*Article*

# Delineating functional regions from road networks: The case of South Korea

## Yifu Ou ⓘ
Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, China

## Euijune Kim ⓘ
Department of Agricultural Economics and Rural Development, Seoul National University, Seoul, Korea

## Xingjian Liu ⓘ and Kyung-Min Nam ⓘ
Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, China

Centre for Urban Studies and Urban Planning, Faculty of Architecture, The University of Hong Kong, Pokfulam, Hong Kong SAR, China

## Abstract
We apply a percolation method to South Korea's road networks and identify major urban clusters. The results show that Korea has developed a highly mono-centric spatial structure, in which the Seoul Metropolitan Area expands to northern Chungnam, far beyond the limits conventionally assumed. A percolation threshold of 1080 m is critical in delineating Korea's functional regions, and the results at this threshold show great similarity to flow-based maps. Similar model outputs support the percolation approach as an alternative to conventional flow-based methods. The similarity is higher for larger clusters, where the supply of road infrastructure better meets demand thanks to size-biased public resource allocation. Also, the similarity between model outputs tends to increase with a time lag, since road network configurations take time to catch up with the demand revealed in traffic flows. These findings speak for the need for increased connectivity in Korea's smaller cities.

## Keywords
percolation, functional regions, road networks, urban hierarchies, spatial boundary delineation, network analysis

**Corresponding author:**
Kyung-Min Nam, Department of Urban Planning and Design, The University of Hong Kong, 8/F, Knowles Building, The University of Hong Kong, Pokfulam Road, Hong Kong, China.
Email: kmnam@hku.hk

## JEL codes

O18, O20, R12, R42, R58

## Introduction

Urban and regional analysis is often conducted on administrative boundaries, but administrative regions may not effectively reflect existing functional complementarity (Dawkins, 2003; Dijkstra et al., 2021; Harrison and Hoyler, 2015). Identifying functional regions matters in dealing with the gap between economic and regulatory boundaries and thus in improving administrative efficiency (Duranton, 2021). However, delineating functional boundaries is not straightforward, since they are dynamic and evolve over time (Brenner, 2003; Storper, 2014).

A conventional approach to estimating the spatial boundaries of functional regions relies on statistical or geo-computational methods using commuting or trade flow data (e.g., Bosker et al., 2021; Brown and Hincks, 2008; Masser and Scheurwater, 1980; Mitchell and Watts, 2010). Although commuting and trade flows convey a good deal of information on cross-city linkage and interdependency, the limited availability of such flow data reduces the applicability of the flow-based approach (Dingel et al., 2021). Dependence on administrative boundaries is another issue, since flow data itself is collected based on political jurisdictions. Finally, flow-based estimation often presumes the total number of functional regions or their central cities before analysis. With these presumptions positing gradualism, however, we may be inclined to neglect those changes in spatial structure that go beyond preexisting configurations, such as new-town development and the evolution of city regions (e.g., Yeh and Chen, 2020).

Fine-resolution raster data, such as nightlight satellite images or population grids, is often used as an alternative to conventional flow data (e.g., Cao et al., 2020; Ch et al., 2021; Dingel et al., 2021; Duranton, 2021; Henderson et al., 2012; Nam and Reilly, 2013). These data sets are widely available for the entire globe, and a morphology-based urban delineation using them reduces the need for strong presumptions regarding clustering patterns. Despite its merit, the sole focus of a morphological approach on physical forms weakens the rationale to treat the identified clusters as functional regions built on connectivity. Geospatial big data sources—such as those on points of interest, taxi/transit services, social media usage, smart-card transactions, and cellular-phone signals—have also been increasingly tested as an alternative to conventional survey-based flow data (e.g., Cai et al., 2017; Houssou et al., 2019; Niu and Silva, 2021; Ratti et al., 2010; Sulis et al., 2018). However, the applicability of such data is bounded by data availability in both spatial and temporal dimensions, and approximating spatial interdependency from big data may also be questioned as to its validity and robustness.

In sum, the limitations of existing methods demand an alternative approach. In recognition of this need, we apply a percolation approach (Arcaute et al., 2016) to estimating functional regions in South Korea based on road networks. On the methodological front, South Korea is chosen as a case example due to the availability of multiple reference maps for comparison. The Korean government has made various attempts to change the existing mono-centric spatial structure in power and resource allocation, which is biased toward the Capital Region, and has widely adopted the concept of functional regions to establish a model for competitive local regions (Government of the Republic of Korea, 2001; 2019). Most of the estimated Korean functional regions are based on flow data, and the comparison of our results with these regions is crucial in evaluating the performance of the percolation model. In addition, the Korean context, where various regional policies have been tested at multiple levels, can also help us demonstrate a set of policy

implications that can be drawn from reasonably delineated functional boundaries. Finally, our study of the Korean case can also enrich the empirical literature, as existing percolation-based urban analysis leans toward the cases of the West and China (Arcaute et al., 2016; Huynh et al., 2018; Long et al., 2016).

## Empirical approaches to delineating functional regions

Conventional flow datasets used to identify functional regions include commuting (Brown and Holmes, 1971; Duranton, 2015, 2021; Farmer and Fotheringham, 2011; Masser and Scheurwater, 1980; Mitchell and Watts, 2010), migration (Brown and Hincks, 2008; Masser and Scheurwater, 1980), and travel flows (Kwon and Seo, 2018). These datasets are often built on large scale sample surveys (Mu and Yeh, 2020), and flow-based approaches using them are subject to a few key limitations in application. One limitation is high data requirements and limited data availability. Reliable inter-city flow data and other complementary sub-national datasets do not exist in many parts of the developing world (Dingel et al., 2021). The dependence of estimated results on administrative boundaries also limits the applicability of the flow-based approach. Flow data itself is collected on political jurisdictions and is often released to the public after some degrees of aggregation (Mitchell and Watts, 2010). The unit of analysis is thus constrained by the units used for the original surveys or aggregated data, which significantly affects estimation results. Finally, too strong a condition presumed for flow-based estimation restricts the approach. Conventionally, flow-based methods presume the total number of functional regions or their central cities *before analysis*, since the absence of the presumption would make flow analysis for all possible origin-destination pairs an extremely computation-intensive task. This presumption of positing gradualism in spatial structure—that is, presuming that the number of functional regions is stable over time—may incur a substantial cost in estimated outcomes by overlooking the possibility of more radical changes in urban systems.

In recognition of these limitations, more recent studies have tested alternative geospatial big data sources, such as point-of-interest coordinates (Gao et al., 2017; Niu and Silva, 2021), taxi/transit service usage (Houssou et al., 2019; Wu et al., 2019), socio-media check-in (Cai et al., 2017; Jia et al., 2019; Poorthuis, 2017), smart-card transactions (Sulis et al., 2018; Zhong et al., 2014), and cellular-phone signals (Chen et al., 2020; Ratti et al., 2010). These alternatives may help reduce the limitations inherent in conventional survey-based flow data. However, they at best offer only approximate information on cross-regional interdependency, since they are not collected for that particular purpose. Also, the compiled data of this sort may not be available for the place and time needed for a particular study. Even when the data of interest is available, access to it is often strictly controlled due to privacy concerns (Smith et al., 2012).

