information (reported in WKT) each point is associated the following attributes: the *number of branches* and the *type* of intersection, and the *angular distance* Δ to the corresponding regular intersection.

Note that the intersection classes defined in Table 1 are not disjoint. This results in the same intersection occurring up to three times in our database, once for each category. Imagine the case of an intersection where both roads and paths converge. For example, we may have 3 roads and 1 path. This intersection appears twice in our dataset: as a path and as a road. Since roads are accessible by both pedestrians and cars but paths are only accessible by cars, we have a 4-way path intersection and a 3-way road intersection. A similar concept applies to the categories of road and car intersections. The relation of the number of ways (denoted as n_{class}) between the intersections that overlap is $n_{car} \geq n_{road}$ and $n_{path} \geq n_{road}$.

In our database we also keep trace of the ways that form intersection branches: their geometry (also converted to OGC standard), the original OSM highway tag, and a relation to the intersections that they generate. This information is not accessible through the current version of the application, but will be made available in future releases.

[4] PostgreSQL 9.6 with PostGIS 2.3.2, the processing application is implemented in Rust 1.23.0.
[5] Nominatim is a search engine for OSM data, see wiki.openstreetmap.org/wiki/Nominatim.

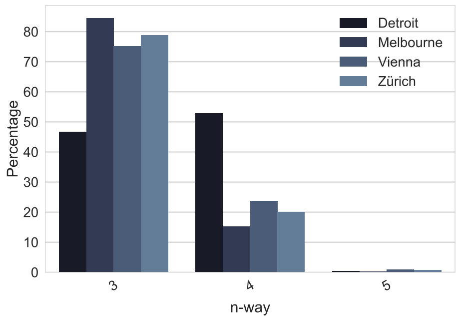


Figure 3 Distribution of the intersections as the number of branches n varies.

	Detroit	Melbourne	Vienna	Zürich
3-way	**46.76%**	**84.49%**	**75.16%**	**78.88%**
4-way	**52.84%**	**15.20%**	**23.74%**	**20.13%**
5-way	0.36%	0.29%	0.93%	0.82%
6-way	0.04%	0.02%	0.13%	0.14%
7-way	0.002%	0.002%	0.02%	0.02%
8-way	0.002%	-	0.01%	0.004%
10-way	-	-	-	0.004%
Total	40929	191508	75644	26286

Table 2 Distribution of intersections over number of ways for the four cities.

4 Use Case: Detroit, Melbourne, Vienna, and Zürich

In this section we present and discuss intersection data obtained with our framework for four exemplary cities and showcase how this data can be used during the design process of navigational experiments. In Section 4.1 we compare the four different cities, while in Section 4.2 we focus on local differences within a single city.

4.1 Comparative Study

We used our framework to extract intersection data for Detroit (USA), Melbourne (Australia), Vienna (Austria), and Zürich (Switzerland). While the framework allows for extracting intersection data concerning different types of streets (cf. Section 3.1), for this case study we focus on *paths* and *roads* (i.e., set of all walkable streets).

Table 2 reports the distribution (as percentages) of intersections as the number of branches n varies. From this data we can derive several interesting insights. First and foremost it has to be noted that for all the cities in exam almost the entirety of intersections are 3-ways and 4-ways. This becomes even more evident by looking at the graphical representation of the data reported in Figure 3. While this fact may seem trivial, it is still surprising the cumulative percentage that these two intersection categories reach together – ranging from 98.9% for Vienna to 99.7% for Melbourne. This pattern seems to recur everywhere in the world. Indeed, we found it in many other cities (Athens, Rome, Kathmandu, Washington DC, Paris, and London, among others) that we analyzed with our framework in a preliminary analysis for this work. This pattern consistently (only with minor differences) repeats across different cities, independently of their very heterogeneous morphology, history, and age.

