A computational approach for categorizing street segments in urban street networks based on topological properties

Hsiao-Hui Chen, Olaf Mumm* and Vanessa Miriam Carlow

ISU - Institute for Sustainable Urbanism, Technische Universität Braunschweig, Braunschweig, Germany

Street classification is fundamental to transportation planning and design. Urban transportation planning is mostly based on function-based classification schemes (FCS), which classifies streets according to their respective requirements in the pre-defined hierarchy of the urban street network (USN). This study proposes a computational approach for a network-based categorization of street segments (NSC). The main objectives are, first, to identify and describe NSC categories, second, to examine the spatial distribution of street segments from FCS and NSC within a city, and third, to compare FCS and NSC to identify similarities and differences between the two. Centrality measures derived from network science are computed for each street segment and then clustered based on their topological importance. The adaption of clustering, which is a numerical categorization technique, potentially facilitates the integration with other analytical processes in planning and design. The quantitative description of street characteristics obtained by this method is suitable for development of new knowledge-based planning approaches. When extensive data or knowledge of the real performance of streets are not available or costly, this method provides an objective categorization from those data sets that are readily available. The method can also assign the segments that are categorized as “unclassified” in FCS to the categories in the NSC scheme. Since centrality metrics are associated with the functioning of USNs, the comparison between FCS and NSC not only contributes to the understanding and description of the fine variations in topological properties of the segments within each FCS class but also supports the identification of the mismatched segments, where reassessment and adjustment is required, for example, in terms of planning and design.

KEYWORDS
street segment categorisation, network analysis, multiple centrality assessment, cluster analysis, sustainable urban mobility

1 Introduction

For cities, well-functioning urban street networks (USN) are essential in connecting places and people (Noori et al., 2020). In transport planning, the categorization of streets provides a sound basis for guidelines for managing and maintaining the street system (Dong et al., 2013). A widely used approach to classify streets is the functional-based classification scheme (FCS), mainly focusing on the mobility and accessibility of streets (FHWA, 2013). FCS characterizes the role of each street in the overall urban transportation network (FHWA, 2013). Since the
flows of motorized traffic that the network elements (i.e., street segments) carry have an impact on the quality and liveliness of the street system (Yerra and Levinson, 2005), a critical criterion for classifying streets is the planned and expected annual average daily traffic volume (FHWA, 2013). Accordingly, arterial roads are typically roadways with high motorized traffic volumes, which by definition “serve a large percentage of travel between cities and other activity centres” (FHWA, 2013). By contrast, local roads typically have relatively low traffic volumes (FHWA, 2013).

However, the common classifications based on functional attributes, as in FCS, largely neglect street segments’ topological properties within a network. Without considering topological attributes, the capacity of FCS to model movement patterns or traffic flows may be undermined (Paul, 2015). For example, a local road, which is expected to have low levels of traffic volumes according to the definition of FHWA, may actually play an important role in the network if we examine its topological characteristics from the perspective of configuration analysis.

This study introduces a computational approach for network-based categorization of street segments (NSC) that provides the quantitative description of the topological characteristics of streets segments. Centrality measures derived from network science are computed for each street segment, which are then clustered based on their topological importance. The adaptation of clustering, which is a numerical categorization technique, supports the creation of an objective categorization scheme of street segments. In a concluding step this study compares the centrality measures based NSC with the FCS description of the USN of the study area (i) to identify the deviation between the two schemes and (ii) to more accurately distinguish different characteristic clusters within each FCS class. This allows a pertinent evaluation with regard to questions of planning and design, to further distinguish the heterogeneity within each FCS class and increase the information, such as accessibility and connectivity, in each classification.

Graph theory has been commonly applied to understand USN. The centrality measures derived from network analysis have been applied to extract spatial-structural properties and the topological importance of nodes and edges in the network (Hillier, 1996; Porta et al., 2006; Zhong, Arisona, Huang, Batty and Schmitt, 2014; Berli, Ducruet, Martin and Seten, 2020).

Previous studies, from both the fields of space syntax and network analysis show that investigating the structure of USNs by “collective network properties”, which simultaneously take the relations between all spaces in the network into consideration (Hillier, 1996; Hillier, 2012) rather than one or few (Serra et al., 2016), ultimately contributes to gaining an understanding of urban function, such as a) traffic flows (Jiang, 2009; Kazerani and Winter 2009) and the movement rates in different thoroughfares (Penn et al., 1998; Hillier and Iida, 2005; Serra et al., 2015), b) movement patterns (Jiang and Liu, 2009), c) the spatial distributions of residential or retail areas (Schwander et al., 2013; Ravulaparthi and Goulais, 2014) and d) the human way-finding capacity (Crucitti et al., 2006; Svetuk et al., 2016; Tiarasari and Kartidjo, 2021) that are network-constrained in urban areas.

To this end, the present study builds on previous studies, but expands upon them by adding the following crucial aspects.

