Cities 55 (2016) 9–21

Contents lists available at ScienceDirect

Cities

journal homepage: www.elsevier.com/locate/cities

Urban function connectivity: Characterisation of functional urban streets with social media check-in data

Yao Shen *, Kayvan Karimi

Space Syntax Laboratory, Bartlett School of Architecture, University College London, England, United Kingdom

ARTICLE INFO

Article history:
Received 9 November 2015
Received in revised form 15 March 2016
Accepted 25 March 2016
Available online 13 April 2016

Keywords:
Network accessibility
Connectivity
Social media check-in data
Land use
Street network
Urban design

ABSTRACT

Social media check-in data, one type of crowdsourcing open data about individual activity-related choices, provides a new perspective to sense people's spatial and temporal preference in urban places. In this paper, through the analysis of the interaction between these scored places on streets, we aim to advance our knowledge of network accessibility with social media check-ins to portray urban structure and related socioeconomic performance more explicitly. By conceptualising an interface graph to reflect the interplay between land-use points and the co-visual paths, we propose a novel framework to characterise the urban streets with land-use connectivity indices that are measured with a new type of place-function signature. A "3-Ds" model is introduced to package three principal dimensions of urban function network, including accessible density, accessible diversity and delivery efficiency, as one integrated index that works towards a comprehensive understanding of function connectivity from each street's midpoints to all reachable land-use points. Streets are further partitioned to the annotated function regions based on function connectivity in different types of active land-use. The results of preliminary studies in the city of Tianjin, China show that the proposed metrics can explicitly describe the inherent function structure and the regions' typology across scales. Compared with space syntax measurements at the same radius for describing the variation of empirically observed house price, the integrated metric can improve the predictability of statistic models sufficiently, and each specified index is confirmed to be statistically significant by controlling other factors. Overall, this research shows that the usage of ubiquitous big social media data can enrich the current description of the urban network system and enhance the predictability of network accessibility on socioeconomic performance.

1. Introduction

The recent growth in the provision of location-based services has inspired people to share their location preferences in social media networks, through which the ubiquitous user-generated information of location choices has produced a new generation of 'human knowledge' regarding urban spaces at a fine-grained resolution (Wyly, 2014). Although such a provision offers new opportunities for research, the use of social media data has its limitations, including sampling problems, context-related uncertainty, and lack of theoretical composition (Boyd & Crawford, 2012). However, the finer resolution of these datasets has the potential to enable people to ask different questions than those based on conventional data that is oftentimes aggregated and out of date (Shelton et al., 2015). As a new type of fine-scaled datasets that contains detailed information about urban land-use, Points-of-Interest (POIs) and Check-in data in modern social media are a new focus in urban studies, resulting in research that identifies urban regions with POIs and taxi trajectories (Yuan et al., 2012), that determines the characteristics of urban parcels using vector cellular automata (Liu & Long, 2015), that maps urban areas of cities (Long et al., 2015), etc. Simultaneously, geo-tagged data have been increasingly discussed with reference to modelling human mobility patterns. Studies have applied individual-based check-in datasets to map mobility patterns (Hasan et al., 2013), to calibrate the parameters of distance decay (Wu et al., 2014), to infer daily activity clusters (Hasan & Ukkusuri, 2014; Jiang et al., 2012), to validate retail store replacement (Karamshuk et al., 2013) and to analyse urban structure (Long & Thill, 2015; Ratti et al., 2010; Zhong et al., 2015). All these efforts imply that now the open 'big data' can represent the variations of place in people's minds. However, no prior study has examined the function connections among various land-uses based on social media check-in data, other than Liu et al. (2016), who classified land-use clusters using spatial interaction patterns between parcels. The aforementioned studies have instead focused more on origin–destination patterns and have not considered the impact of physical layout on the detailed spatial interaction between real urban public spaces in their models that limit the contribution of location-based 'big data' to urban morphology, land-use planning and urban design studies.

* Corresponding author at: Bartlett School of Architecture, University College London, 140 Hampstead Road, NW1 2BX London, United Kingdom.
E-mail address: y.shen.12@ucl.ac.uk (Y. Shen).

http://dx.doi.org/10.1016/j.cities.2016.03.013
0264-2751/

An important issue that should be noted in any urban design and planning task is that streets are the fundamental spatial elements for movements; they interlink urban functions physically and cognitively (Gehl, 2011; Jacob, 1993). Urban function locations, as the destinations of urban activities, interact with one another through a network of urban public spaces and formulate a network-based land-use system. Consequently, the distance-based and cognitive proximity between urban land-uses influence not only the movement but also the choices of land-use locations (Geurs et al., 2015). A vibrant urban location can revitalise the urban context, in which it is embedded; additionally, the proximity of active urban locations can increase the popularity of the places that are connected to them. Nevertheless, few examples have focused on the importance of the role of streets and the physical layouts in the traditional models of accessibility, which are based on assumptions that do not recognise spatial heterogeneity in the urban space (Batty, 2009). This gap has been addressed with the configurational studies, such as the work undertaken by researchers from the space syntax community. These studies have demonstrated that the network properties of urban grids can adequately capture the influence of cognitive efforts on pedestrian movement patterns (Hillier et al., 1993), car volume distributions (Hillier & Iida, 2005), land-use distributions (Penn & Turner, 2004; Scoppa & Peponis, 2015; Shen et al., 2013) and other socioeconomic issues (e.g., Hillier, 2007; Karimi, 2012; Vaughan, 2007).

There has been some criticism of configurational studies for not considering the impacts of land-use distribution or other attractors on the spatial network analysis (Ratti, 2004), but these criticisms seem to lack a certain depth in understanding of theoretical and methodological propositions of space syntax (Hillier & Penn, 2004). In fact, the space syntax model is capable of offering significant potential for further development precisely because it links cognitive costs to navigational energy expenditures in spatial analysis (Kim & Penn, 2004). However, the topological/geometrical interaction between the land-uses through streets is an important dimension in which to scrutinise the underlying structures of functional streets that are typically neglected in conventional studies on land-use distribution and accessibility (Geurs et al., 2012). Recently, by taking into account the reachable densities of activities distribution in the space syntax model, Ståhle et al. (2005) developed a toolbox called 'place syntax' to calculate accumulated opportunities within the buffers defined by the metric/topological/geometrical radius. With the emphasis on the value of perceived density, Marcus and his team suggested that the space syntax model could be extended to a more general concept, the 'spatial capital', with the possibility of translating the urban form to other social, economic and cultural capitals (Berghauser Pont & Marcus, 2014; Marcus, 2010). Simultaneously, other areas of focus include modelling the interplay between reachable metric distance and directional distance to enhance standard space syntax in predicting human pedestrian patterns (Ozbil et al., 2011; Peponis et al., 2008), analysing transit riderships (Ozbil et al., 2009), and modelling the pattern of commercial frontage (Scoppa & Peponis, 2015). These studies implicitly considered the detailed possibility of improving the existing space syntax model, but did not propose a systematic perspective.

In this paper, we propose an original method for computing urban function connectivity by considering the spatial interaction between the scored urban spaces and partitioning the urban streets based on the composition of the defined spatial interactions. Social media check-ins are used to infer the significance of a place for a specific type of active urban function, to weight respectively the *accessible density* and *accessible diversity*, and to measure the *delivery efficiency* and the so-called *urban function connectivity*. A statistical data mining approach is adopted to characterise urban streets with the similar composition of function accessibilities for different types of land-uses. The proposed method is applied in a case study in Tianjin and its feasibility is verified by confirming the enhancements of the predictability of the statistic models that capture explicitly the variation of the house prices.

2. The method

2.1. Preliminary definition

In this study, *urban function connectivity* (*UFC*) is defined as the relatedness information between land-uses through the street networks, representing the sense of function potentials from every street's midpoint to all the reachable land-use points. This particular form of connectivity, therefore, is constructed on the basis of the street network where urban land-uses are assigned spatially. An *urban function region* (*UFR*) is identified as a group of places where the properties of function connectivity for different active land-uses are similar. Apart from the conventional definition of the functional region for comparing economic development in regional studies (Antikainen, 2005; Williams et al., 2010), we use this format of UFRs to refer to the clusters of streets within which urban functions operate similarly. Given this definition, we introduce an alternative approach to partition urban space from the bottom up by considering the spatio-functional relationships in a specified land-use system.

The land-use system in this study is conceptualised as a *path-point model* (PPM), or as a '*network interface model*' (NIM) to abstract the co-existential relationship between urban function points and the visual paths as graphs. In such a model, scored urban function locations (points) are assigned to the nearest paths based on their spatial interlinkage which is identified as the interface between buildings and public spaces (Alexander et al., 1977; Hillier & Hanson, 1984). By converting the spatial relationship between the main elements in PPM/NIM to edges and nodes, the land-use system can be transformed to an interface graph/network. The land-use locations and the directly visible street segments are defined as 'function nodes' and 'segment nodes', respectively, whereas the interfaces (the directly physical relationship) between nodes – including the entrances from the street to the locations and the intersections between the roads – are identified as 'entrance edges' and 'intersection edges'.

Fig. 1 illustrates the basic conceptual method used to construct an interface graph step-by-step. We first prepare the necessary data maps including the road network and land-use pattern so that these entities can be transformed into function and segment nodes with the entrance edges in an interface map on the basis of their interface connections. In the following stages, the dual graph of the interface network is created by converting the street junctions to the intersection edges that connect the segment nodes and assigning the cognitive cost at every junction to the graph as the weights of those intersection edges. The cognitive cost for the intersection edges is specified as the angular change at each junction according to space syntax theory (Dalton, 2000, 2001; Haq, 2003; Hillier & Iida, 2005; Kim & Penn, 2004; Turner, 2001a) and earlier evidence in the field of cognitive neuroscience and way-finding (e.g., Bailenson et al., 1998, 2000; Crowe et al., 2000, Montello, 1991). Using angular-weighted adjusted graphs in a simple land-use system, we represent the manner in which the angle change through a journey along the shortest path is calculated (Fig. 2). As the current evidence suggests that humans are not sensitive to very slight directional change (Figueiredo, 2009), a cut-off angle is used to filter the imperceptible angular deviations (α) from straight lines to enable a more appropriate approximation of the real movement decision making. Urban streets are the basic spatial units for the function connectivity model, as they are the real conduits for human movement.