The second insight that we can derive from this data relates to the ratio between the number of 3-way and 4-way intersections. In this respect, we notice that Melbourne, Vienna, and Zürich present a very similar trend with the majority of intersections being 3-ways, although with slightly different ratios between the number of 3- and 4-ways: approx. 5.5 for Melbourne, 3.2 for Vienna, and 3.9 for Zürich. Conversely, Detroit shows the opposite trend, with the number of 4-ways slightly bigger than that of 3-ways. This may indicate, for example, a more blocked structure of the city.

In the following we analyze the further discriminant introduced in this work to classify intersections: the similarity to *regular intersections* (see Definition 1). As discussed in Sections 2 and 3.2, we measure this by the angular distance $\Delta(I^n, R^n)$ between a generic n-way intersection I^n and the corresponding regular intersection R^n. For the case of 3-ways,

City	Min	P_{25}	P_{50}	P_{75}	Max
Det	∼0%	0.58%	1.96%	15.98%	99.99%
Mel	∼0%	1.03%	4.31%	21.59%	99.65%
Vie	∼0%	1.51%	6.08%	22.16%	99.87%
Zur	∼0%	2.09%	7.54%	23.48%	99.76%
Δ-range: $[0°, 180°]$					

(a) 3-way to regular T, delta percentiles.

City	Min	P_{25}	P_{50}	P_{75}	Max
Det	∼0%	0.23%	0.59%	2.11%	50.00%
Mel	∼0%	0.63%	2.37%	8.05%	83.69%
Vie	∼0%	0.91%	3.17%	9.69%	85.81%
Zur	∼0%	1.44%	4.45%	11.41%	64.12%
Δ-range: $[0°, 360°]$					

(b) 4-way to regular 4-way, delta percentiles.

Table 3 Distribution of 3-way (4a) and 4-way (4b) intersections for the four cities (normalized).

we compare against regular T intersection instead. Moreover, given that for the cities in exam 3-ways and 4-ways combined cover almost the totality of the number of intersections, we will only focus on those.

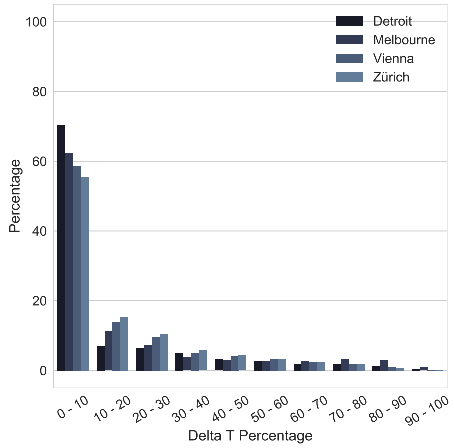
Tables 4a and 4b report descriptive statistics for 3-ways and 4-ways, respectively. The numbers reported are percentages referring to the value range that the angular distances can take on. This is called Δ-range and denotes the difference between the minimum (Δ_{min}) and maximum (Δ_{max}) angular distances from a generic intersection to its regular counterpart. Obviously, the minimum is always zero ($\Delta_{min} = 0°$), which corresponds to a perfect match with the regular intersection. Conversely, Δ_{max} depends on the number of branches (n) of the intersection at hand and corresponds to the angular distance of the (theoretical) worst-case scenario where all the branches of an intersection collapse on top of each other:

$$\Delta_{max} = \sum_{i=1}^{\lfloor \frac{n-1}{2} \rfloor} (2i\alpha) + ((n-1) \bmod 2)\,\pi \tag{2}$$