Building on the methodologies from the field of space syntax (Hillier and Iida, 2005; Turner, 2007), previous studies (Barthelemy, 2017; Berghauser-Pont et al., 2019) aim to develop a street categorization based on only a single network centrality measure, betweenness centrality, for the quantitative description of streets. In the current research we emphasize, that for complex spatial structures like urban street networks multiple centrality measures are necessary for their analysis and description.

Works applying the Multiple Centrality Assessment (Porta et al., 2006; Porta et al., 2013) evaluate different features of street segments utilizing a multi-variable approach to investigate street characteristics separately by individual measures (Zhang and Li, 2011) (see Supplementary Figure S1 as example), instead of applying the clustering technique to multiple centralities and simultaneously taking the relations between all street segments in the network into consideration. Although classification based on separate elements allow comparison between different spatial features, the interrelation between elements is lacking (Berghauser-Pont et al., 2017). To fill this gap, the current study adopts a more integrated approach: the street segments are categorized by a cluster analysis, which combines the different centralities simultaneously.

Other studies apply the clustering technique to integrated multi-variables (though not multiple centrality measures) to define street types (cf. Berghauser-Pont and Haupt, 2010; Gil et al., 2012; Serra, 2013; Barthelemy, 2017; Berghauser-Pont et al., 2017; Berghauser-Pont et al., 2019; Boeing, 2020). This allows for a comparative analysis of different cities to “give an objective summary of their spatial structure, identifying common as well as unique traits” (Berghauser-Pont et al., 2017).

Yet other studies aimed at expanding FCS for improved Multimodal Designs (Stamaitidis et al., 2019) to identify the location for introducing “multimodal corridors” (Tsigdinos et al., 2020; Tsigdinos et al., 2021). However, the topological characteristics of street segments in USN were overlooked. Noori et al. (2020) aim to fill this gap by using deep learning techniques to develop a classifier model and predict the functional class of streets. The current research does not classify multiple city networks or focus on examining the predicting ability of centrality measures. Instead, our analysis applies centrality measures to create a network-based categorization scheme (NSC).

The objectives of this study are: (a) to develop an automatized, and reproducible method for applying network analysis and machine learning for the categorisation of street segments by computing and clustering multiple centrality dimensions, which evaluate their accessibility, connectivity, and intermediary capacity, as well as the importance of the neighbouring segments; and (b) to compare and discuss the method revealing the heterogeneous topological characteristics within each homogeneous FCS classification. The overall objective of the development of the method is motivated by its use in deriving possible approaches for sustainable development of urban street networks, e.g., by identifying possible traffic hotspots—both conflict-prone or those with high potential for, e.g., a new bike infrastructure.

The paper is structured as follows: The "Materials and Methods" section provides an overview of the study area (the city of Braunschweig) and its USN, including the spatial distribution of...
2 Materials and methods

2.1 Case study: study area and data source

The NSC method was developed and tested for the city of Braunschweig in Lower Saxony, a Federal State of Germany. The city has a population of 0.25 million and an administrative size of 192 km². The city of Braunschweig was selected as case study due to the availability of relevant data and access to additional resources such as official transport models. The spatial research boundary is delineated by the administrative boundary of the city. The prime data source for extraction of the USN is OpenStreetMap (OSM) (OpenStreetMap contributors, 2021). OSM describes roads by function and importance (key:highway) (cf. OpenStreetMap contributors, 2023). Although it does not depict official classifications of planning authorities due to lack of comprehensive official data this study utilizes the definition of the functional classification indicated by OSM. In Braunschweig, the values of the key:highway are motorway, primary, secondary, tertiary, residential, and unclassified roads. The total number of segments and overall length of each classification in Braunschweig are provided in Supplementary Table S1.

In Germany, roads are officially grouped into motorways, country roads, main roads, collector roads, and access roads. Furthermore, they are described by the level of their linking function, which depends on whether the road connects big cities, smaller towns or villages (there are 6 levels: from 0 to V). This official classification is provided in the Guidelines for Integrated Network Design (Richtlinien für integrierte Netzgestaltung, RIN) adopted in 2008 by the Research Society for Road and Traffic Engineering (FGSV, 2008). Combining the level and the category of a road allows for determining the road’s design class (Entwurfsklasse)—e.g., level III (regional) country road. So far, Germany does not have an official (publicly available) data set that assigns the design classes to the existing road network (Holthaus and Thiemermann, 2022). For this reason, the classification in OSM forms the basis for the present study.

Figure 1 presents the USN within the administrative borders of Braunschweig differentiated by the OSM highway keys. It shows that the northern and southern parts of the city are mainly connected by streets with the function of a motorway (Figure 1B), whereas its eastern and western parts are mainly connected by primary roads (Figure 1C). Street segments with the key:highway "residential" facilitate smaller-scale communication between the neighbourhoods and often form segment clusters, especially at the outer edge of the study area (See Figure 1F).