Notably, scores or any other information can be used to weight the function nodes to capture the various levels of the significance of urban functions. In this research, check-ins and POIs derived from social media service providers are adopted to present the diverse types of urban activity locations and the proxies for the relative preferences of people in urban destinations. By adding weights for the function nodes, many aspects of function connectivity can be addressed to develop a comprehensive and robust methodology.

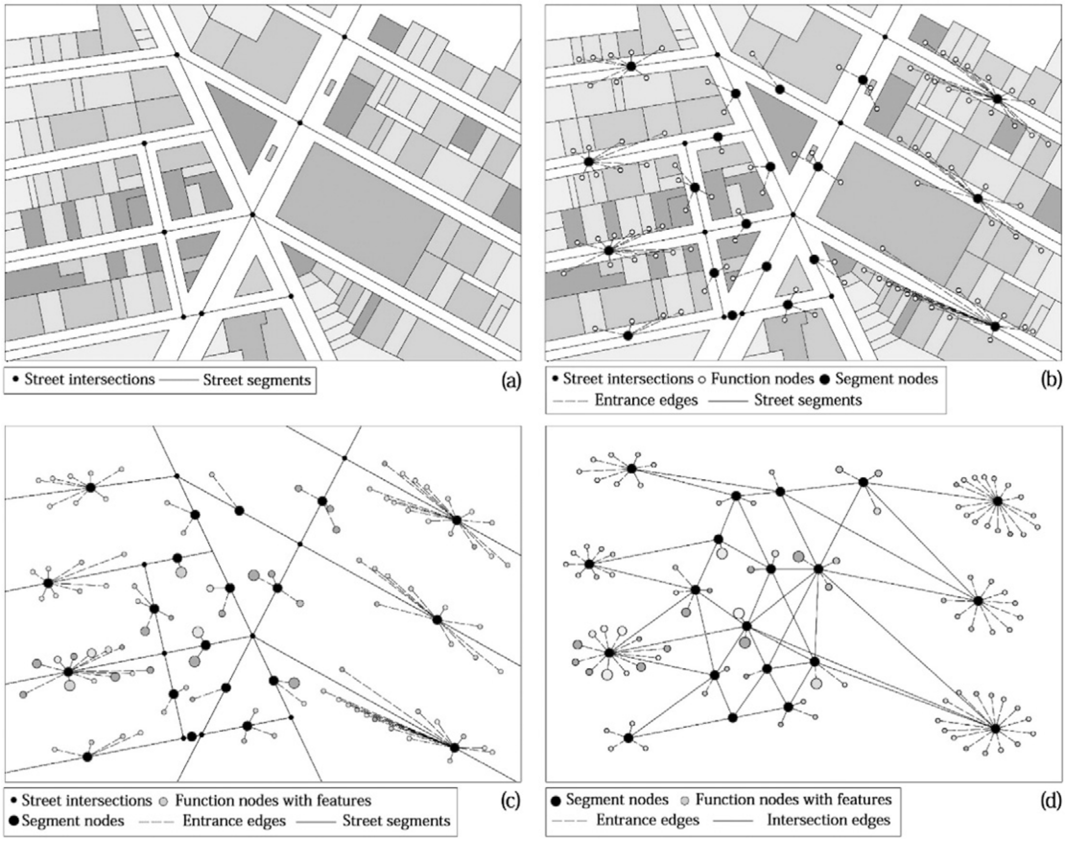

Fig. 1. The PPM (or NIM): (a) road network and land-use distribution; (b) the interface map of street network featured by land-uses; (c) the interface map of street network and POIs and their popularity estimated by social media data; (d) the dual graph of the interface network as a public space network and POIs. (The greyscale of POIs refers to the typology of activities, and the size of the points shows the check-in intensity.)

2.2. The framework for characterising urban streets

We introduce a stepwise framework to identify the various dimensions of UFC and UFRs, which contains several main modules, including data preparation, interface graph formation, function connectivity computation and function regional characterisation (Fig. 3).

2.2.1. Data preparation

In the first module, the dataset is processed using a standard GIS procedure. The initial road network dataset should be cleared and readjusted to an *angular segmental map* that corresponds better to reality based on visibility and walkability. An important part of this process is to transform the road network to an 'axial model', which will then be

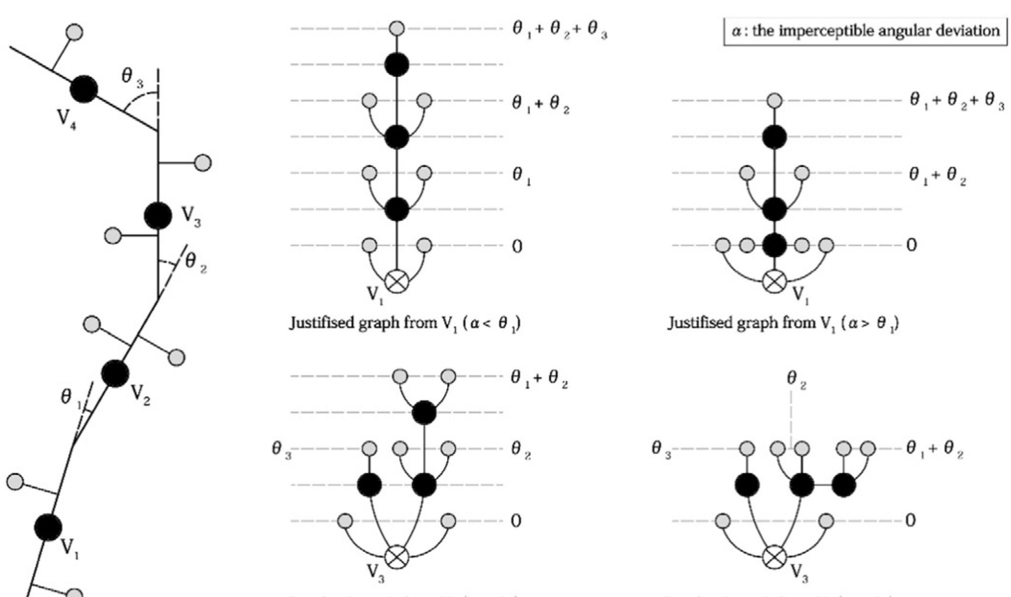

Fig. 2. The shortest path through a network and its associated angular justified graphs from different segment nodes (V_1 and V_2) with different settings for cut-off angles (($\alpha<\theta_1$) and ($\alpha>\theta_1$)).

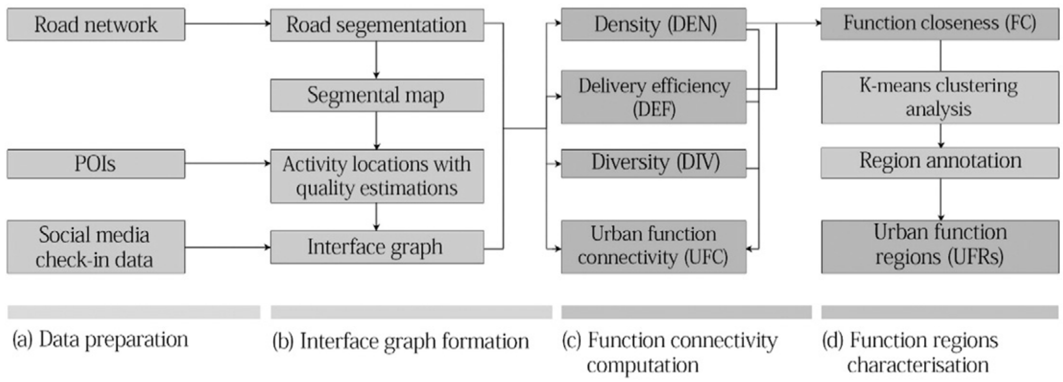

Fig. 3. Stepwise framework for identifying urban function connectivity (UFC) and urban function regions (UFRs).

segmented to create the segment model needed for this study. In previous space syntax studies, the segment models created from an axial model has been shown an efficient method of capturing the movement and navigation in cities. The road network data are first simplified and then split at the real road intersections. In order to avoid the large curves of the road network, they are transformed to straight segments according to the degree of their curviness (Fig. 4), following a segmentation method suggested by Liu and Jiang (2012) to convert street central lines to axial segments. Specifically, we use the deviating distance from the base line that links the two endpoints of all segments to the farthest vertex in the segment lines to reflect the curviness of segments. The curved segments will be cut at the farthest vertex for calculating deviating distance if their deviating distance is longer than the average (Jiang & Liu, 2010). This process will be repeated until all curves are transformed. In so doing, an objective description of *angular segmental map* is formed based on the notion of visibility.

The POI dataset is collected and geocoded with the street network. The POIs are then reclassified as the required main types of urban activities. The social media check-in data are then linked with the POIs based on the tags and coordinates after filtering the fake points, including the check-in locations placed outside of the study area and the locations that have misfits between the coordinates of the check-in point and cell phone GPS for generating clean data, which reflects all real usage of the land-use locations.

2.2.2. Interface graph formation

To draw the interface graph and perform the related computation, we combine spatially the segment map and the social media check-in data on the GIS (Geographic Information System) platform. The POIs are inferred with their check-ins features and snapped to the most proximal segments, whereby the interface relationship can then be appropriately modelled. We have used 15° as the cut-off angle for defining the perceptible angular change and calculate the effective accumulated angular change to a reachable destination as a numeral variable to reflect the cognitive cost between a place and the functions accordingly. Given that humans can only easily recognise significant differences between two turns, angular step depth – a discrete description of the angular change – would be more appropriate for describing a sensible angular change for humans. We define the angular depth at every angular intersection as an integer that rounds up to the quotient, in which the numeric angular change is divided by a designated interval. In this study, we assume that 45° is the project interval for defining the angular depth. For instance, if the angular change at intersection A and B are 35° and 95°, the angular depth for these two angular transits will be 1 and 2, respectively.