Many studies taking a morphological approach use fine-resolution raster data, such as nightlight satellite images and population grids (e.g., Cao et al., 2020; Ch et al., 2021; Duranton, 2021; Fan et al., 2014; Henderson et al., 2012; Li and Zhou, 2017; Liu et al., 2019; Zhou et al., 2015). These datasets, which cover the entire globe for relatively recent years, have merit in terms of their wide availability, and they help researchers avoid subjective presumptions on the number or spatial limits of major urban clusters. A raster-based morphological approach, however, primarily focuses on physical clustering patterns while largely neglecting key drivers underlying them (Batty, 2007; Nam, 2017). Simply being located in proximity may not necessarily represent functional complementarity or interdependency among multiple spatial units, and thus treating the identified clusters as functional regions lacks a rationale. In addition, this approach requires a subjective choice of threshold light-intensity or population-density values to define urban areas, and the chosen threshold can have substantial impacts on analysis results (Dijkstra et al., 2021; Henderson et al., 2021; Nam and Reilly, 2013). Nightlight intensity as a proxy for urbanization also has limited

applicability in less developed countries or in smaller cities, as it is often too weak to allow re-searchers to distinguish urban areas (Dingel et al., 2021; Harari, 2020; Song et al., 2020). The compromised raster-data quality of the developing-country portion may be another issue. Pop-ulation grids are often estimated from census data, and thus the quality of the former is constrained by that of the latter (Doxsey-Whitfield et al., 2015; Li and Nam, 2017).

In this context, a percolation approach based on road network or comparable stock data may be a more practical alternative. Arcaute et al. (2016) is a pioneering study that applies percolation logic in physics to identifying urban regions using British road network data. The primary rationale for a percolation analysis is that the density of existing road networks reflects the intensity of activity levels, such as commuting or trade flows, although a time lag may exist between them (Arcaute et al., 2016). In this sense, substituting the former for the latter in identifying functional boundaries secures some ground. Furthermore, road network data is not restricted to geographical units or scales, and is more widely available with easy access options through globe-wide collaborative open projects, such as OpenStreetMap (e.g., Boeing, 2020).[1] Accordingly, the percolation method has been since applied to other regions, such as China and the U.S. (Long, 2016; Huynh et al., 2018), as well as to an international comparison (Cao et al., 2020).

Despite the merits of the percolation approach, it is not clear how well the percolation-based delineation of functional regions approximates flow-based outputs. Can a stock-based approach produce results reasonably close to those based on a flow-based approach? So far, few studies have offered reasonable answers to these questions; this motivates our study.

## Methodology

### Percolation analysis of road networks

The primary data used for our percolation analysis is Korea's national road network data, whose comparables for other countries are widely available through national official channels or open sources. In addition to such low data requirements, the percolation method has merit in that it does not require a presumption about the number of clusters or their spatial limits.

The first step of our analysis is to create an undirected graph of Korea's road networks. The graph is based on point-and-line data, which marks the location of highway junctions and intersections (nodes) and offers information on connectivity between nodes (links). A key bridging step between
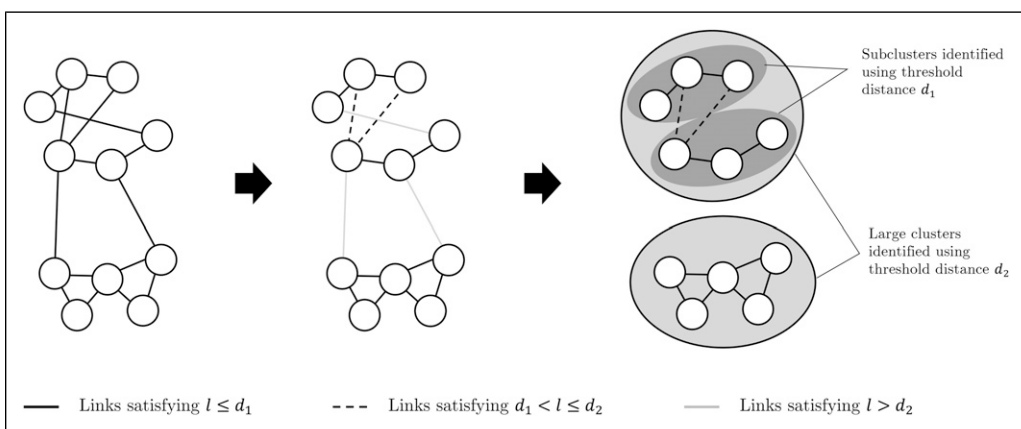


**Figure 1.** Simplified illustration of the percolation method. Note: refers to the length of the link. Source: Created by the authors.

the original data and the graph is to construct an adjacency matrix that identifies each pair of neighboring nodes and determines the length of the links defined by each pair. Once the undirected graph is prepared, a thresholding process is applied to it. As is done in Long (2016) and Masucci et al. (2015), the percolation algorithm is illustrated in a toy example (Figure 1). The procedure begins with a reasonably large threshold distance ($d_1$) and compares it with the length of each link ($l$). Those links satisfying $l \leq d_1$ are considered connected, while those failing to meet the condition are treated as disconnected and are deleted from the graph. This modified graph displays a clustered pattern of nodes with the number of clusters (i.e., a group of connected nodes) and the size of each cluster (i.e., the number of nodes included in each cluster). This threshold distance–imposing process is iterated with a set of smaller thresholds ($d_2, d_3, \cdots, d_n$) at certain intervals to see how the clustering patterns and their hierarchies evolve over the changed conditions. By design, the number of clusters tends to increase with shorter threshold distance. Clusters identified as having a smaller distance threshold (e.g., $d_1$) form a subset of those identified as having a larger threshold (e.g., $d_2$); thus, changes in clustering patterns over a set of distance thresholds convey information on the national urban system.

This iterative percolation process continues until it identifies a reasonable number of functional regions. There is no fixed rule regarding the choice of the appropriate threshold distance, which may differ by country and data resolution. However, an overly large threshold distance may identify entire nodes as a single cluster; conversely, too small a threshold distance can detect only a limited set of central cities.

The estimated clustered nodes are then projected onto a map of Korea's administrative boundaries to delineate functional regions (see Figure S1 for Korea's administrative boundaries). We use city and county-level boundaries for the mapping, since city-level municipalities in Korea (i.e., city, county, and district) are the lowest level administrative units having publicly elected leaders and forming the basis of local autonomy (Park et al., 2019).

## Model's performance assessment

To assess our model's performance, we first pick three reference maps and compare our estimated results with them (see Section "identified functional urban regions" for further details). There is no consensus on Korea's regional boundaries in a functional sense, since they may differ depending on the data and method used for analysis, not to mention their dynamic nature. However, the three maps chosen are built on a conventional flow-based approach using similar functional-region concepts (e.g., labor markets) and are widely referred to within academic and policy circles when delineating functional regions. For the comparison of our results with the reference maps, we use the Jaccard index, which is commonly used to measure the degree of similarity in a pair of model outputs (Farmer and Fotheringham, 2011; Liu et al., 2019; Rybnikova et al., 2021). The Jaccard index $J(A,B)$ for two regions $A$ and $B$ measures the share of overlapping portion of the two regions out of their pooled area ($= \frac{|A \cap B|}{|A \cup B|}$). It returns a value ranging in [0, 1], and a higher score means a greater degree of similarity.