2.2 Indicators

We define street segment categories (NSC categories) as segments with similar topological characteristics regarding the intermediary capacity, accessibility, connectivity, and the importance of its neighbouring segments. USN can be modelled with two methods. In the “primal graph” streets are considered as edges, whereas in the “dual graph” streets are considered as nodes. In this research, the smallest research unit is the street segment, defined as a line segment included in the network analysis, which is the “edge” of the “primal graph”, with uniform characteristics located between two nodes. In the network graph, it is represented as an edge between two nodes. The network-based categorization of street segments proposed in this research is based on the integrated centrality measures, including betweenness centrality, closeness centrality, degree centrality, and PageRank centrality.

1 Betweenness centrality (Cb)

\[ C_b(u) = \sum_{i,j \neq u} \frac{m_{ij}(u)}{m_{ij}} \]

where \( C_b(u) \) is the \( C_b \) of a given segment \( u \), \( N \) is the total number of segments in the network graph, \( m_{ij} \) is the number of shortest paths connecting \( i \) and \( j \), while \( m_{ij}(u) \) is the number of shortest paths connecting \( i \) and \( j \) and passing through \( u \).

2 Closeness centrality (Cc)

\[ C_c(u)= \frac{1}{\sum_{n=1}^{N} S(u, v)} \]

where \( C_c(u) \) is the \( C_c \) of a given segment \( u \), \( N \) refers to the total number of segments in the network graph, and \( S(u, v) \) refers to the length of the shortest distance between the segment \( u \) and the other segment \( v \).
3 Degree centrality ($C_d$)

$C_d$ (Freeman, 1979) reflects connectivity by calculating the number of neighboring segments to which each segment is directly connected. The more ties a segment has, the more critical it is in the network (Crucitti et al., 2006). In a network graph with $N$ segments, the maximum possible degree value of a segment is $N-1$, and the $C_d$ of segment $u$, $C_d(u)$, is calculated by the following formula:

$$C_d(u) = \frac{d_u}{N-1} \quad (3)$$

where $C_d(u)$ is the $C_d$ of a given segment $u$, $d_u$ is the number of segments directly connected to the segment $u$, $N$ is the total number of segments in the network graph, and $N-1$ is the maximum possible degree value of a segment in the network.

4 PageRank centrality ($C_p$)

$C_p$ (Page et al., 1999) ranks the segment by the importance of its neighboring segments. A segment is important if it links to another important segment. $C_p$ can be formally expressed as

$$C_p(u) = \frac{1 - d}{N} \sum_{v \in B_u} \frac{C_p(v)}{\text{NumLinks}(v)} \quad (4)$$

where $C_p(u)$ is the $C_p$ of a given segment $u$, $d$ is the probability of randomly moving to another segment, $N$ is the total number of segments in the network graph, $B_u$ is defined as the set of segments that link to segment $v$, $v$ is each possible segment that connects to segment $u$, and $\text{NumLinks}(v)$ is the number of links on segment $v$.

2.3 Workflow of the proposed NSC method

The proposed method of NSC compromises seven steps (see Supplementary Figure S2):

Step 1. Define the location and boundary of the study area.

Step 2. Construct the network graph from OSM within the defined boundaries.

The function of `graph_from_place()` in the OSMnx python library is used to access the OSM repository and query the OSM Application Programming Interface (API) Overpass to automate the
Characteristics of the USN in Braunschweig

The spatial and statistical distributions of the computed centrality measures are visualized to describe the USN’s structural and topological characteristics. The spatial distribution of these measures presents the locations of the topologically important and, therefore, more “central” segments in the entire network, whereas the statistical distribution summarizes the structural and topological characteristics of the Braunschweig USN.

1 Betweenness Centrality (Cb)

Supplementary Figure S3A presents the results of the distribution of intermediary capacity, measured by Cb (see also Table 1). It shows that the frequency distribution of Cb clearly follows the power law behaviour, which means that, in the case of Braunschweig, a large number of segments have small Cb, and a small number of segments have large Cb.

The spatial distribution of segments with Cb higher than the 75 percentile is coloured red in Supplementary Figure S1A. These segments have higher intermediary capacity. Because the Cb is calculated globally over the whole network, segments with higher Cb are more critical in traveling over the entire city. They connect different areas of the city.

2 Closeness Centrality (Cc)

Supplementary Figure S3B and Table 1 present the results of the distribution of accessibility, which is measured by Cc. The frequency distribution of Cc follows a normal distribution and the majority of the segments concentrate between −0.566 and 0.584, which is also shown by the 25 and 75 percentiles (Table 1).

The spatial distribution of the street segments with Cc higher than 75 percentiles is coloured in dark blue in Supplementary Figure S1B. These segments are more critical regarding their accessibility to all other segments. Supplementary Figure S1B exhibits a clear pattern that segments in the centre of the study area have higher Cc, while segments in the outer areas tend to have lower Cc. Apart from the expected lower centrality of these areas, this also indicates an overwhelming sensibility of the outer segments to the edge effects (Gil, 2017), as explained in section 2.3.