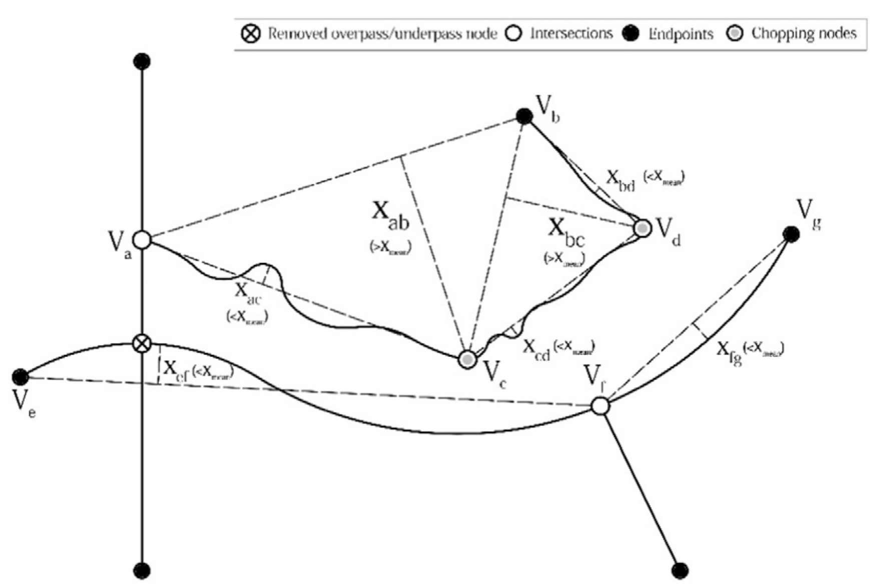

Fig. 4. Re-definition of the road central line data to angular segment map (X is a deviating distance for an included curve/segment. The segments with a deviating distance greater than the mean value are cut at the farthest vertices in the segments, and the divided segments are re-evaluated in terms of their curves and further cut until their deviating distance are smaller than the mean value.)

2.2.3. UFC computation

We consider UFC based on three principal dimensions that are summarised as '3-Ds' model, in which accessible density, diversity and delivery efficiency are calibrated. In attempting to achieve comprehensive understanding of the interplay among these three aspects, we package them as an integrated index to balance the methodological complexity and the simplicity of result interpretation.

2.2.3.1. Density (DEN). The DEN index measures the accumulation of scored urban activities from each street within a defined radius through the reachable shortest paths. Assuming there are K types of active land-uses, the accessible function density for the segment node i at radius r would be aggregated as $DEN_{(i,r)}$:

$$DEN_{(i,r)} = \sum\nolimits_{k=1}^{K}\sum\nolimits_{j=1}^{J} O_{(j,k)} \times W_{(j,k)}, \quad \{dist(i,j) < r\} \tag{1}$$

This summation considers the function nodes that are assigned to street segment edges and weighted based on the social media check-in scores. In the equation above, r is the defined radius, and $W_{(j,k)}$ is the specified weight for the function node j in type k. The scores for the function nodes are relativised according to the list of defined land-use types including retail, catering, and hotel, and measured as the normalised check-ins which can be presented as $W_{(j,k)} = \frac{\log C_{(j,k)}}{\log C_k^{max}}$, where $\log C_{(j,k)}$ represents the log-normalised check-ins for the specific function node j in type k, and $\log C_k^{\max}$ denotes the log-normalised value of the maximum check-ins for all the function nodes in the built graph.

2.2.3.2. Diversity (DIV). The DIV index measures the balance degree of all reachable weighted urban activities from the original street within a given radius. Urban studies have found that the concentration of diverse land-uses contributes to urban vitality and sustainability (Jacobs, 1961; Coupland, 1997) with less energy consumption (Newman & Kenworthy, 1999; 2011) and higher level of social cohesion (Pendola & Gen, 2008). Diversity can be measured in several ways, but the most popular methods include the dissimilarity index (Cervero & Kockelman, 1997) and the entropy method (Chuvieco, 1999). In this study, information entropy is applied to measure the diversity of urban function nodes from segment node i at the radius r, and represented as $Div_{i,r}$ ($Div_{i,r} = [0, 1]$). Further, a normalisation process has been applied to enable the different types of activities to be comparable. A direct way apply such a process is to convert the absolute density to a relative density by dividing the accessible weighted density in type k for each segment node by the maximum value of the accessible density of land-use of the same type at the same radius for all the segment nodes within the study area ($NaDEN_{(i,k,r)} = \frac{DEN_{(i,k,r)}}{DEN_{(k,r)}^{max}}$). The computation of accessible diversity can be formally represented as follows.

$$DIV_{(i,r)} = \frac{-\sum\nolimits_{k=1}^{K} P_{(i,k,r)} \times \ln\left(P_{(i,k,r)}\right)}{\ln(K)}, \quad \{dist(i,j) < r\} \tag{2}$$

$$P_{(i,k,r)} = \frac{NaDEN_{(i,k,r)}}{\sum\nolimits_{k=1}^{K} NaDEN_{(i,k,r)}} \tag{3}$$

The presence probability ($P_{(i,k,r)}$) of the function nodes in type k at radius r for segment nodes i is measured by its empirically observed frequency of normalised density ($NaDEN_{(i,k,r)}$) among all K types of land-uses.

2.2.3.3. Delivery efficiency (DEF). The DEF index measures the mean angular shallowness to all the reachable urban activities from the original street within a given radius through the shortest paths. This index is the reciprocal of the angular step depth, revealing the cognitive efficiency of land-use delivery from all reachable functions to the original street segments beyond the same energy expenditure that is measured in the light of the metric length of the streets. This index can be formally expressed in the following equation.

$$DEF_{(i,r)} = \frac{N_{(i,r)}}{\sum\nolimits_{k=1}^{K}\sum\nolimits_{j=1}^{J} ASD_{(i,j,k)}}, \quad \{dist(i,j) < r\} \tag{4}$$

In the equation above, $ASD_{(i,j,k)}$ shows the angular step depth from segment node i to function node j in type k within the buffer area defined by radius r, and $N_{(i,r)}$ is the summation of the accessible functions at the same radius. Notably, the average angular step depth is inverted in this measurement so that the segment node, which is 'closest' to all reachable function nodes at metric radius r, will have the highest efficiency.

2.2.3.4. Urban function connectivity (UFC). The UFC index is a composite measurement that measures the degree to which the dense and diverse urban activities are accessible with less angular step depth within a given radius. Here, three principal dimensions in the 3-Ds model reflecting the impacts of opportunity accumulations, function composition and cognitive distance are incorporated into the final UFC index ($UFC_{(i,r)}$) which can be calculated formally as follows:

$$UFC_{(i,r)} = {DEN_{(i,r)}}^{DIV_{(i,r)}} \times DEF_{(i,r)}, \{dist(i.j) < r\} \tag{5}$$

Here, the impact of the interplay between density and diversity on function connectivity is quantified by a power function, which recently has been applied as an elasticity parameter in measuring job accessibility (Cheng & Bertolini, 2013). In the light of the foregoing, the product of these two factors (${DEN_{(i,r)}}^{DIV_{(i,r)}}$) will be 1, when $DIV_{(i,r)}$ is equal to 0, and will be $DEN_{(i,r)}$ if $DIV_{(i,r)}$ is 1.

2.2.4. Urban function regions (UFRs) characterisation

Urban streets are connected to different types of functions, in which the UFRs are characterised by the function connectivity in different land-use types. In this essence, we apply the statistical data mining approach to quantitatively measure the similarity of the function connectivity composition. Specifically, we use a k-means clustering analysis to partition the urban street segments and then annotate each cluster according to the unique composition of function connectivity.

2.2.4.1. Function angular closeness (FAC). The FAC index is a particular form of the function connectivity with the focus on the specified type of land-use, and it measures the angular agglomeration of the urban function of a certain type through the shortest reachable urban paths within a given radius. The computation logic of this index follows the idea of establishing so-called angular closeness which is computed as the quotient in which node counts are divided by the mean angular step depth in the space syntax model. Mathematically, this metric can be identified in a straightforward way as follows.

$$FAC_{(i,k,r)} = DEN_{(i,k,r)} \times DEF_{(i,k,r)}, \{dist(i,j) < r\} \tag{6}$$

In the equation above, $DEN_{(i,k,r)}$ refers to accessible density of function nodes in type k from segment node i at the radius r, and $DEF_{(i,k,r)}$ captures the angular delivery efficiency of these functions.

2.2.4.2. Urban function region (UFR). Within the family of statistical data mining approaches, many algorithms can be used to address the question of grouping multi-dimensional data as clusters. These algorithms include hierarchical clustering, two-step clustering, and the self-organisation map (SOM). In this study, k-means clustering for several states is employed by using the FACs of each street as the vectors' dimensions due to its efficiency in handling large-sized numerical datasets (Bishop, 2006). In this method, streets maintaining similar function connectivity in all the defined types will be redistricted to several function regions. As its name implies, k-means clustering intends to group objects into predefined k clusters where every object in the same cluster will have the nearest mean. Consequently, the objective of k-means clustering in this study is to minimise the total intra-cluster variance, which is measured by the squared errors. As a type of iterative descent clustering algorithm, k-means clustering can be summarised as follows:

$$J_{(c,r)} = \min_{C} \sum_{l=1}^{L} \sum_{C(i)=l} dist\left(V_i, \overline{V}_l\right), \{dist(i,j) < r\} \tag{7}$$

where $J_{(c,r)}$ is an objective function for a given cluster assignment C at a radius r, $C(i)$ refers to the label that the observations have, $\overline{V}_l$ is the mean vector for the l th cluster, and V_i is a multi-dimensional vector illustrating the co-presence of function accessibility of various land-uses ($V_i \in (FAC_{(i,1,r)}, \text{FAC}_{(i,2,r)}, \ldots\ldots, \text{FAC}_{(i,K,r)})$). This process will be repeated iteratively until the grouping results are stable with a minimised sum of squares.