A second set of tests compares central cities in each region identified by population density and by road network closeness centrality, and investigates the correlation between road network–based closeness centrality and socioeconomic indicators (e.g., population density, local income tax, and $PM_{2.5}$). Two road network closeness centrality indices are tested. One is $C_i$, a standardized measure of the "global" closeness centrality of node $i$, as adopted in Porta et al. (2009) and Wang et al. (2011). The other is a "local" closeness centrality index, $C_k$, which measures the averages of $C_i$ within a given administrative unit $k$.

## Data

The main data used for our percolation analysis is Korea's official road network shapefiles constructed for 2018 and distributed by the National Geographic Information Institute (2021). The original dataset is comprised of >10 million nodes and links with rich attribute information, including road width/length and pavement condition. Of the nodes and links, we ignore minor road sections with widths ≤4 m, which are not appropriate for vehicular traffic, and consider a total of 634,426 nodes and 770,182 links for analysis.

Other complementary data sets are from various sources. Socioeconomic variables, such as city and county-level population and local income tax, are constructed from the Korean Statistical Information Service (KOSIS) database (Statistics Korea, 2022). Population-weighted, annual mean $PM_{2.5}$ levels in 2019 are computed using the $36'' \times 36''$ $PM_{2.5}$ concentration grids developed by Hammer et al. (2022) and the $60'' \times 60''$ LandScan population grids distributed by the Oak Ridge National Laboratory (2022).
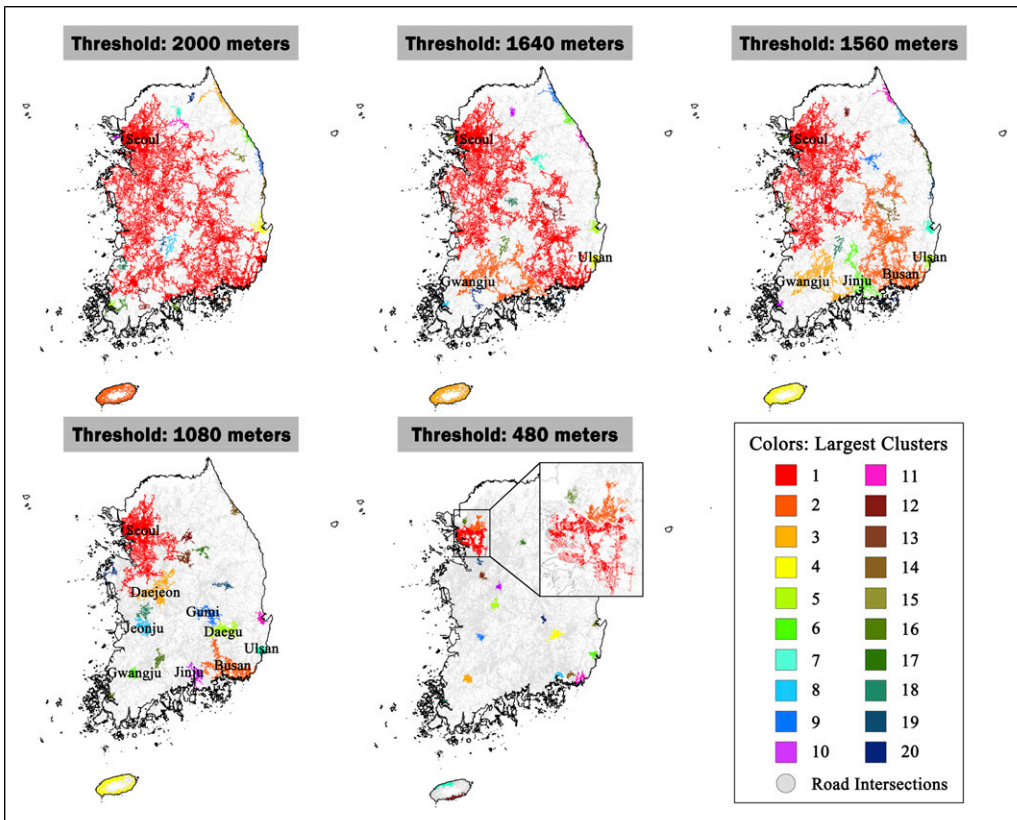


**Figure 2.** Largest clusters of Korea's road network intersections identified at different percolation thresholds. Source: Created by the authors.
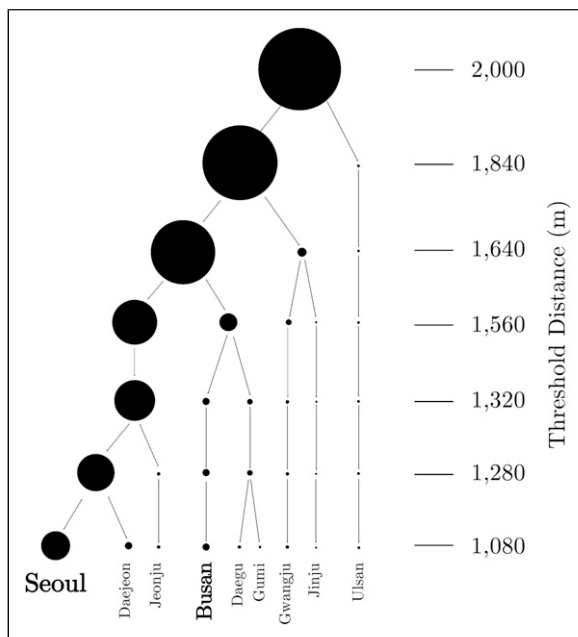
**Figure 3.** Korea's major urban clusters and their hierarchies. Note: Jeju Island is excluded from analysis. Clusters represented as circles are labeled after their central cities. The size of each circle is proportional to the number of road junctions included in the corresponding cluster. Source: Created by the authors.

## Results

### Korea's urban hierarchies

Following the approach used by Arcaute et al. (2016), we highlight key sets of intersection clusters identified at different threshold distances and their corresponding urban geographical patterns (Figures 2 and 3). Overall, our analysis results clearly demonstrate Korea's mono-centric urban hierarchy, in which the Seoul Metropolitan Area (SMA) occupies a dominant position in the national urban system. When a percolation process begins from an initial threshold distance of 2000 m with a 40-m decremental interval, only a single giant cluster is detected within the threshold range of ≥1880 m. Seoul, showing the highest centrality index, may be viewed as the central city of this cluster, which contains >85% of Korea's entire road intersections analyzed. The northeast of Gangwon and Gyeongbuk Provinces is largely excluded from the cluster, revealing their weak network connectivity arising from unfavorable topography. The Taebaek mountain range is vertically stretched across the east of the Korean peninsular along Gangwon and Gyeongbuk Provinces, and this geographical condition functions as a barrier to linking to Korea's main road networks.

The Seoul-centered single cluster begins to branch off when Ulsan and Gwangju—both of which are province-level metropolitan cities—are separated at the thresholds of 1840 m and 1640 m, respectively. Then, five major urban clusters emerge at the threshold distance of 1560 m: Busan is split from the Seoul-centered main cluster, Jinju is disjointed from the Gwangju cluster, and Ulsan remains independent. This level of connectivity is of particular importance in Korea's urban hierarchy in that Busan—Korea's second largest city and most important container seaport—is identified as a separate cluster. The Seoul–Busan axis was a primary economic corridor for Korea's initial growth pole development strategy, and Busan has served as a center of the Southeastern

Maritime Industrial Region—one of the two mega-city regions essential for Korea's export-oriented industrialization (Kim et al., 2014; Nam, 2009).