3 Degree Centrality (Cd)

Supplementary Figure S3C present the results of the connectivity distribution, measured by Cd. The majority of the segments have a Cd between −0.7334 and 0.7608, which is also shown by the values of 25 and 75 percentiles in Table 1.

The spatial distribution of segments with high Cd is presented in Supplementary Figure S1C, where segments with a Cd higher than 75 percentiles are coloured in green. It is noteworthy that, in contrast to Cc, the segments in the neighbourhoods have higher values of Cd, whereas the segments forming the backbone of the urban street network, such as the motorways or primary roads, have a value of Cd lower than 75 percentiles (segments coloured in pink). This means that the segments in-between the neighbourhoods have lower Cd, while segments within the neighbourhoods have higher Cd. This indicates that one of the structural and topological characteristics of the Braunschweig USN is that the segments used to travel between the neighbourhoods are often not connected to many other segments. Within the neighbourhood, the segments are well-connected to many other segments.

4 PageRank Centrality, (Cp)

Supplementary Figure S3D shows that an exceptionally large number of the segments concentrate in 5.3330e-17, the average Cp of Braunschweig. This means that most segments are linked to the segments whose importance is at the average level. Only a small proportion of segments are linked to important and very unimportant segments.

The Cp spatial distribution is presented in Supplementary Figure S1D, where segments with a Cp higher than 75 percentiles are coloured in orange. A unique pattern in the study area is that even if a segment is on the edge of the network, which is the peripheral area, its own level of importance is increased as long as it is connected to an important segment.

The computed centrality measures are the base for the following categorization of the street segments by means of cluster analysis.

Step 5. Cluster analysis for segment categorization based on centrality measures

The categorization of segments of similar characteristics was performed through Hierarchical Cluster Analysis (HCA). HCA is an
unsupervised machine learning approach for building a hierarchy of clusters. In contrast to other clustering approaches it has the advantage of not requiring to determine the final number of clusters beforehand, thus providing open and unbiased results. HCA is applied to the raw segment data to form clusters based on common factors, i.e., the four centrality measures, among various segment data points. In this research we utilised the AgglomerativeClustering function of the sci-kit-learn python library to carry out the HCA. The parameter value of affinity, which indicate the metric used to calculate distance between instances, is “Euclidean”. The “Ward’s method” is chosen to be the aggregative method. The function of “hopkins()” in the pyclustertend python library is imported to calculate the Hopkins’s statistic, which provides the numerical proof of the clusterability of the data. The resulting Hopkins score of 0.026651, which is positive and between 0 and 1, indicates that the data is not uniformly distributed and, therefore, clustering can be useful to categorize the observations. Also, the “silhouette_score()” function of the sci-kit-learn python library is imported to calculate the silhouette coefficient, which is used to evaluate the goodness of the chosen number of clusters. The results suggest that the best number of clusters should be two, according to silhouette score of 0.53. However, this limits the description of a complex USN in particular with regard to the assessment in terms of urban planning and design. Therefore, a varying number of clusters formed by means of HCA were examined for their feasibility (see, e.g., Supplementary Figure S5). We randomly selected 50 street segments across the entire USN. Based on high resolution satellite imagery (IFF, 2020), we visually assessed these and neighbouring segments in each cluster along the specifications of the street layout. In this contextual assessment of the clusters, six clusters proved to be conclusive, which has the silhouette score of 0.29. In addition, the FCS also describes six categories, indicating the need for a greater number of clusters and furthermore facilitating comparison between the NSC and FCS. A dendrogram, in which the euclidean distances of 0.2 was selected as the threshold value (indicated by the horizontal line above the axis), is shown in Supplementary Figure S4. Cutting the hierarchical tree into six clusters returns a vector of cluster labels indicating their memberships. This vector is then attached to the original segment data frame for visualization and summary statistics in the next step.

Step 6. Visualize the spatial and statistical distribution of NSCs to describe each NSC’s characteristics.

The cluster analysis results are mapped utilizing QGIS, and street segments are coloured by their categories to visualize the spatial distribution of the NSC categories (see Figure 3). Descriptive analysis of the centrality values is further facilitated by a diagrammatic representation in order to examine the statistical distribution of the four centrality measures within the individual NSC categories (see Figure 2).

Step 7. Compare the NSC with FCS.

Comparison between the function- and network-based schemes is carried out by plotting the stacked bar chart of the percentage of NSC within each FCS (see Figure 4). The cumulative distribution function (CDF) of the four centrality measures of each category within FCS and NSC is presented in Figure 5. The ks_2samp() function in the python module of scipy.stats is then performed to carry out the 2-sample Kolmogorov–Smirnov test (KS), which allows us to compare FCS and NSC and examine whether they have the same distribution.

