XXVIII International Seminar on Urban Form
ISUF2021: URBAN FORM AND THE SUSTAINABLE AND PROSPEROUS CITIES
29th June – 3rd July 2021, Glasgow

A Taxonomy of Towns: An unsupervised machine learning approach to classify towns in England and Wales.

Dr Kimon Krenz[1], Prof Alan Penn[2], Dr Tasos Varoudis[3]

[1,2,3] Space Syntax Laboratory, The Bartlett School of Architecture, Faculty of the Built Environment, UCL, United Kingdom

Abstract

In 2019, the UK Government pledged to focus on levelling up the UK's underperforming regions. As part of this effort, the Government recognised towns as a vital component in the hierarchical network of regional urban systems. Towns not only link cities to the wider hinterland but can also be places of innovation in their own right. Yet struggling towns, due to their dependence on cities and the surrounding region, often lack the fundamentals to support a strong local economy. Such towns face major socio-economic challenges once economic trajectories decline, including ageing population; lack of existing skills necessary to attract new firms; lack of education; less direct investments; and hindering spatial configurations. Given this context, the aim of this paper is to establish a classification of towns to understand their similarities to support targeted investment and to offer comparative characteristics for policy evaluation. To this end, this study develops a new classification of all towns in England and Wales across a variety of socio-spatial and economic domains. The analysis includes 1,178 urban areas with a population between 5,000 and 225,000. Specifically, we employ 105 workplace-based and residence-based demographic and economic variables of the 2011 Office for National Statistics Census for England and Wales and combine these with newly developed spatial variables on network similarities on the basis of network topology, geometry and centrality metrics. These variables are aggregated on boundaries of the built-up area of towns utilising centroid-based ONS lookup tables. We account for differences in distribution and scale through data transformation and standardisation. We then employ a K-means unsupervised cluster algorithm to establish a two-tier class system, of which the first is presented in this paper. The result is 6 distinctive Supergroups of towns. We further provide descriptive characterisations of each Supergroup and insights into the importance of individual variables.

Keyword: *towns, morphology, machine learning, policy evaluation, geography*

Introduction

The UK faces a policy imperative to 'level-up' society, and in particular to address regional disparities. Levelling up has in the past focussed on regional variations in economic performance such as the 'productivity gap' between London and other regions, however more recently the inequalities embedded in UK society have been highlighted first by the EU referendum which showed stark regional differences in identity, and now by the global pandemic where different health outcomes both regionally and between different communities of ethnic origin. Underlying this variation is seen between prosperous and less prosperous places and groups. 'Levelling-up' is now used in a policy context to refer to both social and economic rebalancing within a place-based view. One set of policies focuses on the perceived decline of towns and especially their high streets. Significant sums are being invested in towns around the country in order to help regenerate so-called 'left behind' places. This has brought a new impetus to attempts from the early 20th century to develop a taxonomy or classification of kinds of town. Unless we can understand the differences between different places it is hard to know how best to tailor investment to the local context, nor to evaluate the results of investment through comparison of different investments between similar as opposed to dissimilar places. This raises the fundamental question: when each town is essentially different to all others how can we go about comparing them?

Background

There has been considerable debate over the years about how to characterise human settlements such as towns. Are these best considered in social terms through the demographics of their residents, in economic terms as places of productive activity and employment, or in morphological terms as arrangements of buildings and open space? Clearly, all three, the social, economic and morphological may also be related through the history of the production of the built environment. As settlements grow periods of expansion in population related to a changing local economy can be seen to form fringe belts of morphological difference. Analysis of these has provided a central component of MP Conzen's morphological analysis (2009). Equally, periods of economic or institutional decline can lead to the desertion of buildings and the disappearance of whole settlements. Maurice Beresford has documented the lost villages of mediaeval England (1985). However, neither of these have attempted a systematic classification perhaps because morphological properties seem poorly suited to classifying in any but the simplest of terms, and the morphological tradition has concentrated instead on describing the unique form and history of individual settlements rather than to seek to establish any kind of taxonomy. Early classifications in the geographic literature have focused instead on administrative or functional differences. Marcel Aurousseau (1921) classified towns into administrative, defensive, cultural, production-towns, communication and recreation. His classification has the benefit of simplicity but at the expense of missing complexity; in this classification, a town can only be one thing. This approach was developed progressively through the use of employment data and industry sector classifications in the American context by Chauncy Harris (1943), Healy et. al (1957), and Howard Nelson (1955) amongst others. By making use of data, they progressively developed the classification of functional type, but always focussed on towns and cities thought of as places of economic production and transaction. A third approach was developed by Bill Hillier and Julienne Hanson (1984) who proposed a mechanism through which urban aggregations considered as a rule restricted random processes, could take on specific morphologies. By writing a linguistic notation to define the relationships between built form and open space they demonstrated a set of morphological 'types' which aggregations could take up. Next, they developed methods to represent and measure the pattern properties of morphologically different settlements so as to allow comparison. Using these analytic methods, they then carried out a series of empirical studies of urban function along with a theoretical framework linking social, economic and morphological dimensions. The research reported in this paper makes use of these analytic methods to bring town morphology within the ambit of a socio-economic classification. In this paper, we describe a method of iterative unsupervised clustering on three discrete data sets to find distinct categories amongst all English and Welsh towns.