The identification of Korea's most densely urbanized regions is evident at the threshold of 1080 m. At this level of connectivity, at least 15 major clusters are identified in South Korea's main territory (excluding Jeju Island), and the gap in size between the SMA and the other clusters is still large. Six out of Korea's eight province-level cities (and their vicinity), representing the country's most densely urbanized areas, are recognized as independent clusters, except for Incheon merged to Greater Seoul and Sejong merged to Greater Daejoen. One point to be noted is that the estimated boundaries of the Greater Seoul area are stretched further to the north of Chungnam Province. This may highlight the mismatch between the functional and administrative boundaries in Korea's context, given that the SMA is commonly defined as the aggregation of Seoul, Incheon, and Gyeoggi, complying with their political jurisdictions while excluding Chungnam. However, our results demonstrate that the SMA and Chungnam are well connected and may form an even larger Greater Seoul region from a functionality perspective. In fact, Daejeon and its vicinity, located in the south of Chungnam, are part of the Greater Seoul area at thresholds of ≥1120 m, forming a much greater Seoul-centered urban cluster than conventionally assumed.

Finally, sub-regional urban areas are found at the threshold of 480 m. Only Seoul and its close vicinity form a sizeable cluster at this level of connectivity, while a few other local cities are identified as quasi-isolated points, rather than as a cluster. This finding suggests that thresholds ≤480 m in Korea's context are too restrictive to consider functional complementarities at a regional level. Thresholds ≤480 m even bisect the city of Seoul into north (Gangbuk) and south (Gangnam) across the Han River, and may be more suited to examine connectivity at district or lower levels. As a comparison, Long (2016) identifies individual cities in China with a threshold of 500 m.

## Identified functional urban regions

To assess the model's performance relative to conventional flow-based approaches, we first choose three reference maps of Korea's functional regions estimated from relatively recent flow data (Kim et al., 2012; Kim and Woo, 2019; Lee and Lee, 2015) and then compare our results with them (Figure 4). Out of multiple percolation threshold choices available, we choose the 1080 m case for comparison, since the major clusters identified at this threshold include all of Korea's most densely urbanized regions. In fact, the mean Jaccard index for Korea's five major functional regions identified peaks at the threshold of 1080 m (0.68), presenting the highest overall similarity at this level of connectivity (Table 1). One caveat is that the three chosen reference maps themselves present substantially different estimation results, arising from dissimilarities in data sources, methods, and time periods of interest. It is hard to tell which one of them is superior to the others in the absence of official functional-region boundaries in Korea's context. Thus, our assessment focuses on the similarity between our results and conventional flow-based approaches as a whole.

Overall, the network-based functional region classification shows a fairly high similarity to flow-based estimates (Table 2). For all major clusters identified, the mean Jaccard index comparing our results with each reference map ranges in [0.53, 0.66]. This means that, on average, 53%–66% of the pooled area is overlapping between the two different sources compared. The similarity is greater when our results are compared with reference maps using older flow data. The Jaccard index against Kim et al. (2012) using the 2007 linked-trip data is 0.66, but the coefficient against Lee and Lee (2015) using the 2010 linked-trip data and against Kim and Woo (2019) using the 2015 linked-trip data decreases to 0.55 and 0.53, respectively. A time lag in transport infrastructure investment seems to underlie this tendency. Existing road infrastructure can be better aligned to past flow patterns than current ones, as it was built in accordance with the past or then-forecasted demand.
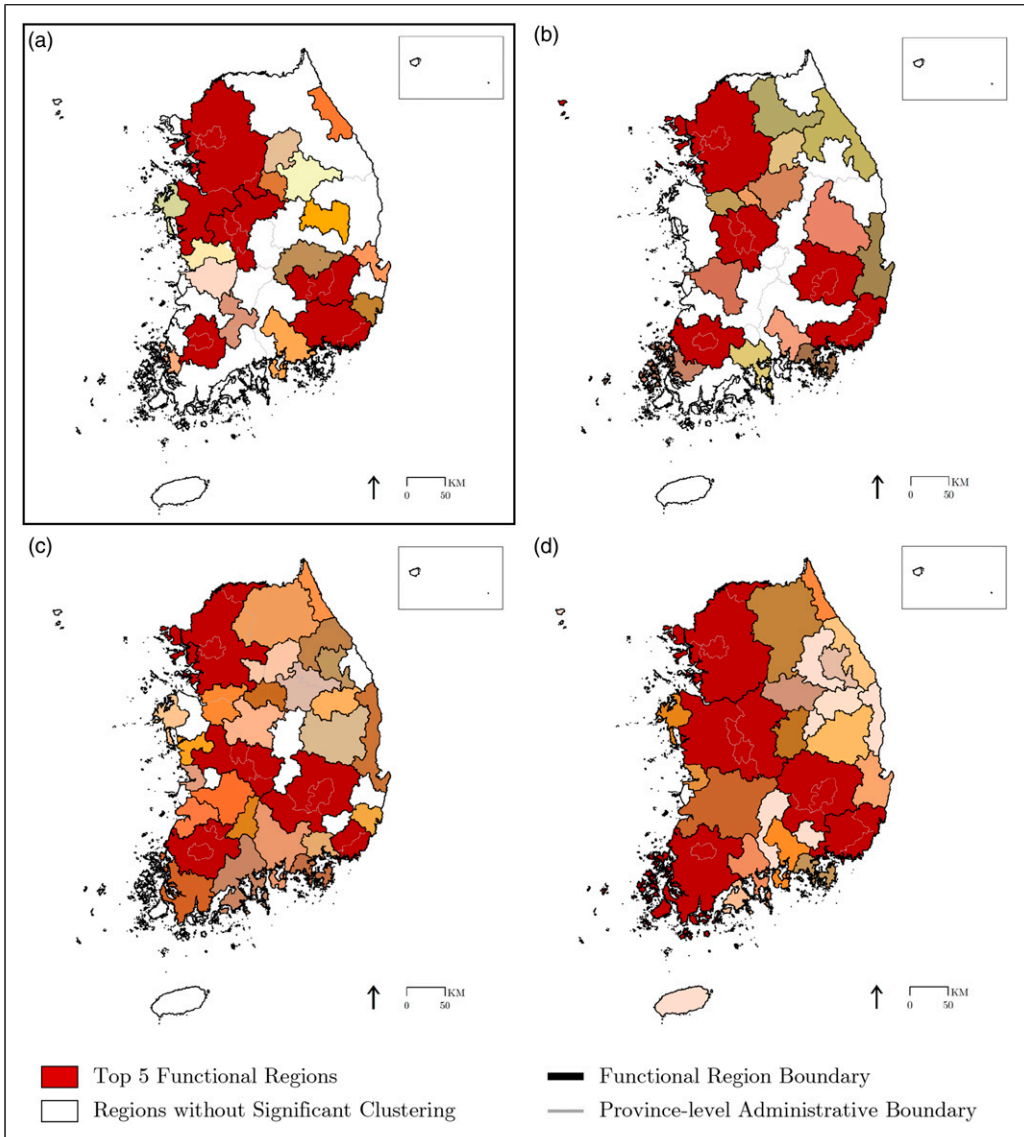
**Figure 4.** Functional regions defined at 1080 m threshold, compared with reference cases: (a) Our results; (b) Reference map from Kim et al. (2012); (c) Reference map from Lee and Lee (2015); (d) Reference map from Kim and Woo (2019). Note: Identified functional regions are highlighted in non-white colors. Source: Created by the authors.