3 Results

3.1 Characteristics of the NSC categories

The segments are grouped into six clusters as a result of the HCA. This forms the basis for the further definition of the NSC categories, also with regard to spatial aspects such as segment length relevant from the perspective of urban and transport planning. The characteristics of each cluster forming the NSC categories are described by statistical distribution presented in Table 2(a) and Figure 2. The spatial distribution of all categories is mapped in Figure 3. A description of the characteristics is also provided in Table 2(b). The following steps illustrate the process of defining and detailing the NSC from the initial clusters created.

In the course of the detailed analysis of the clusters, it became evident that one of the distinct and recognizable factors is the average length of the segments in each cluster. This is also plausible from the perspective of urban (transportation) planning, e.g., with regard to connectivity, and it can be seen spatially that the clusters with the shorter segments describe high-density networks, for example, in the inner-city area (see also Figure 3). Therefore the NSC groups are divided into two groups based on their average length. NSC 1 to 4 are characterized as long segment groups, and NSC 5 and 6 are represented as short segment groups using the city-wide average length (137 m) as a threshold.

| TABLE 1 Descriptive statistics of the z-score of the four computed centrality measures. |
|-------------------------------------|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                                     | Mean             | Standard deviation | Min          | 25%          | 50%          | 75%          | Max          |
| Betweenness Centrality             | −2.7232e-17      | 1.0              | −0.3577      | −0.3451      | −0.2984      | −0.1064      | 14.9193      |
| $C_p$                              |                  |                  |               |              |              |              |              |
| Closeness Centrality               | 3.6310e-17       | 1.0              | −7.0932      | −0.5659      | −0.01195     | 0.5841       | 2.8153       |
| $C_c$                              |                  |                  |               |              |              |              |              |
| Degree Centrality                  | 2.9048e-16       | 1.0              | −2.9747      | −0.7334      | 0.01372      | 0.7608       | 2.2550       |
| $C_d$                              |                  |                  |               |              |              |              |              |
| PageRank Centrality                | 5.3330e-17       | 1.0              | −3.8071      | −0.1199      | 0.0678       | 0.1103       | 7.5850       |
| $C_p$                              |                  |                  |               |              |              |              |              |
The results of network-based with the function-based categories

Next, the long- and short-segment groups are further differentiated based on the values of their four centrality measures. Table 2(a) and Figure 2 shows that average $C_b$, $C_c$, and $C_d$ values decrease from NSC 1 to NSC 6. On the other hand, the average values of $C_d$ increase gradually from NSC 1 to NSC 6.

Among all the categories, NSC 1 has the highest average values in the three centrality indices $C_a$, $C_b$, and $C_c$. This means that segments in NSC 1 have the highest level of intermediary capacity, accessibility, and connection with very important neighbouring segments. However, it also has the lowest average $C_d$ (0.00030), which means that they are connected with the smallest number of other segments and, therefore, has the lowest connectivity. As shown in Figure 3B, segments in NSC 1 play an essential role in connecting the northern and southern areas of the city. Figure 3A shows that NSC 1 is only connected to segments in NSC 2, which explains the low average $C_d$. On the whole, NSC 1 can be characterized as long segments with low intermediary capacity, accessibility, connection with very important neighbouring segments, and low connectivity.

Among the long segments, NSC 2 and 3 have medium intermediary capacity (average $C_b = 0.06343$ and 0.0273) and accessibility (average $C_c = 0.03515$ and 0.032688). The main difference between NSC 2 and 3 concerns their connectivity, where NSC 2 has a lower average $C_d$ (0.00035) than NSC 3 (0.00043). Another difference can be found in the connectivity distribution, $C_d$ as shown in Figure 2C. Compared with NSC 3, NSC 2 has lower connectivity, which means that segments in NSC 2 are connected to a smaller number of segments. By contrast, the broader distribution range of $C_d$ in NSC 3 indicates that segments belonging to this category are connected to a large and small number of other segments. In sum, NSC 2 and 3 are characterized as long segments with medium intermediary capacity and accessibility, whereas NSC 2 has lower connectivity than NSC 3.

Among the long segment groups, NSC 4 has the lowest level of intermediary capacity (average $C_b = 0.00872$) and accessibility (average $C_c = 0.03025$). However, the level of connectivity, measured by average $C_d$ (0.00045), is the highest among the NSCs with long segments. This means that the segments of NSC 4 are connected to more other segments, compared to NSC 1, NSC 2, and NSC 3 segments. As shown in Figure 3E, the ring road along the outer edge of the historical city center and the roads connecting central and outer areas start to appear in NSC. Their roles in the network system can be defined as connecting the neighbourhoods distributed across the city, and at least one of the nodes of these segments is connected to multiple segments. This explains why its connectivity is high. Therefore, NSC 4 is characterized as long segments with low intermediary capacity, accessibility, and high connectivity.