Methodology

To address the outlined policy need and lack of available methods for the comparison of towns, this research combines existing methodological work on the morphological classification of the built environment through spectral clustering of street networks (Varoudis and Penn, 2021) with established areas classification through socio-economic and demographic Census information (Cockings, Martin and Harfoot, 2015; ONS, 2015; Gale *et al.*, 2016). This is done at the geographic scale of 'towns' in England and Wales utilising built-up area (BUA) geographies (ONS, 2013). Specifically, we propose to aggregate i) new spatial variables describing differences in street networks morphologies, and ii) residential-based and iii) workplace-based 2011 Census information on social, economic and demographic characteristics at the town level using BUA geographies. The aim is to capture a more comprehensive picture of a towns everyday population and spatial character through the combination of residential- and workplace-based data with morphological information. All variables are converted into percentage shares of the respective characteristic to maintain comparability across areas of different sizes and scales. Furthermore, to account for differences in data distribution, scale, multicollinearity and magnitude between variables we remove all variables featuring collinearity, high skewed distributions and subsequently transform and scale all variables. This dataset is then clustered into distinctive groups using the unsupervised machine learning technique k-means clustering. The k-means cluster technique is a widely used unsupervised machine learning technique that features a series of advantages over alternative techniques, such as it is appropriate for complex and large datasets, it has substantially less time complexity compared to hierarchical algorithms, and it is easier to interpret. K-means cluster solutions are produced using the R package *kmeans* and the Hartigan-Wong algorithm (Hartigan and Wong, 1979). The Hartigan and Wong algorithm defines the total within-cluster variation as the sum of squared Euclidean distances between features and the corresponding centroid as:

Equation 1:

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2$$

Where x_i is a data point belonging to the cluster C_k, and μ_k is the mean value of all points assigned to the cluster C_k. Cluster solutions were produced for 3 to 10 clusters and these were further subdivided into 2 to 6 clusters (we have not included the results of the subdivision in this paper due to space limitations) following the approach by Cockings et al. (2015). All solutions were evaluated employing a combination of 30 indices (Charrad *et al.*, 2014) together with mappings to determine whether solutions produced meaningful outputs. Finally, we provide initial descriptions for each group and elaborate on their distinctive characteristics.

Defining Towns Through Built-up Areas

There has been a long-standing debate about which area should and can be classified as 'town', in opposite to 'city' or 'village' respectively (Baker, 2018). In England and Wales, towns have traditionally been settlements that feature a market or fair, and hence differing through their economical function to surrounding villages and the wider hinterland. Since then, the term town has been used to describe various technical and legal categories of administrative and governmental boundaries. It is not the aim of this work to address the dispute about the classification of settlements, instead, we will utilise an existing geographic demarcation and its classification by the Office for National Statistics (ONS, 2013) called 'built-up areas'. Built-up areas are defined as "as land which is 'irreversibly urban in character'" which reflects settlements that are either villages, towns or cities. The 4-stage process identifies cells within a 50m grid of the British National Grid System, that fall within the minimum coverage threshold including buildings, tarmac and metalled surfaces as well as primary gardens (ONS, 2013, p. 4). Adjacent cells are then dissolved into continuous polygons. Polygons that are within 200m of each other are linked and sub-divisions identified, and areas of less than 200,000 m^2 are removed (ONS, 2013, p. 5). Finally, the remaining areas a named and classified into villages, towns and cities based on size through an automated process (see Figure 1 for visualisation of BUA boundary of the town Rothwell). Towns are those BUAs that feature a population between 5,000 and 225,000. For this paper, we select all BUAs and sub-areas in England and Wales that have been classified as towns, i.e., 1,186 geographies. We further disregarded towns that are not part of the continuous mainland, resulting in a final selection of 1,176 towns.