When focusing on upper-tier cities, the mean Jaccard index is much higher, presenting a range of [0.69, 0.81] for the top three urban clusters. These numbers even exceed the mean Jaccard index from a pairwise comparison of the three reference maps themselves, ranging in [0.64, 0.77] (Table 3). This finding suggests that road network density in large urban clusters well represents inter-city travel patterns, and thus a percolation-based approach in this setting can be a particularly suitable alternative to flow-based delineation. Among the top three clusters, the Greater Seoul region shows the highest mean Jaccard index, ranging in [0.73, 0.87], followed by Daegu with a range of

**Table 1.** Changes in mean Jaccard index over percolation distance thresholds.

| Threshold Distance (m) | Central Cities of the Five Major Clusters | | | | | |
| | Seoul | Busan | Daegu | Daejeon | Gwangju | Mean |
|---|---|---|---|---|---|---|
| 1120 | 0.71 | 0.72 | 0.73 | n/a | 0.62 | n/a |
| 1080* | 0.82 | 0.72 | **0.73** | **0.49** | **0.62** | **0.68** |
| 1040 | 0.86 | 0.69 | **0.73** | **0.49** | 0.55 | 0.66 |
| 1000 | 0.84 | **0.77** | 0.67 | 0.42 | 0.55 | 0.65 |
| 960 | 0.85 | 0.76 | 0.62 | 0.42 | 0.48 | 0.62 |
| 920 | 0.85 | 0.76 | 0.62 | 0.42 | 0.48 | 0.62 |
| 880 | 0.85 | 0.76 | 0.62 | 0.42 | 0.48 | 0.62 |
| 840 | 0.85 | 0.67 | 0.62 | 0.42 | 0.48 | 0.61 |
| 800 | 0.87 | 0.67 | 0.62 | 0.42 | 0.48 | 0.61 |
| 760 | 0.87 | 0.67 | 0.62 | 0.45 | 0.48 | 0.62 |
| 720 | 0.88 | 0.63 | 0.62 | 0.39 | 0.41 | 0.59 |
| 680 | **0.89** | 0.63 | 0.56 | 0.39 | 0.41 | 0.58 |
| 640 | 0.87 | 0.63 | 0.56 | 0.39 | 0.41 | 0.57 |
| 600 | 0.82 | 0.59 | 0.50 | 0.37 | 0.41 | 0.54 |
| 560 | 0.80 | 0.59 | 0.50 | 0.37 | 0.34 | 0.52 |
| 520 | 0.80 | 0.59 | 0.45 | 0.37 | 0.34 | 0.51 |

Note: All figures displayed in this table take the mean of the Jaccard coefficients for each region between our results and each of the three reference maps. The top five clusters are not defined at the thresholds of ≥120 m, since the Daejeon cluster is merged into the Seoul cluster. The maximum Jaccard index values for each region are in bold.

**Table 2.** Jaccard coefficients between our estimates and three reference maps.

| Region | Kim et al. (2012) | Lee and Lee (2015) | Kim and Woo (2019) |
|---|---|---|---|
| Five Major Urban Clusters | | | |
| Seoul | 0.87 | 0.73 | 0.87 |
| Busan | 0.75 | 0.67 | 0.75 |
| Daegu | 0.81 | 0.68 | 0.68 |
| *Mean (Top 3 Clusters)* | *0.81* | *0.69* | *0.77* |
| Daejeon | 0.63 | 0.33 | 0.50 |
| Gwangju | 0.75 | 0.69 | 0.41 |
| *Mean (Top 5 Clusters)* | *0.76* | *0.62* | *0.64* |
| Other Regions | | | |
| Ulsan | (N/A) | 1.00 | (N/A) |
| Jeonju | 0.86 | 0.63 | 0.33 |
| Gumi | (N/A) | (N/A) | (N/A) |
| Jinju | 0.60 | 0.50 | 0.80 |
| Pohang | 0.50 | 0.50 | 0.67 |
| Wonju | 1.00 | 0.67 | 0.29 |
| Chungju | 0.25 | 0.50 | 0.33 |
| Gangneung | 0.40 | 0.25 | 0.25 |
| Mokpo | 0.50 | 0.33 | (N/A) |
| Namwon | (N/A) | 0.25 | (N/A) |
| *Mean (Excluding Top 5 Clusters)* | *0.59* | *0.51* | *0.44* |
| **Mean (All Clusters Detected)** | **0.66** | **0.55** | **0.53** |

Note: A threshold distance of 1080 m is applied to percolation-based urban clustering.

**Table 3.** Cross-referenced Jaccard coefficient matrix for Korea's five major urban clusters.

| Region | This Study vs. | | | Kim et al. (2012) vs. | | Lee and Lee (2015) vs. |
|---|---|---|---|---|---|---|
| | Kim et al. (2012) | Lee and Lee (2015) | Kim and Woo (2019) | Lee and Lee (2015) | Kim and Woo (2019) | Kim and Woo (2019) |
| Seoul | 0.87 | 0.73 | 0.87 | 0.62 | 0.93 | 0.62 |
| Busan | 0.75 | 0.67 | 0.75 | 0.84 | 0.84 | 0.90 |
| Daegu | 0.81 | 0.68 | 0.68 | 0.53 | 0.53 | 0.40 |
| **Mean (Top 3)** | **0.81** | **0.69** | **0.77** | **0.66** | **0.77** | **0.64** |
| Daejeon | 0.63 | 0.33 | 0.50 | 0.92 | 0.55 | 0.59 |
| Gwangju | 0.75 | 0.69 | 0.41 | 0.74 | 0.77 | 0.67 |
| **Mean (Top 5)** | **0.76** | **0.62** | **0.64** | **0.74** | **0.77** | **0.67** |

[0.68, 0.81] and Busan with a range of [0.67, 0.75]. Comparable coefficients for Gwangju and Daejeon on the lower top-five cluster list decrease substantially, ranging in [0.41, 0.75] and [0.33, 0.63], respectively. Lower index values for Gwangju and Daejeon—which are younger province-level cities promoted in the late 1980s—may be attributable to their relatively low positions in Korea's urban hierarchy compared with Seoul, Busan, and Daegu—which became province-level cities in 1949, 1963, and 1981, respectively. Given Korea's size-biased pecking order in public investment, a lower position in national urban hierarchies results in a larger gap between demand (reflected in flows) and supply (reflected in networks) in road infrastructure.