NSC 5 and 6 share the common feature of having short segments with the lowest level of intermediary capacity, as they both have the lowest average $C_b$ (0.00099 and 0.00078). Furthermore, they both connect to segments that are not very important, which is indicated by their lowest average $C_d$ (both are 0.00009). On the other hand, their highest average $C_d$ (0.00047 and 0.00048) among all NSC categories means that the number of segments they are connected to is the largest, which means NSC 5 and 6 have the highest connectivity. The most obvious feature that can distinguish NSC 5 and 6 is their level of accessibility, $C_c$, as shown in Figure 2B. NSC 5 has higher average $C_c$ than NSC 6. This means that NSC 6 is more likely to be found in the outer peripheral areas than NSC 5, as shown in Figures 3G, H. Therefore, NSC 5 and 6 can be described as short segments with low intermediary capacity, connecting to neighbouring segments with low importance and high connectivity.

3.2 Comparison between the network-based and the function-based method

The comparison between the FCS and the NSC is divided into three steps. First of all, the percentages of different NSC categories within each FCS class are presented in Figure 4 to distinguish the topological heterogeneity within each FCS class and increase the information such as accessibility and connectivity in each class. Secondly, the cumulative distribution function (CDF) of the four centrality measures within each category in FCS and NSC is visualized in Figure 5. Finally, the KS is carried out to examine whether FCS and NSC have the same distribution and how similar these two schemes are.

1 Comparing the results of network-based with the function-based categories

NSC 1 and the majority of NSC 2 segments belong to motorways in FCS (Figure 4; Supplementary Table S2), reflecting the characteristics of NSC 1 and 2 as long segments with high intermediary capacity, accessibility, connection to very important neighbouring segments and low connectivity, all of which are the typical features of motorways.

Secondly, more than 36% of the motorways and primary roads are categorized as NSC 5 and 6, characterized by short segments with low intermediary capacity, connecting to neighbouring segments with
There is a mismatch between some features of these segments and the typical expectations regarding the motorways and primary roads. Such roads are expected to carry heavy loads of traffic volume, but topologically their centrality values are low. These segments should be the location for collecting further data regarding the actual traffic volume.

At the same time, more than 30% of the tertiary and 3% of the residential roads in the NSC 3 and 4 have middle levels of importance in terms of intermediary capacity and accessibility. Thus, an inevitable mismatch between the expected traffic volume and the topological features is also observable with regard to these roads.

Finally, segments belonging to the "unclassified" group in the FCS are perhaps the most interesting for further investigation. Based on the FCS, it is unclear how much traffic should be expected for these segments. However, based on the NSC, more than 10% of these segments can be clearly categorized as belonging to NSC 3 and 4.

In sum, the segments with conflicting or indecisive categorizations between FCS and NSC require further investigation. On-site data collection potentially allows the discrepancy to be validated or determined.

### Comparing cumulative frequency distribution of centrality measures within each category

Further investigation in this section aims to explore to what extent the FCS deviates from the NSC. The cumulative distribution function of the four centrality measures within each category in the FCS and the NSC is used to visualize their distribution patterns. To quantify the differences between the FCS and NSC categories, descriptive statistics and specific characteristics of each NSC category are provided below.

<table>
<thead>
<tr>
<th>NSC Category</th>
<th>Length</th>
<th>Betweenness centrality: Intermediary</th>
<th>Closeness centrality: Accessibility</th>
<th>Degree centrality: Connectivity</th>
<th>PageRank centrality</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSC 1</td>
<td>Long</td>
<td>High</td>
<td>High</td>
<td>Lowest</td>
<td>Most connected to important neighbouring segments</td>
</tr>
<tr>
<td>NSC 2</td>
<td>Medium</td>
<td>Medium</td>
<td>2nd lowest connectivity among long segments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSC 3</td>
<td>Medium</td>
<td>Medium</td>
<td>2nd highest connectivity among long segments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSC 4</td>
<td>Low</td>
<td>Low</td>
<td>Highest connectivity among long segments</td>
<td></td>
<td>Least connected to important neighbouring segments</td>
</tr>
<tr>
<td>NSC 5</td>
<td>Short</td>
<td>low intermediary, high connectivity, connecting to neighbouring segments with low level of importance and high connectivity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSC 6</td>
<td>Short</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### TABLE 2

(a) Descriptive statistics and (b) specification of the characteristics of each NSC category.