One well-known problem of k-means clustering is the problem of cluster validity. In other words, we must evaluate the results and select the optimised number of clusters, which can hardly be decided before the analysis (Halkidi et al., 2001). Some metrics have been developed in previous studies for validating cluster numbers, such as, Dunn's Index (Dunn, 1973), Davies–Bouldin index (Davies & Bouldin, 1979), Silhouette Index (Rousseeuw, 1987), Xie and Beni's Index (Xie & Beni, 1991) and others. In this paper, we use Dunn's Index and Silhouette Index as the validation measurements to evaluate the most proper number of clusters (Brun et al., 2007). The former index emphasises maximising the inter-cluster distances and minimising the intra-cluster distances, whereas the latter index focuses on the clustering strength of each observation by measuring the mean compactness and separation of clusters.

2.3. The settings

Because the proposed method is a trade-off approach, the *radius* refers to the metric distance thresholds applied to select the set of functions from the entire system to be analysed from the root segments. In this work, the distance of the radius is measured along the street segments. Four radius thresholds are specified to represent the spatial scales of the analysis, namely, 500 m (super-local scale), 1000 m (local scale), 2500 m (lower semi-local scale), 5000 m (higher semi-local scale) and 10,000 m (global scale).

Active land-uses in this study are defined as the complementary land-uses that are more likely to be linked by urban travels and thereby contribute to emergent movement patterns. Unlike mixed-use developments, which seek a balance of all land-uses, the active uses in this study are based on function complementarity between non-residential land-uses (Hess et al., 2001). Complementary land-uses (active land-uses in this work) include *retail*, *catering*, *hotel*, *office*, *school*, *social services*, *hospital*, *recreation*, *culture*, *park and transport* according to the main activity types that are distinguished in the social media. Although the overall effects of the mixture of complementary land-uses through streets are of great concern, the check-in behaviour for different active land-uses will exhibit different frequencies. Consequently, the way in which activities are classified will influence the weighting results for those functions thereby impacting the final results of function connectivity. Therefore, classifications of POIs should consider the internal similarity of check-in behaviours for different types of land-uses in order to score the specific function appropriately. For example, retail and catering are two distinct categories because the probability that various shops are checked in is approximately 10%, whereas the same check-in likelihood for restaurants of different types is generally about 35%. We should also distinguish the culture land-use from recreation as an independent type due to the fact that about 80% of culture amenities are scored, whereas only 15% of other recreation facilities are featured in social media. In summary, we use complementary functions and check-in behaviours as two critical criteria to optimise the classification list of active land-uses and to secure the data reliability by controlling the bias.

3. Study area and data specifications

3.1. Study area

In this study, the central area in the Tianjin Metropolitan Area (TMA) is used as the case study for our empirical investigation. Tianjin functions as a major economic centre in north China, with a very strong connection with the capital, Beijing. With approximately 12 million inhabitants, Tianjin ranks the third city in China in terms of population. The study area is located in the centre of the administrative districts and covers 100 census survey units, i.e., 2081 km^2 with a population of nearly 10 million. The central area here reflects the spatial context in which the inner city is embedded. Tianjin is presented as test case to implement the proposed method for evaluating and visualising the interplay between spatial configuration and active urban functions. To better represent the street-based results, we select a rectangular area of approximately 48 square kilometres in the centre of the defined study area to scrutinise the complexity of the functionality information associated with the street network (Fig. 5).

3.2. Road network and social media check-in data

The road network used in this study was obtained from Tianjin's Surveying and Mapping Bureau in October 2014. There are 67,603 road segments within the boundary of the study area after converting the original road network to a segmental map (Fig. 5-a). The central area of Tianjin maintains a higher road network density and smaller block sizes than the suburban areas. A total of 127,258 POIs were obtained from a well-known social media platform, Weibo, which is the equivalent of Twitter in the Chinese context. This dataset was obtained from Weibo's streaming API in December 2014. At the same time, we also gathered 3,012,970 records of tagged tweets of 136,842 users across 35,220 avenues in the same areas. By joining the geo-tagged tweets in Weibo, nearly 30% of the POIs are featured with the check-in intensity. All of these check-in data cover the large built-up areas within the study area, suggesting that human activities are highly related to urbanisation (Fig. 5-b).

The POIs were then ranked based on the number of check-ins and users for all POIs, aiming to identify popularity of places in the public social media map (Fig. 6). The log–log plot patterns show that the regularity patterns follow a similar scaling law, in which there are far more POIs with fewer check-ins than POIs with many check-ins, suggesting that popular places in the public's minds will be more likely to attract more people and encourage them to revisit. In both diagrams, almost all of the upper half of the checked-in POIs follows a power law distribution, which reflects both the preferential selection process of people who visit urban places and the fact that less popular places are known to everyone. Therefore, the findings here illustrate that the scaling law represents people's preferences and the real usage of urban sites, which should be addressed in the accessibility modelling.

For computing the land-use inter-complementarity, the check-in locations are reclassified into eleven types of active land-uses that were previously defined (Table 1). Retail locations occupy the highest number of POIs, followed by catering, office and recreation, whereas most check-in records are concentrated in transport, education, catering and shopping categories. The check-ins without clear POI definitions have been removed from the dataset to generate a real site-related land-use map. In this sense, the temporal functionality distribution of an urban space can be properly derived from the ubiquitously available social media check-in data in our framework.

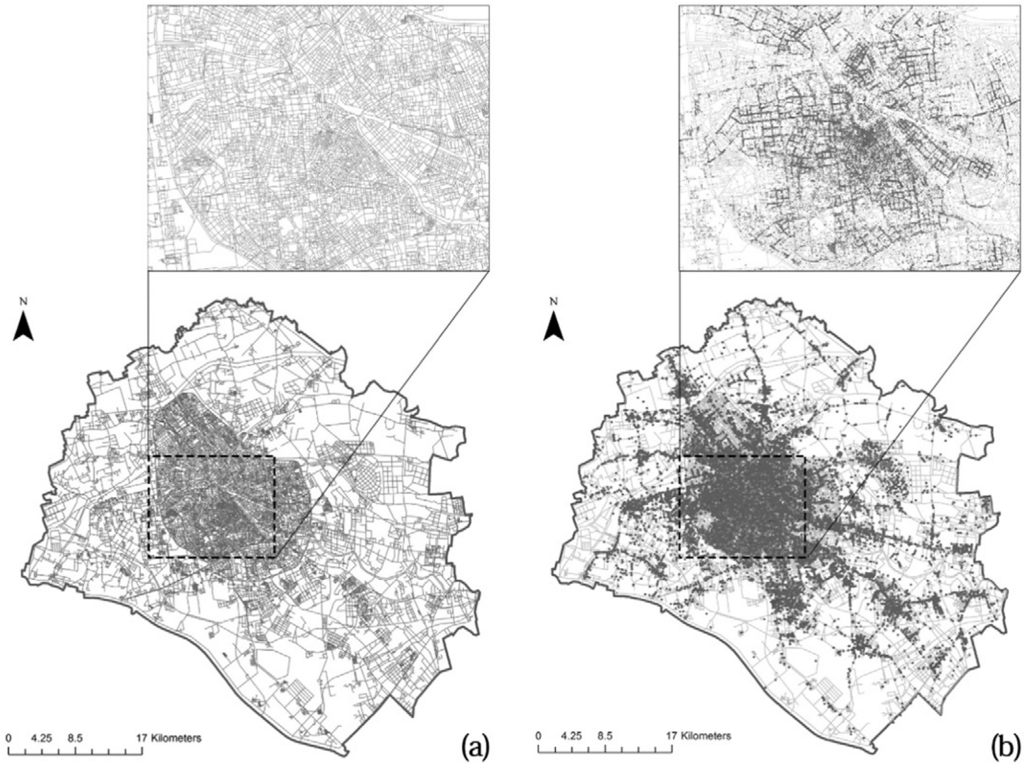

Fig. 5. Spatial girds (a) and POIs (b) in the study area.

4. Empirical results

4.1. Urban function connectivity for streets

The formerly defined indices in various aspects of UFC at different spatial scales have been mapped in central Tianjin (Fig. 7). The density maps (DEN) present the transformation of metric function agglomeration across scales from the polycentric local centre structure to a relatively homocentric one. The Central Business District (CBD) is empirically observed as the place with a multi-level synergy to generate the sense of multi-scaled centrality for density. Moreover, the maps of urban diversity (DIV) at different levels suggest a general similarity between land-use agglomeration and land-use mixture. However, many mismatched areas with different densities and diversity values can be clearly observed, where the land-use mixture is significantly high but the function density is relatively low, particularly at the super-local (500 m) and lower semi-local scales (1000 m). Meanwhile, the street-based urban centres highlighted by the diversity centrality are more polycentric than the ones highlighted by function density, indicating that functional mixture can occur at different degrees of function clustering. It implies that diversity adds valuable information to local centres, thereby depicting the urban structure more explicitly.

The results of the computed accessible density and diversity are relatively patch-like because they are based on the network metric distance without taking into account the cognitive dimension. In contrast, the results of the delivery efficiency (DEF) of street-based urban functions present network patterns that distinguish the active primary streets with more angularly connected active functions from the secondary streets that have deep angular connection in the denser built environment, illustrating the influence of the geometric properties of spatial structure on connecting land-uses through the street system. It can be argued that the primary streets with less angular distance to urban functions captures the super-block structure in modern cities or the continuous main roads in historic cities, whereas the secondary streets reflect the urban communities, in which land-uses are metrically proximal but angularly distanced from the original streets. In particular, the long and straight roads are more likely to be captured as the primary network on local scales, whereas the historically developed roads with a high degree of self-adjustment within the process of urban transformation are typically identified as the active primary structure at the larger

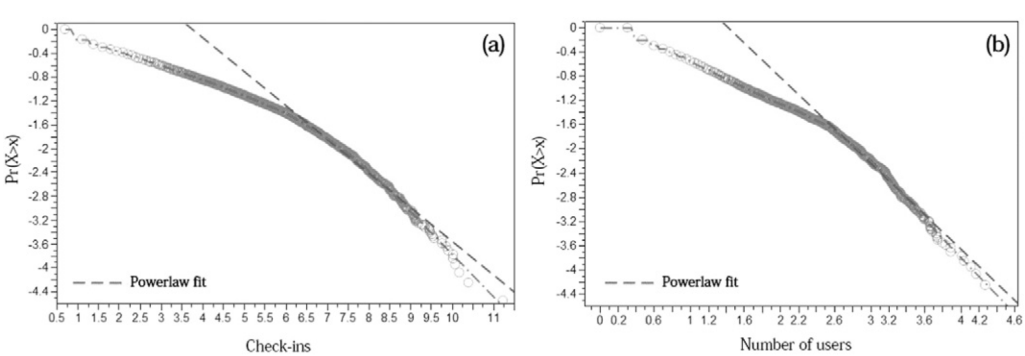

Fig. 6. Log–log plot of the probability of POIs against check-in number (a) and uses (b).