Lower Jaccard coefficients for even lower-tier urban clusters seem to support our reasoning. When the five largest clusters are excluded, the mean Jaccard index for the remaining cities drops to [0.44, 0.59]. The decreasing similarity reflects that the gap between supply of road infrastructure and actual demand for it is greater in lower-tier cities. As discussed earlier, lower-tier cities have a disadvantage in developing dense road networks under a size-biased national resource allocation system, and suffer from an undersupply of road infrastructure and weaker connectivity.

## Road network centrality

The central cities in each functional region, displayed in Figure 5(a), may be identified from either population density or closeness centrality indices computed from the road network data. Those with high population density include all higher tier cities (province-level cities or provincial capitals), which are generally perceived by the public as central cities in each region (Figure 5). Overall, those with high closeness centrality indices overlap with those with high population density, suggesting that road network density can produce reasonably similar results in identifying central cities. The single exception is Busan, which is located in the southeast corner of its hinterland. Gimhae, a regional sub-center closer to the geographical center of the region, reveals higher centrality than Busan. One reason is that, by design, closeness centrality indices built on the shortest distance concept present a bias toward a geographical center (Mitchell and Griffin, 2021). Busan's unique geographical condition—a port city embracing 41 islands where around half the land is mountainous and not suitable for development—is another factor that has pushed the center of the regional road network away from its city proper.
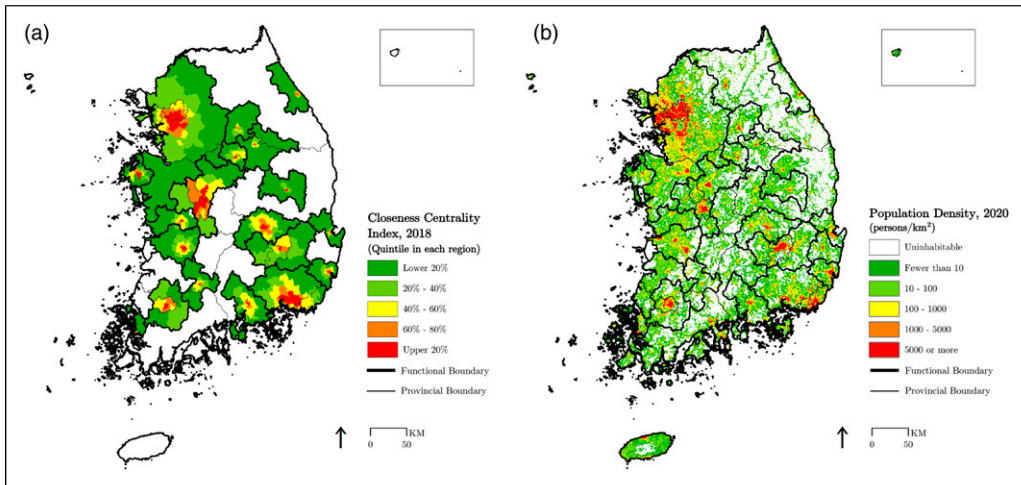
**Figure 5.** Regional centers based on road network and population density: (a) Distribution of closeness centrality, 2018; (b) Distribution of population density, 2020. Source: Created by the authors based on data from the National Geographic Information Institute (2021) and the Oak Ridge National Laboratory (2022).

**Table 4.** Correlation coefficients between centrality measures and socioeconomic attributes.

| Centrality | Population Density | Income Tax | $PM_{2.5}$ |
|---|---|---|---|
| City Level ($C_k$) | 0.72** | 0.32** | 0.79** |
| Point Level ($C_i$) | 0.50** | n/a | 0.69** |

Note: *$p < 0.05$; **$p < 0.01$.

The predictive power of closeness centrality for population density may be extended to other socioeconomic indicators, including those listed in Table S1 in the Supplementary Material. Overall, closeness centrality and the three socioeconomic variables tested (population density, local income tax, and $PM_{2.5}$) show positive correlations ranging in [0.32, 0.79] (Table 4). For example, correlation coefficients for population density are 0.72 at the city level (vs. $C_k$) and 0.50 at the point level (vs. $C_i$) at the 1% level. These coefficients present either strong (>0.67) or moderate (0.36–0.67) correlation (Taylor, 1990). The correlation coefficient for local income tax (vs. $C_k$)—a proxy for local income levels and welfare conditions—is 0.32 and significant at the 1% level. This coefficient, located near the lower bound of the "moderate" range, is substantially smaller than the population density case, suggesting that economic conditions involve a more complex set of contributors. Finally, of the three variables tested, closeness centrality is most strongly correlated with $PM_{2.5}$ concentrations—a proxy for total energy consumption and activity levels. The coefficients are 0.69 at the city levels and 0.79 at the point level, both of which fall into the "strong" correlation class, and are significant at the 1% level.

## Further discussion

As discussed in earlier sections, our analysis of percolation in Korea's road networks demonstrates that this approach is useful in understanding national urban hierarchies and well approximates the delineation of flow-based functional regions. In particular, a percolation approach has strength in its

low data requirement and wide applicability as well as in its independence from subjective presumptions regarding regional boundary limits.

However, there are several caveats in the application of the approach. First, natural conditions can distort percolation-based delineation. For example, a single functional region may be detected as multiple isolated sub-regions under a reasonable percolation threshold, if rivers or mountain ranges passing through increase the physical distance among road junctions within the region. Second, inconsistent data quality can undermine the robustness of percolation application results (e.g., Boeing, 2020). It is not rare for open GIS data to be inconsistent in quality even within a single country, let alone across countries—detailed road segments are represented in only a few large cities while they are omitted or aggregated elsewhere. Finally, strong government interventions in public infrastructure investment may increase the gap between actual demand (e.g., traffic volume) and supply (e.g., roads) and thus weaken a key assumption of the percolation approach: that the latter reflects the former with a certain time lag. Then, this can be a source of divergence between flow- and percolation-based delineation results.

## Conclusions

In this study, we apply a percolation approach to delineating functional regions and assess its predictive power relative to conventional flow-based approaches, using South Korea as a case example. Our results show that the percolation concept offers a useful tool for understanding urban hierarchies, and its application to widely available network stock data can well approximate the flow-delineated functional regions subject to high data requirements. The rationale for potential substitution of such stock data for flow data is the former's predictive power for the latter—the configuration of existing networks reflects the actual demand revealed in flow data.

A percolation analysis of the Korean road networks demonstrates that South Korea has developed an extremely Seoul-centered mono-centric structure, and the spatial coverage of the SMA is much greater than conventionally assumed. At a reasonable percolation threshold defining Korea's major urban clusters, the SMA—commonly defined as an aggregate of Seoul, Incheon and Gyeonggi—is stretched to the north of Chungnam Province, going beyond the southern limit set by Gyeonggi Province. The expansion of the SMA's functional boundaries has significant implications for Korea's regional policy and growth control over the Capital Region, but it can be overlooked by conventional flow-based methods. For example, all three flow-based reference maps chosen for comparison limit the maximum functional boundaries of the SMA to Gyeonggi Province borders *before* their analysis, positing the conventional view. Presumptions of this kind lack a rationale and can lead to an underestimation of the actual change in urban systems.