<table>
<thead>
<tr>
<th>NSC Category</th>
<th>(C_b)</th>
<th>(C_c)</th>
<th>(C_d)</th>
<th>(C_p)</th>
<th>Length (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entire study area</td>
<td>-2.72E-17</td>
<td>3.63E-17</td>
<td>2.90E-16</td>
<td>5.33E-17</td>
<td>137.87</td>
</tr>
<tr>
<td>NSC 1</td>
<td>1.00008</td>
<td>1.00008</td>
<td>1.00008</td>
<td>1.00008</td>
<td>188.09</td>
</tr>
<tr>
<td>NSC 2</td>
<td>11.07996</td>
<td>2.004781</td>
<td>-1.34465</td>
<td>3.102888</td>
<td>884.18</td>
</tr>
<tr>
<td>NSC 3</td>
<td>2.06078</td>
<td>0.55207</td>
<td>0.71572</td>
<td>2.61676</td>
<td>643.56</td>
</tr>
<tr>
<td>NSC 4</td>
<td>5.11387</td>
<td>1.73304</td>
<td>-0.92313</td>
<td>0.33594</td>
<td>406.75</td>
</tr>
<tr>
<td>NSC 5</td>
<td>1.99691</td>
<td>1.11581</td>
<td>-0.30213</td>
<td>0.19387</td>
<td>271.59</td>
</tr>
<tr>
<td>NSC 6</td>
<td>0.95385</td>
<td>0.58772</td>
<td>0.94852</td>
<td>1.93397</td>
<td>461.61</td>
</tr>
<tr>
<td>NSC 1</td>
<td>0.39487</td>
<td>0.505561</td>
<td>-0.13242</td>
<td>0.114004</td>
<td>177.28</td>
</tr>
<tr>
<td>NSC 2</td>
<td>0.31635</td>
<td>0.75614</td>
<td>0.96265</td>
<td>1.25615</td>
<td>269.87</td>
</tr>
<tr>
<td>NSC 3</td>
<td>-0.27248</td>
<td>0.564619</td>
<td>0.002011</td>
<td>-0.05537</td>
<td>119.04</td>
</tr>
<tr>
<td>NSC 4</td>
<td>0.10503</td>
<td>0.48230</td>
<td>1.00145</td>
<td>0.96455</td>
<td>126.27</td>
</tr>
<tr>
<td>NSC 5</td>
<td>-0.29024</td>
<td>-0.80204</td>
<td>0.102035</td>
<td>-0.04173</td>
<td>116.26</td>
</tr>
<tr>
<td>NSC 6</td>
<td>0.08675</td>
<td>0.79458</td>
<td>0.97546</td>
<td>0.76475</td>
<td>130.22</td>
</tr>
</tbody>
</table>
FIGURE 3
Map of USN colored by the identified categories through the proposed network-based method: (A) NSC 1 to 6, (B) NSC 1, (C) NSC 2, (D) NSC 3, (E) NSC 4, (F) NSC 5 and (G) NSC 6.

FIGURE 4
Visualization of distribution of different NSC categories within each FCS.
and the NSC, the KS, which is a nonparametric method in statistics for comparing two samples, is then carried out to measure the greatest distance and detect the difference in both the location and the shape of the cumulative distribution functions of the FCS and the NSC.

The null hypothesis of the KS is that the two samples were drawn from the same distribution, and there is no difference between the FCS and the NSC. The resulting KS value is evaluated by the $p$-value (see Table 3) to decide whether this null hypothesis is rejected. The significance level of the $p$-value is
set at 0.05. If \( p > 0.05 \), it is assumed that the null hypothesis cannot be dismissed, which means that the FCS and the NSC are sampled from the same distribution. This is the case with the C2 of category I, the C3 of category III, and the C4 of category II (see Figure 5; Table 3). This means that among all the categories in the FCS and the NSC, only these three pairs can be sampled from the same distribution for a 5% significance, and the distributions of these three pairs are equal. In other words, the topological similarity between the FCS and the NSC only exists in categories I and III in terms of the C2 and in category II in terms of the C4. For all the other categories, there is no topological similarity between the pairs of the FCS and the NSC because their \( p < 0.05 \), so the null hypothesis is rejected.

TABLE 3 KS value and \( p \)-value (in bracket) from KS test. Between FCS and NSC, the topological similarity only exists in categories I and III because their KS values are not significant enough, and the null hypothesis that the two samples are from the same distribution need to be accepted.

<table>
<thead>
<tr>
<th></th>
<th>Cb</th>
<th>Cc</th>
<th>Cd</th>
<th>Cp</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.9223*** (5.551e-16)</td>
<td>0.3996* (1.887e-03)</td>
<td>0.2734 (7.688e-02)</td>
<td>0.6184*** (5.511e-08)</td>
</tr>
<tr>
<td>II</td>
<td>0.9963*** (0.000e+00)</td>
<td>0.6146** (0.000e+00)</td>
<td>0.2094* (1.918e-02)</td>
<td>0.0890 (7.761e-01)</td>
</tr>
<tr>
<td>III</td>
<td>0.8914*** (4.441e-16)</td>
<td>0.3521*** (4.441e-16)</td>
<td>0.0926 (7.159e-02)</td>
<td>0.1333*** (2.096e-03)</td>
</tr>
<tr>
<td>IV</td>
<td>0.5736*** (2.998e-15)</td>
<td>0.2415*** (4.441e-16)</td>
<td>0.1979*** (1.643e-14)</td>
<td>0.0826** (6.786e-03)</td>
</tr>
<tr>
<td>V</td>
<td>0.0945*** (8.212e-12)</td>
<td>0.5299*** (5.551e-16)</td>
<td>0.0678*** (2.717e-06)</td>
<td>0.1332*** (5.551e-16)</td>
</tr>
<tr>
<td>VI</td>
<td>0.1875** (5.560e-03)</td>
<td>0.5360*** (3.331e-16)</td>
<td>0.1566* (3.270e-02)</td>
<td>( k = 0.2171*** (7.517e-04) )</td>
</tr>
</tbody>
</table>

Note: *** \( p < 0.001 \), ** \( p < 0.01 \), * \( p < 0.05 \).