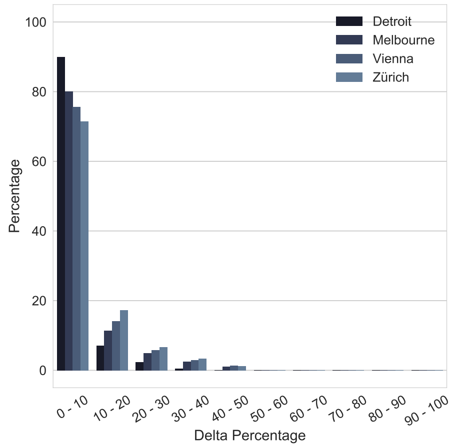
For an understanding of this formula imagine to align any branch of the regular intersection to the first branch of the n-way at hand. Subsequently, take a pair of unmatched branches from the generic intersection and rotate them (one clockwise and the other counterclockwise) by $\alpha = \frac{360}{n}$ to match the first pair of unmatched branches of the regular intersection. Now repeat for the second pair of unmatched branches. In this case, we will have to rotate 2α in order to find the first pair of unmatched branches of the regular intersection. Generalizing this operation we obtain the formula in Equation 2. For 3-ways and 4-ways we have Δ-ranges equal to $[0°, 240°]$ and $[0°, 360°]$, respectively. The Δ-range for 3-ways when compared against the regular T intersection is equal to $[0°, 180°]$.

Figures 4a and 4b plot in greater details the distribution of 3-ways and 4-ways as the angular distance varies over the Δ-ranges for the regular T intersection and the regular 4-way, respectively. The figures show that the majority of the intersections are very similar to their regular counterparts (which aligns nicely with Klippel's set of wayfinding choremes [22, 21]), with Detroit and Zürich representing extreme cases. The intersections of Detroit are the most regular, with approximately 70% of its 3-ways and 90% of its 4-ways showing an angular distance below 10% to the regular T intersection (i.e., 18°) and the regular 4-way (i.e., 36°), respectively. Conversely, Zürich is the least regular, with approximately 55% of its 3-ways and 70% of its 4-ways showing an angular distance below 10% to the regular T intersection (i.e., 18°) and the regular 4-way (i.e., 36°), respectively. Melbourne and Vienna are located in between these extremes, with Melbourne being slightly more regular than Vienna with respect to both 3-ways and 4-ways.

These findings can be used, for example, during the design of navigational experiments to select paths that adhere to the structure of the city where the experiments are to be



(a) 3-way to regular T intersection.



(b) 4-way to regular 4-way intersection.

Figure 4 Dsistribution of the angular distance (Δ) for 3-ways **(a)** and 4-ways **(b)** with respect to the regular T intersection and the regular 4-way intersection, respectively. The angular distance (on the x-axis) is reported as a percentage of the different Δ-ranges for 3-ways (i.e., $0° - 180°$) and 4-ways (i.e., $0° - 360°$). The percentage on the y-axis refers to the number of intersections in each bin with respect to the total number of intersections of that type (i.e., 3-way and 4-way). The smaller the value of Δ, the higher the similarity to the corresponding regular intersection.

performed. In this way, we can avoid to select some *atypical* path that may lead to biased results. Assume that for our hypothetical navigational experiment we need a path that comprises 10 intersections. If we were to conduct the experiment with a path matching the characteristics of Detroit, we should select a path in the real world or in a virtual environment that encompasses, e.g., five 3-way and five 4-way intersections. Of the selected 3-ways (resp. 4-ways), three (resp. five) should present a maximum angular distance of 18° (resp. 36°) from the regular T intersection (resp. the regular 4-way). Conversely, if we were to conduct the same experiment with a path matching the characteristics of Zürich, our path should encompass eight 3-way and two 4-way intersections. Of the selected 3-ways (resp. 4-ways), four (resp. six) should present a maximum angular distance of 18° (resp. 36°) from the regular T intersection (resp. the regular 4-way).

Moreover, the availability of intersection data for the entire world easily supports comparative analysis that so far was difficult to control. Imagine to run the same spatial experiment in different cities or areas of the globe. The availability of this data may allow for comparing the different paths and, consequently, for relating and gaining insights on the possibly different experimental results obtained in different locations.

4.2 Local Differences

In this section we discuss local differences within the city of Vienna. We used our framework to run analysis on all 23 districts (DIST) and focus on the two with the highest variation, district 8 and 10.