Overall, our percolation analysis results well approximate existing flow-based estimates, and their similarity tends to increase for larger clusters with some time lag. The mean Jaccard index for Korea's 15 major urban clusters identified from percolation ranges in [0.53, 0.66], and comparable index values increase to [0.69, 0.81] for Korea's top three clusters. In particular, the latter even exceed the coefficients computed from a pairwise comparison of the three flow-based reference maps themselves—[0.64, 0.77] for the top three clusters. This contrasts the mean Jaccard index for smaller cities, which decreases to [0.44, 0.59] if the top five clusters are excluded. This tendency is related to a size-biased pecking order in national resource allocation: the density of existing road networks well reflects the actual demand in big cities, while road infrastructure in smaller cities is relatively undersupplied. Also, the similarity tends to decrease when our results are compared with reference maps using more recent data. This seems be associated with a time lag involved in transport infrastructure investment. It takes time to respond to a demand shock, and existing road networks are likely concurrent with past demand with a lag of a few years.

The main contribution of this study is twofold. From a methodological perspective, we add further empirical evidence to support use of the percolation method and widely available stock data for identifying functional urban regions. The adopted percolation method can beget similar results to those based on flow data and is applicable to virtually all countries, including those developing countries where reliable subnational-level flow data is not available. Another strength of the approach is that it does not require strong presumptions on the total number or spatial limits of major clusters. This provides a great advantage in identifying crucial changes in urban networks that exceed a conventionally expected level, as exemplified by the spatial expansion of the SMA.

From a policy perspective, our results suggest the need for more balanced road infrastructure investment in South Korea. The SMA's dominance in the national urban system and lower Jaccard coefficients for smaller cities suggest that Korea's lower tier functional regions suffer weak connectivity to the main networks and an undersupply of road infrastructure. Improving these aspects seems to be a necessary condition to counterbalance a Seoul–centered mono-centric structure.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## ORCID iDs

Yifu Ou https://orcid.org/0000-0001-8756-8018
Euijune Kim https://orcid.org/0000-0003-4354-5150
Xingjian Liu https://orcid.org/0000-0003-4158-3655
Kyung-Min Nam https://orcid.org/0000-0002-4579-4815

## Supplemental Material

Supplemental material for this article is available online

## Note

1. However, the data quality of OpenStreetMap or other open sources is often inconsistent across regions, and thus can limit the applicability of network-based regional delineation

## References

Arcaute E, Molinero C, Hatna E, et al. (2016) Cities and regions in britain through hierarchical percolation. *Royal Society Open Science* 3(4): 150691.

Batty M (2007) *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*. Cambridge, MA: MIT Press.

Boeing G (2020) A multi-scale analysis of 27,000 urban street networks: every US city, town, urbanized area, and Zillow neighborhood. *Environment and Planning B: Urban Analytics and City Science* 47(4): 590–608.

Bosker M, Park J and Roberts M (2021) Definition matters. Metropolitan areas and agglomeration economies in a large-developing Country. *Journal of Urban Economics* 125: 103275.

Brenner N (2003) Standortpolitik, state rescaling and the new metropolitan governance in western Europe. *disP - The Planning Review* 39: 15–25.

Brown LA and Holmes J (1971) The delimitation of functional regions, nodal regions, and hierarchies by functional distance approaches. *Journal of Regional Science* 11: 57–72.

Brown PJB and Hincks S (2008) A framework for housing market area delineation: principles and application. *Urban Studies* 45(11): 2225–2247.

Cai J, Huang B and Song Y (2017) Using multi-source geospatial big data to identify the structure of polycentric cities. *Remote Sensing of Environment* 202: 210–221.

Cao W, Dong L, Wu L, et al. (2020) Quantifying urban areas with multi-source data based on percolation theory. *Remote Sensing of Environment* 241: 111730.

Ch R, Martin DA and Vargas JF (2021) Measuring the size and growth of cities using nighttime light. *Journal of Urban Economics* 125: 103254.

Chen B, Song Y, Huang B, et al. (2020) A novel method to extract urban human settlements by integrating remote sensing and mobile phone locations. *Science of Remote Sensing* 1: 100003.

Dawkins CJ (2003) Regional development theory: conceptual foundations, classic works, and recent developments. *Journal of Planning Literature* 18: 131–172.

Dijkstra L, Florczyk AJ, Freire S, et al. (2021) Applying the degree of urbanisation to the globe: a new harmonised definition reveals a different picture of global urbanisation. *Journal of Urban Economics* 125: 103312.

Dingel JI, Miscio A and Davis DR (2021) Cities, lights, and skills in developing economies. *Journal of Urban Economics* 125: 103174.

Doxsey-Whitfield E, MacManus K, Adamo SB, et al. (2015) Taking advantage of the improved availability of census data: a first look at the gridded population of the world, version 4. *Papers in Applied Geography* 1(3): 226–234.

Duranton G (2015) Delineating metropolitan areas: measuring spatial labour market networks through commuting patterns. In: Watanabe T, Uesugi I and Ono A (eds), *The Economics of Interfirm Networks*. Tokyo: Springer.

Duranton G (2021) Classifying locations and delineating space: an introduction. *Journal of Urban Economics* 125: 103353.

Fan J, Ma T, Zhou C, et al. (2014) Comparative estimation of urban development in china's cities using socioeconomic and dmsp/ols night light data. *Remote Sensing* 6(8): 7840–7856.

Farmer CJQ and Fotheringham AS (2011) Network-based functional regions. *Environment and Planning A: Economy and Space* 43(11): 2723–2741.

Gao S, Janowicz K and Couclelis H (2017) Extracting urban functional regions from points of interest and human activities on location-based social networks. *Transactions in GIS* 21: 446–467.

Government of the Republic of Korea (2001) *The Fourth Comprehensive National Territorial Plan (2000-2020)*. Anyang: Korea Research Institute for Human Settlements.

Government of the Republic of Korea (2019) *The Fifth Comprehensive National Territorial Plan (2020-2040)*. Sejong: Government of the Republic of Korea. (in Korean).

Hammer MS, van Donkelaar A, Li C, et al (2022) *Global Annual PM$_{2.5}$ Grids from MODIS. MISR and SeaWiFS Aerosol Optical Depth (AOD), 1998-2019. V4.GL.03. Palisades*, New York: NASA Socio-economic Data and Applications Center. DOI: 10.7927/fx80-4n39 (accessed 1 January 2022).

Harari M (2020) Cities in bad shape: urban geometry in India. *American Economic Review* 110(8): 2377–2421.

Harrison J and Hoyler M (2015) *Megaregions: Globalization's New Urban Form?* Cheltenham: Edward Elgar Publishing.

Henderson JV, Nigmatulina D and Kriticos S (2021) Measuring urban economic density. *Journal of Urban Economics* 125: 103188.

Henderson JV, Storeygard A and Weil DN (2012) Measuring economic growth from outer space. *The American Economic Review* 102(2): 994–1028.

Houssou NLJ, Guillaume J-l and Prigent A (2019) A graph based approach for functional urban areas delineation. *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing* 652–658.

Huynh HN, Makarov E, Legara EF, et al. (2018) Characterisation and comparison of spatial patterns in urban systems: a case study of U.S. cities. *Journal of Computational Science* 24: 34–43.

Jia T, Yu X, Shi W, et al. (2019) Detecting the regional delineation from a network of social media user interactions with spatial constraint: a case study of Shenzhen, China. *Physica A: Statistical Mechanics and Its Applications* 531: 121719.