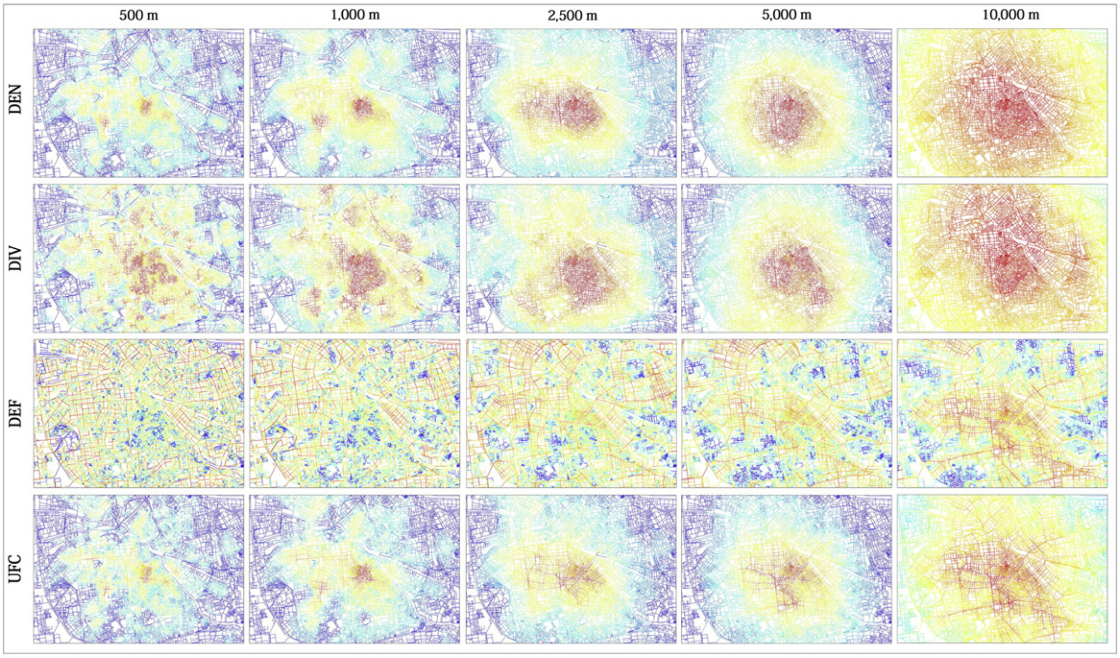

Fig. 7. Urban function connectivity maps (DEN: accessible density index; DIV: accessible diversity index; DEF: delivery efficiency index; UFC: urban function connectivity).

spatial levels. At the local level, the patches of the secondary streets are discretely distributed, but at the global scale, they are aggregated and interlinked by the primary structure. As discussed before, the geometric interaction between land-uses through streets is an important dimension for scrutinising the inherent structures of functional streets that are absent from conventional studies on land-use distribution.

The results of UFC in Tianjin are reported in the last panel of Fig. 7, illustrating how the street structure influences the locality of places in the built environment. Additionally, the changes in the function connectivity patterns between scales can further represent the relationship between functional centres through urban streets. The structures of the function connectivity at larger radii highlights the routes, through which the centres at smaller radii are interconnected to generate a global structure. In the case of Tianjin, the organic historic roads are the critical routes for the spill-over effects of the land-use agglomeration. In economic geography, the agglomeration economy is the driving force for urban development (Fujita et al., 2001). Based on this logic, the UFC pattern can represent developing trends if the function clustering is considered as the evidence of the occurrence of economic agglomeration. For instance, it seems that the local centre structure will move westward and combines the three local centres that emerge at the pedestrian levels. But the global centres might seek expansion to the south. Instead of using a single measurement for one dimension, it is suggested that the UFC based on social media data, might provide an alternative perspective for studying the morphological structure explicitly at the street level.

4.2. Characterisation of urban streets

Land use configuration is represented not only by the function connectivity centralities that are reflected in their positions in the street network, but also by their detailed compositions of connectivity in various land-use types. We rank segments based on the FAC for each activity category across scales and find that urban streets can be selected with diminish regularity patterns which follows a relaxed scaling laws (Fig. 8). With the increase of the scales, the cumulated distributions of FAC for each type of land-use tend to be more similar and cut-off values appear. Below the cut-off values at the global scales, probability of selecting a cell vary significantly, which implies that the significant difference between the highly urbanised areas and the non-urban areas.

Fig. 9 shows the results of the detected UFRs at different spatial scales and the associated information, including the final cluster centres for the detected regions, which are measured by the FAC in each function types, the cluster validation to select the optimised number of clusters and the transformation of clusters across scales. Table 2 summaries the annotations of the emergent clusters at various scales according to the emergent cluster centres represented in Fig. 9.

The UFRs at the super-local level (500 m) represent the most detailed characterisation, reflecting the heterogeneity of the patterns of FACs. Except for cluster C2 (non-central streets), all other clusters represent the central area of Tianjin, because the FACs for various active land-uses are ranked highly in those regions. Three FURs are captured due to the dominance of function connectivity in certain types of POIs. These specific clusters include the education streets (C1) on which most university campuses are located, the cultural streets (C7) with good connectivity to cultural amenities, historic interests and associated parks and the streets for travels and hotels (C8). Good examples in C8 are the well-known train stations in Tianjin (e.g., Tianjin station), which are crowded by hotels and are impacted by their roles as termini for intra-city travels.

Four UFRs are annotated as a group because the average angular closeness in all land-use types are relatively high: the office led central business streets (C3) with the highest ranks of the FACs to offices, public agencies, shops, recreation and restaurants; the developed commercial/recreational streets (C6), which are the destinations for shopping, eating and entertainment activities; the developed business streets (C4), where a large number of retail units, shopping centres, restaurants, offices, public agencies/organisations, recreation and transport nodes are angularly reachable through streets; and the developing business streets (C5), where the average levels of function connectivity are lower than those in C3, C4 and C6. Unlike the conventional ways of identifying function areas, these results demonstrate the complexity of the interplay between fine-gained functions, suggesting the possibility of studying street life at every spatial turn instead of using the aggregated – but somehow meaningless – zones.

At the local scale (1000 m), the number of UFRs is reduced to six, and less specified clusters are recognised. Many government agencies that

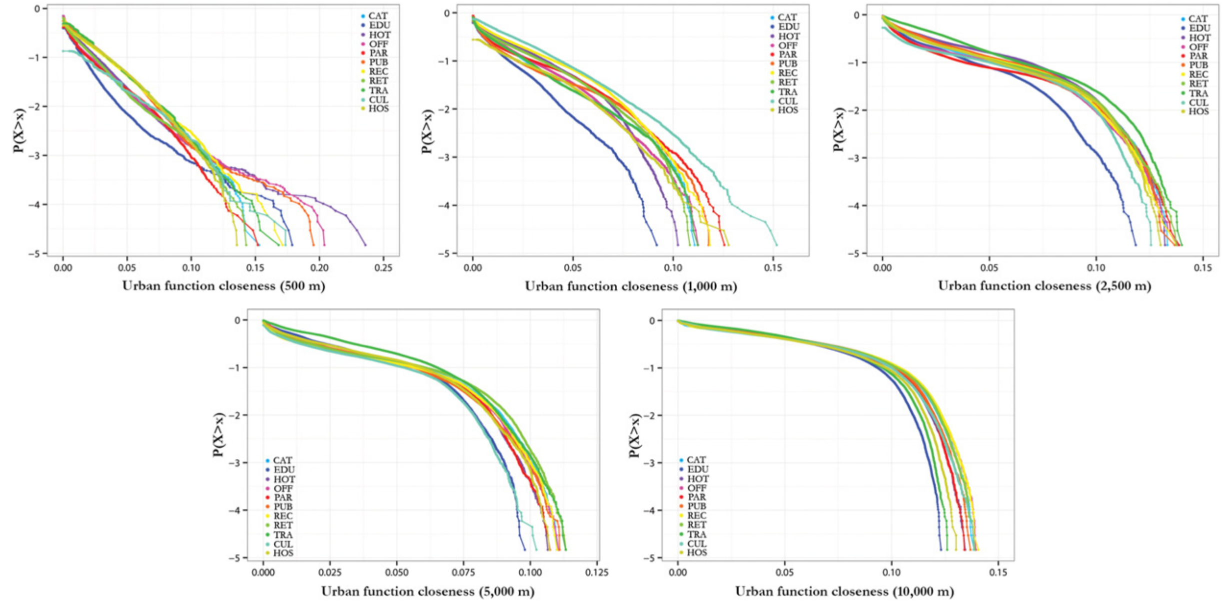

Fig. 8. Cumulated distributions of function closeness for each type of active land-use among urban streets.

offer public health care services, offices and amenities are clustered in the diplomatic and business streets (C5), which were originally located within the historic colonial areas, but have been transformed to become the modern diplomatic centre of the city in the 1970s. The central business streets (C2) are located around the famous pedestrian shopping street 'Binjiang Road', which is known as the main business centre in Tianjin, to form a pervasive rather than discrete cluster at the local level. The developed commercial/recreational streets (C3) are the segments where daily public activities significantly co-exist with parks and hotels. The patterns at this level illustrate the angularly pervasive