Table 5 reports the distribution (as percentage) of the intersections as the number of branches n varies. This allows for easily comparing the statistics of the selected districts against the statistics extracted for whole Vienna. Both the selected districts adhere to

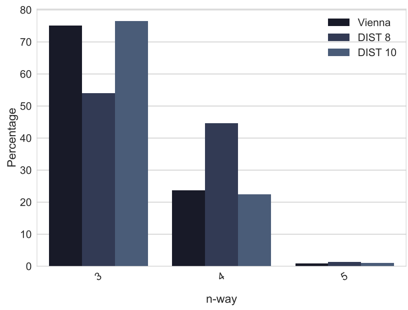


Figure 5 Distribution of the intersections as the number of branches n varies.

	Vienna	DIST 8	DIST 10
3-way	**75.16%**	**53.97%**	**76.46%**
4-way	**23.74%**	**44.63%**	**22.44%**
5-way	0.93%	1.4%	1%
6-way	0.13%	-	0.07%
7-way	0.02%	-	0.03%
8-way	0.01%	-	-
Total	75644	428	7121

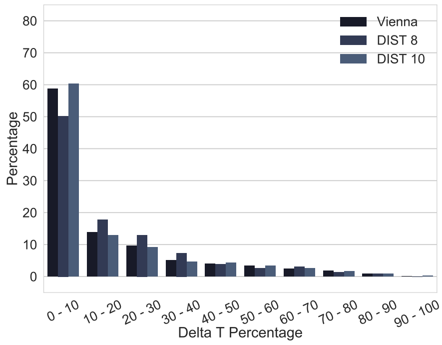
Table 5 Distribution of intersections over number of ways for whole Vienna and the 2 districts in exam.

the overall distribution pattern that we discussed in Section 4.1, with almost the entirety of intersections distributed between 3-ways and 4-ways. The graphical representation of the data (see Figure 5) allows for glimpsing different local patterns for the two districts. Specifically, district 10 exhibits a distribution almost identical to whole Vienna. In contrast, district 8 exposes different distributions, with approximately 20% less 3-ways (resp. 20% more 4-ways) than whole Vienna.

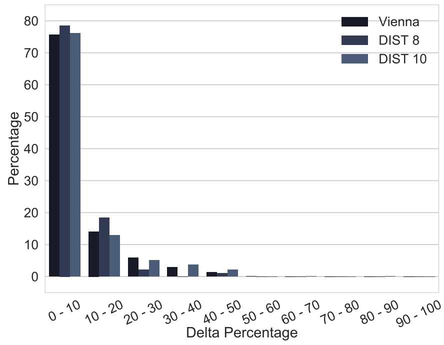
The distribution of 3-way and 4-way intersections can be seen in Figures 6a and 6b as their normalized angular distance varies in the corresponding Δ-ranges – i.e., $[0°, 180°]$ and $[0°, 360°]$, respectively. As for 3-ways, district 8 is the most dissimilar with respect to Vienna, while district 10 exhibits only a small deviation from the distribution of the whole city. The same pattern emerges also for 4-ways.

Assume that we want to replicate in Vienna the navigational experiment discussed at the end of Section 4.1 for which we need to select a path encompassing 10 intersections. If we were to conduct the experiment in district 10, according to the intersection distribution reported in Figure 5, approximately 76% (resp. 22%) of these intersections should be 3-ways (resp. 4-ways). Say, for example, that we choose a path consisting of eight 3-ways and two 4-ways. According to the distribution of Δs in Figures 6a and 6b, of the selected 3-ways (resp. 4-ways), five (resp. 2) should present a maximum angular distance of 18° (resp. 36°) from the regular T intersection (resp. the regular 4-way).

If the experiment was to be conducted in district 8 we could either decide to stick to the statistics of whole Vienna or to the statistics of the district. In the first case we would end up with a selection similar to that of district 10. In the second case we would have to choose differently. If we opt for the first alternative the findings that relate to the structure of intersections could be considered as a step towards generalization to whole Vienna but might apply more loosely to district 8. More generally, the statistical data provided by our framework can be used to find out areas all over the world that expose an intersection structure similar to that of a given area where, e.g., we performed an experiment. This information can be used to replicate the experiment in any of these areas and identify which of the insights we derive from the experiment results are invariant with respect to the intersection structure of the path.