Kim H, Kim DW and Cho JY (2019) Exploring the key communicator role of exosomes in cancer micro-environment through proteomics. *Proteome Science* 17(7): 5–23. (in Korean). DOI: 10.17208/jkpa.2019.12.54.7.5.

Kim D-J, Joo M-J, Min S-H, et al. (2012) *A Study on Global City-Regions in Korea (III)*. Anyang: Korea Research Institute for Human Settlements. (in Korean).

Kim E, Hewings G and Nam K-M (2014) Optimal urban population size: national vs local economic efficiency. *Urban Studies* 51(2): 428–445.

Kwon K and Seo M (2018) Does the polycentric urban region contribute to economic performance? The case of Korea. *Sustainability* 10(11): 4157.

Lee S and Lee H (2015) Delimitation of city-regions based on the method of travel-to-working area and analyzing spatial structure of city-regions. *The Korea Spatial Planning Review* 84: 165–189. (in Korean).

Li X and Nam K-M (2017) One country two "urban" systems: focusing on bimodality in China's city-size distribution. *The Annals of Regional Science* 59(2): 427–452.

Li X and Zhou Y (2017) Urban mapping using dmsp/ols stable night-time light: a review. *International Journal of Remote Sensing* 38(21): 6030–6046.

Liu X, Ning X, Wang H, et al. (2019) A rapid and automated urban boundary extraction method based on nighttime light data in China. *Remote Sensing* 11(9): 1126.

Long Y (2016) Redefining Chinese city system with emerging new data. *Applied Geography* 75: 36–48.

Long Y, Shen Y and Jin X (2016) Mapping block-level urban areas for all Chinese cities. *Annals of the American Association of Geographers* 106(1): 96–113.

Masser I and Scheurwater J (1980) Functional regionalisation of spatial interaction data. an evaluation of some suggested strategies. *Environment and Planning A: Economy and Space* 12(12): 1357–1382.

Masucci AP, Arcaute E, Wang J, et al. (2015) *Logistic Growth and Ergodic Properties of Urban Forms. arXiv preprint arXiv*:1504.07380.

Mitchell A and Griffin LS (2021) *Spatial Measurements and Statistics*. Redlands, CA: ESRI Press.

Mitchell W and Watts M (2010) Identifying functional regions in australia using hierarchical aggregation techniques. *Geographical Research* 48(1): 24–41.

Mu X and Yeh AGO (2020) Regional delineation of China based on commuting flows. *Environment and Planning A: Economy and Space* 52(3): 478–482.

Nam K-M (2009) Spatial integration of corporate r&d and mass production activities: a case study of samsung electronics. *Korea Spatial Planning Review* 62: 125–145.

Nam K-M (2017) Is the spatial distribution of China's population excessively unequal? A cross-country comparison. *The Annals of Regional Science* 59(2): 453–474.

Nam K-M and Reilly JM (2013) City size distribution as a function of socioeconomic conditions: an eclectic approach to downscaling global population. *Urban Studies* 50(1): 208–225.

National Geographic Information Institute, Republic of Korea (2021) Korea National Spatial Data Infrastructure Portal - National Road Network Shapefiles. Available at: http://data.nsdi.go.kr/dataset/20180927ds0062 (accessed on 15 January 2021).

Niu H and Silva EA (2021) Delineating urban functional use from points of interest data with neural network embedding: a case study in greater London. *Computers, Environment and Urban Systems* 88: 101651.

Oak Ridge National Laboratory (2022) LandScan Global Population Distribution Dataset. Available at: https://landscan.ornl.gov/index.php/landscan-datasets (accessed on 1 January 2022).

Park S, Kim SY and Nam K-M (2019) Effectiveness of worksite-based dietary interventions on employees' obesity: a systematic review and meta-analysis. *Nutrition Research and Practice* 13(3): 399–409.

Poorthuis A (2017) How to draw a neighborhood? The potential of big data, regionalization, and community detection for understanding the heterogeneous nature of urban neighborhoods. *Geographical Analysis* 50: 182–203.

Porta S, Strano E, Iacoviello V, et al. (2009) Street centrality and densities of retail and services in Bologna, Italy. *Environment and Planning B: Planning and Design* 36(3): 450–465.

Ratti C, Sobolevsky S, Calabrese F, et al. (2010) Redrawing the map of great britain from a network of human interactions. *PLoS One* 5(12): e14248.

Rybnikova N, Portnov BA, Charney I, et al. (2021) Delineating functional urban areas using a multi-step analysis of artificial light-at-night data. *Remote Sensing* 13(18): 3714.

Smith M, Szongott C, Henne B, et al. (2012) Big Data Privacy Issues in Public Social Media. In: IEEE International Conference on Digital Ecosystems and Technologies (DEST), Paper presented at the 2012 6th Campione d'Italia, Italy, 18-20 June 2012. DOI: 10.1109/DEST.2012.6227909.

Song Y, Chen B and Kwan M-P (2020) How does urban expansion impact people's exposure to green environments? A comparative study of 290 Chinese cities. *Journal of Cleaner Production* 246: 119018.

Statistics Korea (2022) Korean Statistical Information Service (KOSIS) database. Available at: https://kosis.kr/search/search.do (accessed on 15 April 2022.)

Storper M (2014) Governing the large metropolis. *Territory, Politics, Governance* 2(2): 115–134.

Sulis P, Manley E, Zhong C, et al. (2018) Using mobility data as proxy for measuring urban vitality. *Journal of Spatial Information Science* 16: 137–162.

Taylor R (1990) Interpretation of the correlation coefficient: a basic review. *Journal of Diagnostic Medical Sonography* 6(1): 35–39.

Wang F, Antipova A and Porta S (2011) Street centrality and land use intensity in baton rouge, Louisiana. *Journal of Transport Geography* 19(2): 285–293.

Wu K, Tang J and Long Y (2019) Delineating the regional economic geography of china by the approach of community detection. *Sustainability* 11(21): 6053.

Yeh AG-O and Chen Z (2020) From cities to super mega city regions in china in a new wave of urbanisation and economic transition: issues and challenges. *Urban Studies* 57(3): 636–654.

Zhong C, Arisona SM, Huang X, et al. (2014) Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science* 28(11): 2178–2199.

Zhou Y, Smith SJ, Zhao K, et al. (2015) A global map of urban extent from nightlights. *Environmental Research Letters* 10(5): 054011.

**Dr Yifu Ou** is a Senior Research Assistant in the Department of Urban Planning and Design at the University of Hong Kong. His current research interests lie in environmental sustainability associated with spatial and transportation planning and air quality valuation.

**Dr Euijune Kim** is Professor of Regional Science in the Department of Agricultural Economics and Rural Development at Seoul National University. His expertise is in urban and regional economics, spatial and infrastructure policy analysis, and applied general equilibrium modeling.

**Dr Xingjian Liu** with the Department of Urban Planning and Design at the University of Hong Kong. He studies urban and regional development and is a member of the Globalization and World Cities Research Network.

**Dr Kyung-Min Nam** is Associate Professor in the Department of Urban Planning and Design at the University of Hong Kong. His research interests focus on spatial planning/policy and climate change, and he has widely applied econometrics, computable general equilibrium models, and GIS tools to his empirical research.