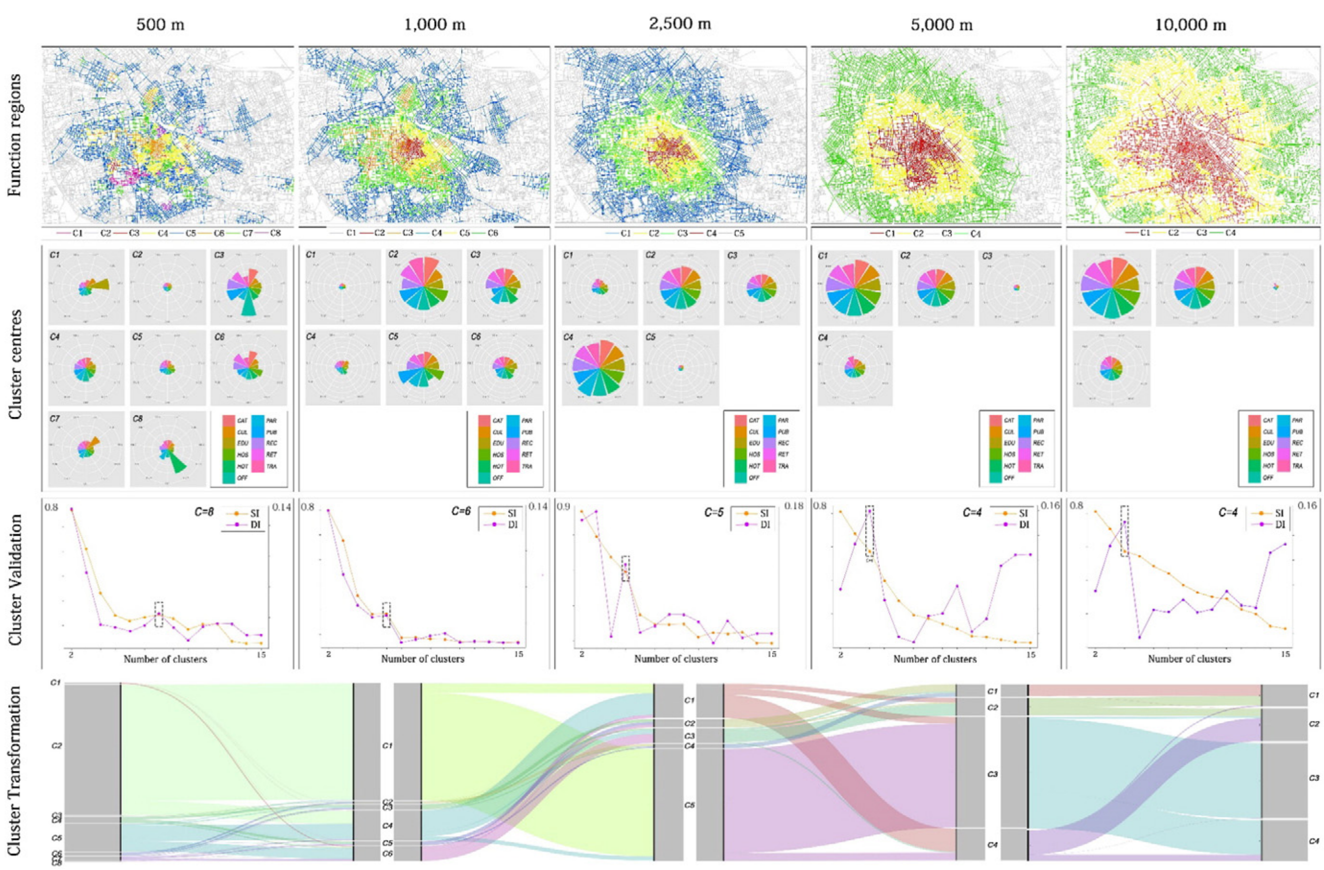

Fig. 9. Characteristics of functional streets (1st panel: the functional regions; 2nd panel: the final cluster centres measured by the intra-cluster average function angular closeness (FAC) in the individual land-use type; 3rd panel: cluster validation diagrams with Dunn index (purple) and Silhouette index (orange) for the number of clusters from 2 to 15; 4th panel: alluvial diagrams of cluster transformation across scales).

Table 1
Social media check-in data types and aggregated information.

Type	Abbreviation	POIs count	Check-in POIs count	Check-in number	Check-in users
Retail	RET	33,429	2884	209,888	123,340
Catering	CAT	27,485	9475	322,583	245,334
Hotel	HOT	2575	1298	87,872	49,614
Office	OFF	17,178	3017	167,510	63,913
Education	EDU	2016	1286	442,558	111,659
Public service	PUB	7277	1446	36,027	18,411
Hospital	HOS	2808	1298	85,523	41,548
Recreation	REC	14,045	2342	125,547	84,076
Culture	CUL	252	207	33,303	17,121
Park	PAR	2126	873	96,637	56,290
Transport	TRA	2025	1467	503,509	224,377

structure of functional regions, showing that land-uses would spill over convexly or relatively linearly according to the geometric features of the locations across space.

With the increase of a study scale, the detected optimised numbers of emerging clusters are further reduced and the distributions become more hierarchical and mono-centric from the city centre to the peripheral areas. At the semi-local scales and at the global scale, five and four clusters are recognised, respectively. These clusters are distinguished by the average degree of FACs in all types of land-uses as the result of similar degrees of accessible diversity at the larger scales. With regard to the transformation of these patterns on a large scale, it suggests that the city centre has shifted from the old core towards the south and the west, and the convex or linear patterns in different areas illustrate the shape of the UFC and the morphological shape of urban development.

The transformation of emerging UFRs across scales is also proof that urban streets, characterised by land-uses and street networks, perform differently at different scales. Evidently, the real urban function is highly mixed in the central area, which consequently makes it relatively difficult for people to discern the function characteristics of urban spaces. In addition, land-use patterns are continuously changing and are formed from the bottom up, piece-by-piece and street-by-street. The UFR is a scale-reliant concept; consequently, a certain function region may be merged into other regions, when the study scale is changed. Although increasing scales would lead to the simplification of urban regions, the geometric connection between the functions significantly influences the patterns of detected regions.

5. Validation and application

For evaluating our proposed method, house price data are adopted for the validity and the discussion of possible applications. Asking house price datasets, were gathered from online websites including 58tongcheng and Soufan in China. These data are building-based and consist of 10,015 samples (Fig. 10). Previous studies have shown that the network accessibility matters significantly in house price modelling, for which hedonic models are predominantly used (Law et al., 2013; Shen & Karimi, 2015; Xiao, 2012). The theoretical foundation of these studies is that locational characteristics are perceived as the environmental externality that will be directly reflected as a part of the price of properties (Boyle & Kiel, 2001). Furthermore, a reliable validation of network centrality should rely on the sample covering most of the study area to prove the universal effectiveness. Due to the sensitivity of housing prices to locational advantages and data coverage, it is argued that the effectiveness of the proposed metrics can be verified by means of the way in which people value the residential properties in this study. By using a comparison of the correlation coefficients in the Unary Linear Regression (ULR) and the t-values of variables in the Multivariable Linear Regression (MLR) analysis, we will explore the effectiveness of the indices in our methods for modelling urban performance in a manner that is favourably comparable to standard centrality metrics.

To validate the results generated by the proposed framework, we compare the predictability of statistical models in asking house price distribution by using the indices proposed in this study with other existing network-based centrality metrics and planar geometric indices. The distance to a predefined central business district (CBD) is selected as planar accessibility, and the segmental closeness and betweenness of the street network (also known as integration and choice in space syntax studies) are used as standard network centrality indices. The validation process is conducted in two scenarios: in the first scenario, each metric used in the comparison is treated as a variable in the ULR model and then the adjusted correlation coefficients are mapped and compared to reveal the predictability of UFC indices at each radius. In the second scenario, density, diversity, and delivery efficiency at a certain radius are entered in the MLR model with the standard network accessibility indices at the same radius as the variables to compare their statistical significance in a head-to-head manner. The metrics of standard network analysis in the space syntax theory are calculated in the Depthmap software developed by Turner (2001b) and Varoudis (2012), while the proposed measurements in the present study are computed in a self-developed toolkit in ArcGIS 10.2.

Four indices of accessibility centralities, distance to the CBD (Euclidean distance to the CBD), closeness, weighted and non-weighted function connectivity with social media scores, are computed and compared as a single variable in the ULR model such that the role of social media check-in data is shown in the first scenario (Fig. 11). Overall, the network measurements perform significantly better than the distance to CBD, which suggests that the definition of CBD is arbitrary and that network centralities can summarise the importance of urban spaces more strongly. Moreover, the UFC weighted by the social media check-ins, correlates with asking house price more significantly than the non-weighted measures or segmental closeness, highlighting the importance of check-in data for inferring the real functionality of point-based urban functions. The best correlation appears at 2500 m for the weighted UFC, with an adjusted R2 value of 0.426, whereas the peak of the correlation (0.367) between non-weighted composite accessibility and the house price is present at 5500 m. This finding demonstrates that the clustering pattern of human activity is more compact than the density of functions, whereby the economic externality of residential properties is captured at a relatively local scale by the function accessibility based on social media data. If we consider the economic externality of urban space as the key aspect of

Table 2
Function regions annotation according to the emerging cluster centres in k-means analysis.