(a) 3-way to regular T intersection.



(b) 4-way to regular 4-way intersection.

Figure 6 Bar plot visualization of the distribution of the angular distance (Δ) for 3-ways **(a)** and 4-ways **(b)** with respect to the regular T intersection and the regular 4-way intersection, respectively. See Figure 4 for reading instructions.

5 Discussion and Conclusion

The framework presented in this work can be considered as an important asset during the design of spatial experiments and to perform spatial analysis. As shown through the case study in Section 4, the framework can be easily used to partially validate a selected route with respect to generalization issues. Since local differences can be found in an urban environment that do not adhere to the overall structure of a city, a country, or even a continent, the choice of a route has to be considered very carefully. Furthermore, by identifying similarities of the selected route at different scales (i.e., from district up to continent scale), one can go a step further and carefully interpret the findings of the experiment (at least those related to features of the intersection distributions) and draw conclusions concerning the reproducibility and comparison with experiments performed in different areas. Of course looking only at the intersections of a route is not sufficient, but necessary. This work can be considered as a further step towards interpreting the results of an experiment concerning generalizability aspects.

Next to the scenario used throughout this paper to exemplify how the results of this work can be utilized, this type of quantitative data can also be useful for a multitude of other purposes. For instance, machine learning approaches could profit from this framework, generating relevant features that can help to describe the spatial phenomena of interest. Another example would include work in the area of transportation, trying to model the access and demand or relevant work in the area of urban planing. Furthermore this framework could also easily be used as part of city modeling softwares, e.g., Esri CityEngine[6], helping to automatically generate look-alike urban environments.

In this paper we presented the raw intersection data that we generated from OSM data through the procedure described in Section 3.2 and show an example of how this data can be used for the design and comparison of navigational experiments. However, according to the specific experiment at hand it might be necessary to clean the raw data in order to accommodate geometrical and perceptual aspects. We identified two cases where the raw data may need to be cleaned before usage. Both cases concern scenarios where two or more

[6] See http://www.esri.com/software/cityengine.

intersections are located very close to each other. If the intersections under consideration are of the same type, this may denote a mapping issue: due to accuracy problems a single intersection in the real world is actually reported as several in OSM. Alternatively, the intersections might actually be correctly reported in OSM, but we may have a perception issue: although we physically have several intersections, they are so close to each other that a person could perceive them as a single intersection.

The other scenario concerns the case where intersections of different types are very close to each other. Specifically, we identified a somewhat problematic pattern where a road intersection is surrounded by a set of path intersections representing sidewalks and zebra stripes. In such situations, we actually have a single intersection in the real world that is identified as several by our framework. This issue is due to the fact that in OSM, sidewalks can either be mapped as separate ways or denoted with an apposite tag on the corresponding road. This means that we cannot know in advance how many times this scenario appears in our data. For this reason we performed a simple buffer and cluster analysis on Vienna to find out the amount of groups of intersections in our data that should actually be considered as a single intersection. We used buffer of different sizes (ranging from 1m to 10m) to identify clusters corresponding to both scenarios: intersections of same type and one road surrounded by path intersections. For the first scenario we found that the number of clusters ranges from 0.04% to 4.8% (resp. from 0.2% to 12.4%) of the road (reps. path) intersections, as we increase the buffer radius from 1m to 10m. For the road-to-path scenario, the number of clusters ranges 0 to 5.7% of the road and path intersections.

Finally, it has to be noted that the implementation of our framework does not compute the data on the fly from OSM data. Rather, a snapshot of the OSM database is taken and intersection data is generated from there. This means that the data provided on the website might not be completely actual, although we do not expect huge discrepancies.