4 Conclusion

The authors introduce a network-based method for categorizing the segments of urban street networks. By calculating the centrality measures of the segments and applying cluster analysis, street segments with similar topological characteristics are grouped together. These categories are characterized and their spatial distribution within a city is analyzed. The results show that six network-based segment categories can be distinguished based on different levels of intermediary capacity, accessibility, connectivity, and levels of importance of the neighbouring segments.

The adaptation of clustering as a numerical categorization technique, facilitates the integration with other analytical and planning procedures in urban (transport) planning and design. One of the advantages of the method is that when extensive data-collection or prior local knowledge of the real performance of streets are not available or too cost- and resource-consuming, it provides an objective categorization from available open data or self-generated data based on e.g., satellite imagery (Verma et al., 2021).

The results of this research can serve as the initial step for further analyses, such as proposing appropriate types of interventions according to street segments’ characteristics also in terms of spatial design and evaluating the resilience and robustness of the network against disruption or congestion. For example, the segments with a high level of topological importance could be analysed in detail by setting up sensor technology to collect data to analyse the traffic flows of various transportation modes (motorized, cycling, walking) to explore whether the topological importance is associated with a higher proportion of all or specific modes of transport. The quantitative descriptions of street characteristics produced by this method is suitable for the comparisons between different urban areas based on which research regarding network qualities in different urban fabrics is possible.

This study furthermore focused on the development of the methodology to compare FCS with NCS to identify the potential linkages and mismatches between the two schemes. The heterogeneity within each FCS class can be identified, and street segments of the same functional class can be further distinguished by their topological importance. The comparison between FCS and NSC contributes to the understanding and description of the subtle differences in the topological properties of the segments within the individual FCS classes. It was found that these are not congruent. For example, in the case of Braunschweig, the segments belonging to “motorway” have very diverse topological characteristics because all six categories of the NSC can be found in this FCS category. Also, among the “unclassified” category in the FCS, more than 10% of segments have the middle level of importance regarding their intermediary capacity and accessibility in the NSC. Finally, the cumulative distribution function results and the KS provide evidence that the FCS significantly deviates from the NSC in the case of Braunschweig. Except for the top two classes (i.e., motorway and primary in FCS and NSC 1 and 2), no topological similarity is found between the pairs of FCS and NSC. With these results, this study intends to initiate and stimulate further discussion on the extent to which street segments should be categorized based on their functionality, topological importance, or both of these attributes combined.

This can potentially serve to describe challenges and opportunities for transportation planning with respect to specific sections of the network. For transport authorities dealing with transport infrastructure investments, maintenance, design, and operation, the quantitative assessment of street segments’ performance is essential. However, in practice, close monitoring of the performance and patterns of movement flows in the entire urban street network is not always possible due the cost- or resource constraints. The proposed approach in the current research and the comparison between the function- and network-based schemes can help to (a) identify mismatched segments (FCS different to NCS); and (b) the most critical central places in the network, where real-time data could be collected with view to monitoring the efficiency and accessibility of the overall transportation networks, and to identify the segments that require further planning within each FCS class.

A concrete example of application could be the examination of residential streets, where speed limits are usually imposed, so that noise levels are generally expected to be lower. However, if the NSC analysis shows that a residential street has a potentially high level of use and connectivity, installing noise sensors in the street section in question enables to check whether the noise level is higher than expected on residential roads.
Another potential application relates to network planning for different modes of transport: if a segment has a high level of topological importance but is currently without e.g., a bike lane, it may be a potential candidate for adding the respective infrastructure to increase the bike network’s overall connectivity and quality. From the perspective of encouraging more sustainable and active transportation modes, increasing the functionality of a street segment may also contribute to higher network connectivity for pedestrians or cyclists.

Overall, the network-based categorization of street segments (NSC-method) can be understood as a (spatially) high resolved tool to analyse and describe urban street networks (USN). In addition to the further investigation and validation of the method as such—e.g. regarding the scalability, advancement of the algorithms and criteria adopted or research on spatial structural effects of different street networks on this categorization method by applying this approach to more—particular data describing the built environment of the respective city, town, urbanized area, and zillow neighborhood.

16 (6), 1019–1030. doi:10.1177/0265813516650678

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbuil.2023.1216888/full#supplementary-material

References


Frontiers in Built Environment frontiersin.org 12