Clusters at 500 m		Clusters at 1000 m		Clusters at 2500 m		Clusters at 5000 m	
C1	Education streets	C1	Non-central streets	C1	Developing business streets	C1	Central business streets
C2	Non-central streets	C2	Central business streets	C2	Highly developed business streets	C2	Developed business streets
C3	Office led central business streets	C3	Developed commercial/recreational streets	C3	Developed business streets	C3	Non-central streets
C4	Developed business streets	C4	Developing business streets	C4	Central business streets	C4	Developing business streets
C5	Developing business streets	C5	Diplomatic and business streets	C5	Non-central streets		
C6	Developed commercial/recreational streets	C6	Developed business streets				
C7	Cultural streets						
C8	Streets for travels and hotels						

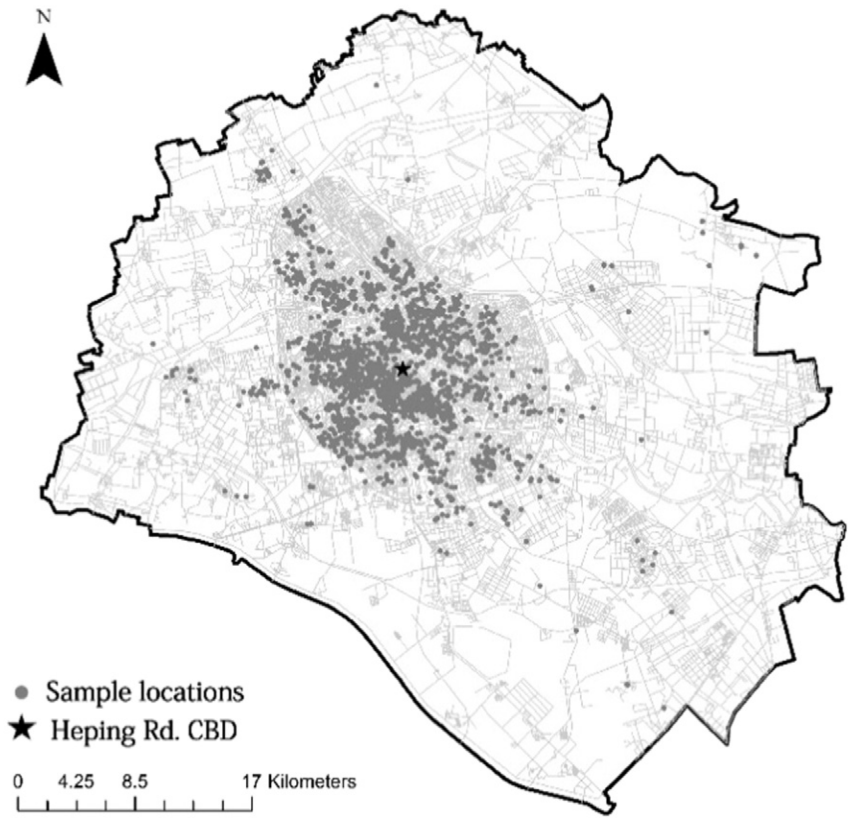

Fig. 10. Asking house price map of Tianjin city (RMB per square).

urban locality, a safe argument can be made that incorporating the geo-tagged social media check-in data with the spatial network can increase the accuracy of modelling place locality and its socioeconomic significance across radii.

The second test shows that all network centrality variables at different radii are statically significant, except for the delivery efficiency at 500 m (Fig. 12), which suggests that location centrality cannot be reflected by a single measurement; instead, it is impacted by the interaction between different types of network centrality variables, which emphasises further the significance of the interaction between spatial and functional elements in the built environment at various scales. The peak of the significance for the accessible density is found at approximately 5000 m. The closeness of the spatial network and the accessible density of urban functions experience a similar trend across all radii, in which function diversity is more significant at the local and semi-local scales from 500 m to 7000 m and becomes less

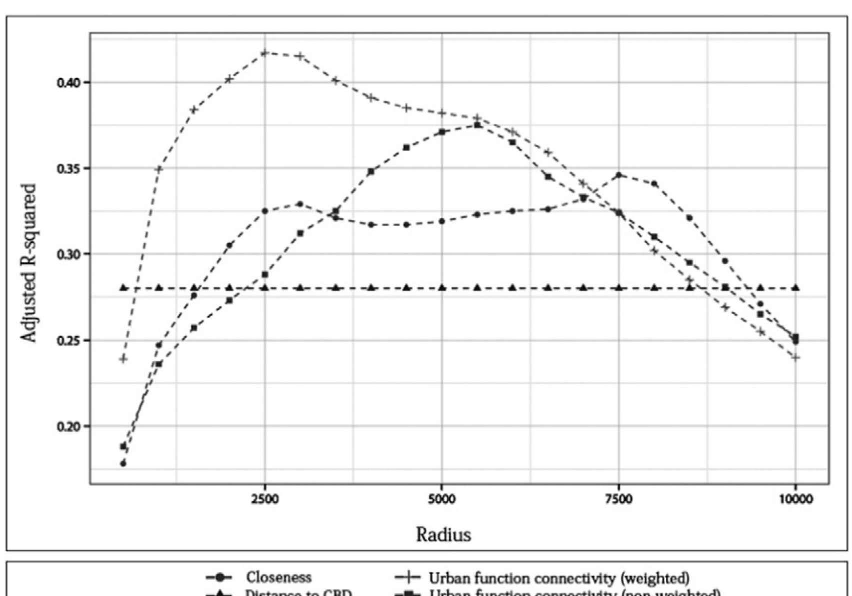

Fig. 11. Comparison of adjusted R-squared values of urban function connectivity with and without social media weights from social media, standard network accessibility index (closeness) and distance to the CBD in ULR models at different radii of modelling the variation of asking house price.

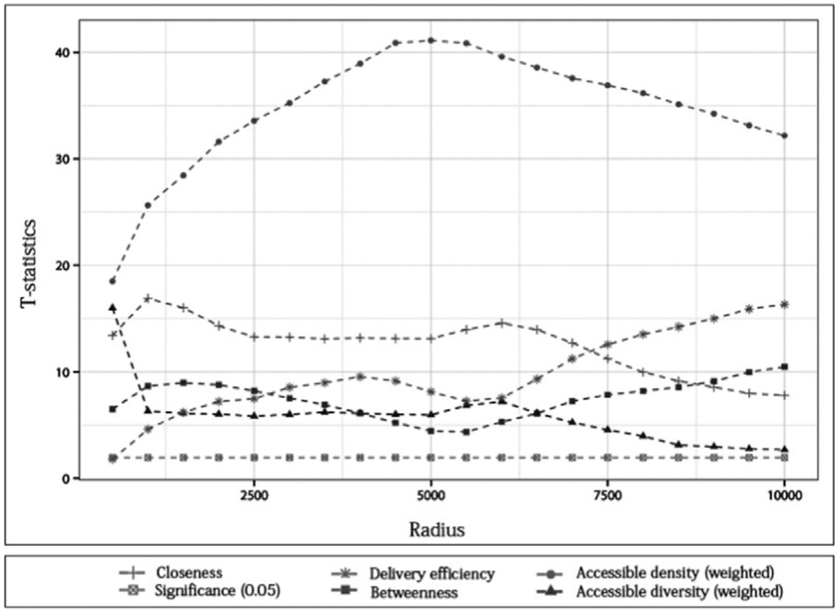

Fig. 12. Comparison of t-values of variables (weighted function density, weighted function diversity, delivery efficiency, angular closeness, and angular betweenness) in MLP models at different radii of modelling the variation of asking house price.

significant as the scale increases. By contrast, the significance of the betweenness variable and delivery efficiency index grows from 7000 m to 10,000 m, indicating the geometric or topological connection between land uses and that the route choice matters more significantly when other functionality information is more aggregated at the macro scale. At all radii below 5000 m, the significance of betweenness continuously falls and the t-value of the functional delivery efficiency rises. The shifting relationship among these variables across scales indicates that the proposed individual index of UFC can provide additional valuable information regarding existing descriptions of the spatial network to infer the housing price variations. The interaction between the spatial configuration and functional system characterises the locality of urban space in terms of modelling the physical and functional externality of housing properties.

6. Conclusion

With the aim to advance our knowledge of network accessibility to portray urban structures and related socioeconomic performance more explicitly in the new data environment, this study has proposed an analytical framework to characterise urban streets with function connectivity indices that are measured with a new type of place-function signature. We introduce a "3-Ds" model to integrate three principal dimensions of UFC patterns that include density, diversity and delivery efficiency into one integrated index that works towards a comprehensive understanding of function connectivity from each street's midpoints to all reachable land-use points. Based on computing a series of urban function angular closeness, a particular form of function connectivity in individual land-use type, urban streets have been grouped as UFRs with the mixture of urban function connectivity in different land-use types.

In an empirical study of Tianjin, the geo-referenced user-generated social media data reveal the sound dynamics of street-based spatio-functional structures and function regions at different radii. Using a current asking house price dataset, the computed results in our framework demonstrate that the integrated urban function connectivity index captures more explicitly the variation of locational externalities than existing network accessibility measurements for predicting the variation of residential properties' value across scales. Meanwhile, the measurements based on the three principal dimensions projected in this study are recognised to be statistically significant, controlling the impacts of spatial accessibility indices at every radius. The result of this study show that the proposed method enhances the understanding of the morphological structure of the land-use system and socioeconomic performance based on location-based social media data.

The main advantage of using this framework lies in its ability to capture the functional information though urban streets more efficiently with increasingly more available urban data. There are several contributions that this approach makes. Firstly, urban streets are characterised and grouped based on the interplay between spatial configuration and the visually interlinked land-use locations that are scored in social media. In this way, we provide a spatio-functional model in which location-based urban 'big data' can be properly utilised in morphological analyses. Secondly, social media check-ins are empirically tested to confirm their role in improving the accuracy of network centrality computation, particularly at the local scales. Thirdly, we extend the standard space syntax model by adding the influence of land-use attractions and improve the predictability of existing network centrality indices on socioeconomic performance. Fourthly, the UFC index balances methodological complexity and the interpretational simplicity of the proposed method, enabling its potential applications theoretically and practically. Finally, the proposed approach can potentially be applied in the urban design process to evaluate the effects of different spatial plans on connecting the land-uses and assess the detailed land-use plan and the allocation of facilities in the spatial and functional contexts, in which the study area is embedded. Subsequently, this approach promotes the advantages of a street-based model for planning and design at fine-grained scales. This, in turn, highlights the potential role of ubiquitous big social media data in an explicit and real-time description of urban systems, and drives further relevant studies.

References

Alexander, C., Ishikawa, S., & Silverstein, M. (1977). *A pattern language: Towns, buildings, construction, Vol. 2*, Oxford University Press.

Antikainen, J. (2005). The concept of functional urban area. *Findings of the Espon Project*, *1*(1).

Bailenson, J. N., Shum, M. S., & Uttal, D. H. (1998). Road climbing: Principles governing asymmetric route choices on maps. *Journal of Environmental Psychology*, *18*(3), 251–264.

Bailenson, J. N., Shum, M. S., & Uttal, D. H. (2000). The initial segment strategy: A heuristic for route selection. *Memory & Cognition*, *28*(2), 306–318.