6 Outlook

Since in our work we focused on regular intersections, we omitted analyses of roundabouts. In the underlying OSM data, roundabouts are modeled as multiple 3-way intersections. Although this might look correct at a first glance, one can argue that roundabouts form a category of its own, or even an n-way intersection, with n equals the number of ingoing and outgoing branches. As this is an open question that needs further investigation and probably a user study to understand how humans perceive roundabouts, we will focus on this problem in the future. Since this framework is not only indented to be used for experimental design, a possible solution could be to transfer the choice to the users of this framework, by providing multiple options on how to handle roundabouts during runtime.

Also, in this work we did not perform any scale-based aggregation of the street geometries (e.g., aggregating two lanes of a street into a single line). Therefore, the results presented in this paper are at the finest level of details allowed by data source. Street aggregation will also yield a reduction in the number of detected intersections as well as a simplification of the resulting intersection network. Future work along this direction may potentially lead to a hierarchical organization of the data that, in turn, may allow for further types of uses and analyses of the intersection data.

In future work we will also focus deeper on network patterns. For instance, what is the most common sequence of intersections for a given length (number of intersections)? What is the typical distance between intersections or intersection types (segment length)? Being able to extract this type of information will further improve the goals set in this paper, allowing to draw even better conclusions and automatically create even more realistic look-alike cities.

References

1 Nizar Alsharif, Sandra Céspedes, and Xuemin Shen. icar: Intersection-based connectivity aware routing in vehicular ad hoc networks. In *Communications (ICC), 2013 IEEE International Conference on*, pages 1736–1741. IEEE, 2013.

2 Behrang Asadi and Ardalan Vahidi. Predictive cruise control: Utilizing upcoming traffic signal information for improving fuel economy and reducing trip time. *IEEE transactions on control systems technology*, 19(3):707–714, 2011.

3 Mohamed Bakillah, Steve Liang, Amin Mobasheri, Jamal Jokar Arsanjani, and Alexander Zipf. Fine-resolution population mapping using openstreetmap points-of-interest. *International Journal of Geographical Information Science*, 28(9):1940–1963, 2014.

4 Andrea Ballatore, Gavin McArdle, Caitriona Kelly, and Michela Bertolotto. Recomap: an interactive and adaptive map-based recommender. In *Proceedings of the 2010 ACM Symposium on Applied Computing*, pages 887–891. ACM, 2010.

5 Jin-Jia Chang, Yi-Hua Li, Wanjiun Liao, and Chau Chang. Intersection-based routing for urban vehicular communications with traffic-light considerations. *IEEE Wireless Communications*, 19(1), 2012.

6 Dragos Costea and Marius Leordeanu. Aerial image geolocalization from recognition and matching of roads and intersections. *arXiv preprint arXiv:1605.08323*, 2016.

7 Ole Henry Dørum. Deriving double-digitized road network geometry from probe data. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL'17, pages 15:1–15:10. ACM, 2017.

8 Said Easa. *Civil Engineering Handbook*, chapter Chapter 63: Geometric design. CRC Press, Boca Raton, FL, 2002.

9 Jacinto Estima and Marco Painho. Exploratory analysis of openstreetmap for land use classification. In *Proceedings of the second ACM SIGSPATIAL international workshop on crowdsourced and volunteered geographic information*, pages 39–46. ACM, 2013.

10 Alireza Fathi and John Krumm. Detecting road intersections from gps traces. In *International Conference on Geographic Information Science*, pages 56–69. Springer, 2010.

11 Mohammad Forghani and Mahmoud Reza Delavar. A quality study of the openstreetmap dataset for tehran. *ISPRS International Journal of Geo-Information*, 3(2):750–763, 2014.

12 Ioannis Giannopoulos, David Jonietz, Martin Raubal, Georgios Sarlas, and Lisa Stähli. Timing of pedestrian navigation instructions. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 86. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

13 Ioannis Giannopoulos, Peter Kiefer, and Martin Raubal. Mobile outdoor gaze-based geohci. In *Geographic Human-Computer Interaction, Workshop at CHI 2013*, pages 12–13. Citeseer, 2013.

14 Ioannis Giannopoulos, Peter Kiefer, and Martin Raubal. Gazenav: Gaze-based pedestrian navigation. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 337–346. ACM, 2015.

15 Ioannis Giannopoulos, Peter Kiefer, Martin Raubal, Kai-Florian Richter, and Tyler Thrash. Wayfinding decision situations: A conceptual model and evaluation. In *International Conference on Geographic Information Science*, pages 221–234. Springer, 2014.

16 Michael F Goodchild. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4):211–221, 2007.

17 Stefan Hahmann, Ross S Purves, and Dirk Burghardt. Twitter location (sometimes) matters: Exploring the relationship between georeferenced tweet content and nearby feature classes. *Journal of Spatial Information Science*, 2014(9):1–36, 2014.

18 Mordechai Haklay. How good is volunteered geographical information? a comparative study of openstreetmap and ordnance survey datasets. *Environment and planning B: Planning and design*, 37(4):682–703, 2010.

19 Mordechai Haklay, Sofia Basiouka, Vyron Antoniou, and Aamer Ather. How many volunteers does it take to map an area well? the validity of linus' law to volunteered geographic information. *The Cartographic Journal*, 47(4):315–322, 2010.

20 Ronald F Kirby and Renfrey B Potts. The minimum route problem for networks with turn penalties and prohibitions. *Transportation Research*, 3(3):397–408, 1969.

21 Alexander Klippel. Wayfinding choremes. In *International Conference on Spatial Information Theory*, pages 301–315. Springer, 2003.

22 Alexander Klippel, Heike Tappe, Lars Kulik, and Paul U Lee. Wayfinding choremes—a language for modeling conceptual route knowledge. *Journal of Visual Languages & Computing*, 16(4):311–329, 2005.

23 Chunxiao Li and Shigeru Shimamoto. A real time traffic light control scheme for reducing vehicles CO 2 emissions. In *Consumer Communications and Networking Conference (CCNC), 2011 IEEE*, pages 855–859. IEEE, 2011.

24 Dennis Luxen and Christian Vetter. Real-time routing with openstreetmap data. In *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems*, pages 513–516. ACM, 2011.

25 Helmut Mayer, Stefan Hinz, Uwe Bacher, and Emmanuel Baltsavias. A test of automatic road extraction approaches. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, 36(3):209–214, 2006.

26 Tobias Meilinger, Christoph Hölscher, Simon J Büchner, and Martin Brösamle. How much information do you need? Schematic maps in wayfinding and self localisation. In *International Conference on Spatial Cognition*, pages 381–400. Springer, 2006.

27 Volodymyr Mnih and Geoffrey E Hinton. Learning to detect roads in high-resolution aerial images. In *European Conference on Computer Vision*, pages 210–223. Springer, 2010.

28 Eva Nuhn, Wolfgang Reinhardt, and Benjamin Haske. Generation of landmarks from 3d city models and osm data. In *Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, France*, pages 24–27, 2012.

29 American Association of State Highway and Transportation Officials. A policy on geometric design of highways and streets, 2011.

30 Martin Raubal and Stephan Winter. Enriching wayfinding instructions with local landmarks. In *International conference on geographic information science*, pages 243–259. Springer, 2002.

31 Kai-Florian Richter and Stephan Winter. Harvesting user-generated content for semantic spatial information: The case of landmarks in openstreetmap. In *Proceedings of the Surveying and Spatial Sciences Biennial Conference*, pages 75–86, 2011.

32 Shunta Saito and Yoshimitsu Aoki. Building and road detection from large aerial imagery. In *Image Processing: Machine Vision Applications VIII*, volume 9405, page 94050K. International Society for Optics and Photonics, 2015.

33 Barbara Tversky and Paul U Lee. Pictorial and verbal tools for conveying routes. *Spatial information theory. Cognitive and computational foundations of geographic information science*, pages 51–64, 1999.