Batty, M. (2009). Accessibility: In search of a unified theory. *Environment and Planning B: Planning and Design*, *36*(2), 191–194.
Berghauser Pont, M., & Marcus, L. (2014). Innovations in measuring density: From area and location density to accessible and perceived density. *NA*, *26*(2).
Bishop, C. M. (2006). Pattern recognition. *Machine Learning*.
Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679.
Boyle, M., & Kiel, K. (2001). A survey of house price hedonic studies of the impact of environmental externalities. *Journal of Real Estate Literature*, *9*(2), 117–144.
Brun, M., Sima, C., Hua, J., Lowey, J., Carroll, B., Suh, E., & Dougherty, E. R. (2007). Model-based evaluation of clustering validation measures. *Pattern Recognition*, *40*(3), 807–824.
Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, *2*(3), 199–219.
Cheng, J., & Bertolini, L. (2013). Measuring urban job accessibility with distance decay, competition and diversity. *Journal of Transport Geography*, *30*, 100–109.
Chuvieco, E. (1999). Measuring changes in landscape pattern from satellite images: Short-term effects of fire on spatial diversity. *International Journal of Remote Sensing*, *20*(12), 2331–2346.
Coupland, A. (1997). *Reclaiming the city: Mixed use development*. Taylor & Francis.
Crowe, D. A., Averbeck, B. B., Chafee, M. V., Anderson, J. H., & Georgopoulos, A. P. (2000). Mental maze solving. *Journal of Cognitive Neuroscience*, *12*(5), 813–827.
Dalton, N. (2001). Fractional configurational analysis and a solution to the Manhattan problem. *Proceedings of the 3rd Space Syntax Symposium, Atlanta*. University of Michigan.
Dalton, N. (2000). *Meanda, Computer Program*.
Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *2*, 224–227.
Dunn, J. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, *3*(3), 32–57.
Figueiredo, L. C. (2009). *Continuity lines: An investigation of urban form through street networks.* (PhD Thesis) University College London.
Fujita, M., Krugman, P. R., & Venables, A. (2001). *The spatial economy: Cities, regions, and international trade*. MIT Press.
Gehl, J. (2011). *Life between buildings: Using public space*. Island Press.
Geurs, K., Montis, A., & Reggiani, A. (2015). Recent advances and applications in accessibility modelling. *Computers, Environment and Urban Systems*, *49*, 82–85.
Geurs, K. T., Krizek, K. J., & Reggiani, A. (Eds.). (2012). *Accessibility analysis and transport planning: Challenges for Europe and North America*. Edward Elgar Publishing.
Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, *17*(2), 107–145.
Haq, S. (2003). Investigating the syntax line: Configurational properties and cognitive correlates. *Environment and Planning B*, *30*(6), 841–864.
Hasan, S., & Ukkusuri, S. V. (2014). Urban activity pattern classification using topic models from online geo-location data. *Transportation Research Part C: Emerging Technologies*, *44*, 363–381.
Hasan, S., Zhan, X., & Ukkusuri, S. V. (2013). Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing* (pp. 6). ACM.
Hess, P. M., Moudon, A. V., & Logsdon, M. G. (2001). Measuring land use patterns for transportation research. *Transportation Research Record*(1780), 17–24.
Hillier, B. (2007). *Space is the machine: A configurational theory of architecture*.
Hillier, B., & Hanson, J. (1984). *The social logic of space*. Cambridge University Press.
Hillier, B., & Iida, S. (2005). Network and psychological effects in urban movement. *Spatial information theory* (pp. 475–490). Berlin Heidelberg: Springer.
Hillier, B., & Penn, A. (2004). Rejoinder to Carlo Ratti. *Environment and Planning B-Planning & Design*, *31*(4), 501–511.
Hillier, B., Penn, A., Hanson, J., Grajewski, T., & Xu, J. (1993). Natural movement-or, configuration and attraction in urban pedestrian movement. *Environment and Planning B-Planning & Design*, *20*(1), 29–66.
Jacob, A. (1993). *Great streets*. USA: Massachusetts Institute of Technology.
Jacobs, J. (1961). *The death and life of Great American cities*. Random House Digital, Inc.
Jiang, B., & Liu, X. (2010). Automatic generation of the axial lines of urban environments to capture what we perceive. *International Journal of Geographical Information Science*, *24*(4), 545–558.
Jiang, S., Ferreira, J., & González, M. C. (2012). Clustering daily patterns of human activities in the city. *Data Mining and Knowledge Discovery*, *25*(3), 478–510.
Karamshuk, D., Noulas, A., Scellato, S., Nicosia, V., & Mascolo, C. (2013). Geo-spotting: Mining online location-based services for optimal retail store placement. *Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 793–801). ACM.
Karimi, K. (2012). A configurational approach to analytical urban design: 'Space syntax' methodology. *Urban Design International*, *17*(4), 297–318.
Kim, Y. O., & Penn, A. (2004). Linking the spatial syntax of cognitive maps to the spatial syntax of the environment. *Environment and Behavior*, *36*(4), 483–504.
Law, S., Karimi, K., Penn, A., & Chiaradia, A. (2013). Measuring the influence of spatial configuration of the housing market in metropolitan London. *Proceedings of the Ninth International Space Syntax Symposium, Seoul*.
Liu, X., & Jiang, B. (2012). Defining and generating axial lines from street center lines for better understanding of urban morphologies. *International Journal of Geographical Information Science*, *26*(8), 1521–1532.
Liu, X., & Long, Y. (2015). Automated identification and characterization of parcels with OpenStreetMap and points of interest. *Environment and Planning B: Planning and Design* (p.0265813515604767).
Liu, X., Kang, C., Gong, L., & Liu, Y. (2016). Incorporating spatial interaction patterns in classifying and understanding urban land use. *International Journal of Geographical Information Science*, *30*(2), 334–350.
Long, Y., & Thill, J. C. (2015). Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing. *Computers, Environment and Urban Systems*, *53*, 19–35.
Long, Y., Shen, Y., & Jin, X. (2015). Mapping block-level urban areas for all Chinese cities. *The Annals of the Association of American Geographers*, *106*(1), 96–113.
Marcus, L. (2010). Spatial capital. *The Journal of Space Syntax*, *1*(1), 30–40.
Montello, D. R. (1991). The measurement of cognitive distance: Methods and construct validity. *Journal of Environmental Psychology*, *11*(2), 101–122.
Newman, P., & Kenworthy, J. (1999). *Sustainability and cities: Overcoming automobile dependence*. Island Press.
Newman, P., & Kenworthy, J. (2011). Peak car use: Understanding the demise of automobile dependence. *World Transport Policy and Practice*, *17*(2), 31–42.
Ozbil, A., Peponis, J., & Bafna, S. (2009). The effects of street configuration on transit ridership. *7th International Space Syntax Symposium*. Stockholm: KTH School of Architecture and the Built Environment.
Ozbil, A., Peponis, J., & Stone, B. (2011). Understanding the link between street connectivity, land use and pedestrian flows. *Urban Design International*, *16*(2), 125–141.
Pendola, R., & Gen, S. (2008). Does "Main Street" promote sense of community? A comparison of San Francisco neighborhoods. *Environment and Behavior*, *40*(4), 545–574.
Penn, A., & Turner, A. (2004). Movement-generated land-use agglomeration: Simulation experiments on the drivers of fine-scale land-use patterning. *Urban Design International*, *9*(2), 81–96.
Peponis, J., Bafna, S., & Zhang, Z. (2008). The connectivity of streets: Reach and directional distance. *Environment and Planning B Planning & Design*, *35*(5), 881.
Ratti, C. (2004). Urban texture and space syntax: Some inconsistencies. *Environment and Planning B: Planning and Design*, *31*(4), 487–499.
Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., ... Strogatz, S. H. (2010). Redrawing the map of Great Britain from a network of human interactions. *PloS One*, *5*(12), e14248.
Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.
Scoppa, M. D., & Peponis, J. (2015). Distributed attraction: The effects of street network connectivity upon the distribution of retail frontage in the City of Buenos Aires. *Environment and Planning B: Planning and Design*, *42*(2), 354–378.
Shelton, T., Poorthuis, A., & Zook, M. (2015). Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning*, *142*, 198–211.
Shen, Y., & Karimi, K. (2015). Understanding the roles of urban configuration on spatial heterogeneity and submarket regionalisation of house price pattern in a mix-scale hedonic model. *Proceedings of the 10th Space Syntax Symposium*. London: University College London.
Shen, Y., Karimi, K., & Xia, Q. (2013). Morphological transformation of historical centres in Tianjin. *Proceedings of the 9th Space Syntax Symposium*. Seoul: Sejong University.
Ståhle, A., Marcus, L., & Karlström, A. (2005). Place syntax: Geographic accessibility with axial lines in GIS. *Fifth international space syntax symposium* (pp. 131–144). Techne Press.
Turner, A. (2001a). Angular analysis. *Proceedings of the 3rd Space Syntax Symposium*. Atlanta: University of Michigan.
Turner, A. (2001b). A program to perform visibility graph analysis. *Proceedings of the 3rd Space Syntax Symposium*. Atlanta: University of Michigan.
Varoudis, T. (2012). DepthmapX MultiPlatform Spatial Network Analysis Software, Version 0.30 OpenSource. http://varoudis.github.io/depthmapX/
Vaughan, L. (2007). The spatial syntax of urban segregation. *Progress in Planning*, *67*(3), 205–294.
Williams, B., Walsh, C., & Boyle, I. (2010). The development of the functional urban region of Dublin: Implications for regional development markets and planning. *Journal of Irish Urban Studies*, *7*, 9.
Wu, L., Zhi, Y., Sui, Z., & Liu, Y. (2014). Intra-urban human mobility and activity transition: Evidence from social media check-in data. *PLoS ONE*, *9*(5).
Wyly, E. (2014). The new quantitative revolution. *Dialogues in Human Geography*, *4*(1), 26–38.
Xiao, Y. (2012). *Urban morphology and housing market.* (PhD thesis) University of Cardiff.
Xie, X. L., & Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *8*, 841–847.
Yuan, J., Zheng, Y., & Xie, X. (2012). Discovering regions of different functions in a city using human mobility and POIs. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 186–194). ACM.
Zhong, C., Schläpfer, M., Arisona, S. M., Batty, M., Ratti, C., & Schmitt, G. (2015). Revealing centrality in the spatial structure of cities from human activity patterns. *Urban Studies* (p.0042098015601599).