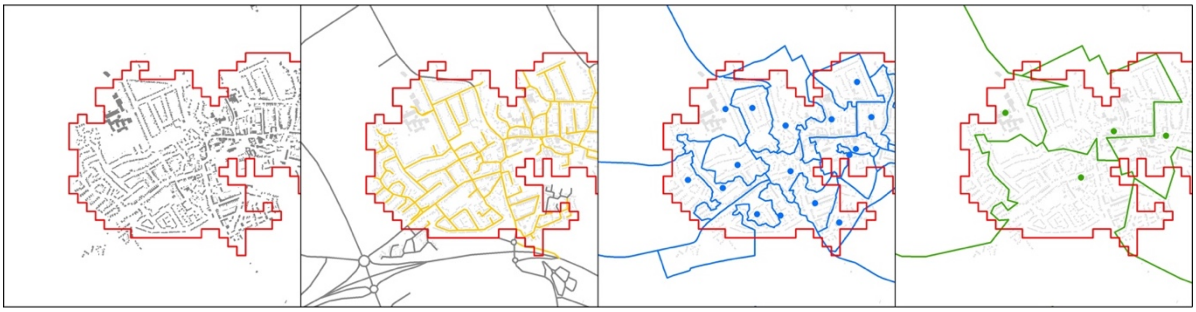

Figure 1. *Figure-ground plan and BUA boundary (red) of the town Rothwell in the East Midlands; Street network selection and BUA boundary (orange); Output area (blue) and population-weighted output area centroids; Workplace zones (green) and population-weighted workplace zone centroids (from left).*

Built environment characteristics

To capture morphological characteristics between different towns, we are using a spectral clustering approach (Varoudis and Penn, 2021). Specifically, we created detailed topo-geometric encodings of the road-central-line dataset of the UK. We concentrated on the angle-scale composite spectra that brings out the inner dynamics of spatial penetration of local neighbourhoods in different scales and dimensions by focusing purely on geometric features. The spectral clustering output is a vector of high dimensionality that can be compared against other vectors for detailed analyses. From a road segment dataset Varoudis and Penn (2021) extracted one local subgraph per street segment of a set size 200, which ensures that any computational analysis is always made between graphs of equal size (Robles-Kelly and Hancock, 2003; Varoudis and Penn, 2021). We then use this subgraph (of N=200) to compute the graph spectra weighted by angular and metric distance. The two vectors of eigenvalues, one from the angular weighting and one from the metric, are then sorted and concatenated to form a composite feature (a vector of size 400). Spectra extracted in this composite form can capture differences between the overall topo-geometric structure of graphs and are unique to a particular spatial configuration (Luo, Wilson and Hancock, 2003; Robles-Kelly and Hancock, 2003; Hanna, 2012; Varoudis and Penn, 2021). Finally, we project composite vectors (of N=400) onto a 400-dimensional feature space creating a point-cloud representing the gradual changes of the spatial morphology of the neighbourhoods. The same point cloud is also used as input to the k-means unsupervised model in order to create distinct clusters of morphological differences. The result is an 8-cluster solution (see Figure 2). We subsequently aggregate the number of streets grouped by each cluster at the town level. This is done by aggregating those streets whose centroid falls within the BUA boundaries (see Figure 1). Finally, we calculate the percentage share of each cluster per town (see Appendix for full variable list).

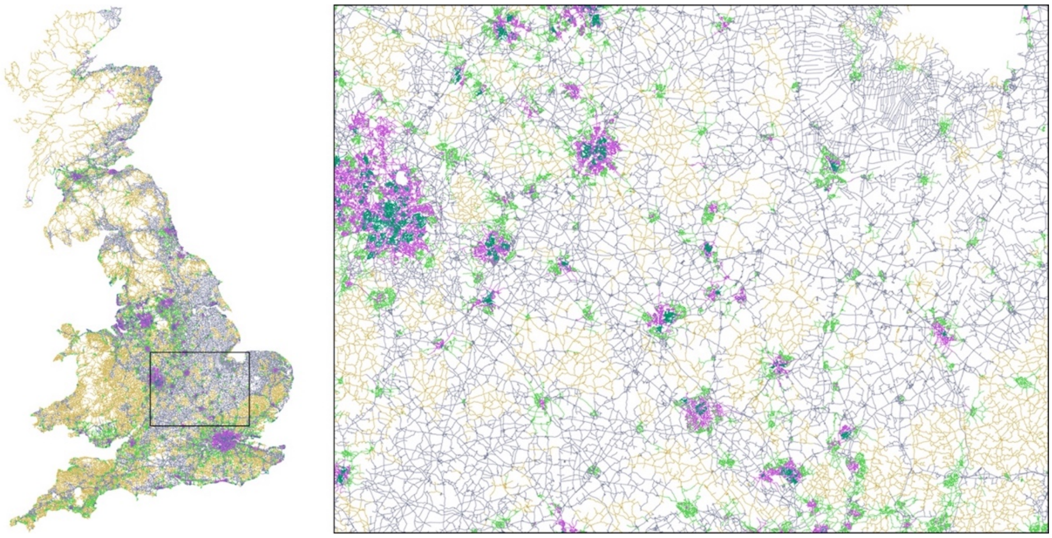

Figure 2. *Morphological classification of UK's street networks. Visualisation of 8-cluster solution showing distinctive topo-geometric groups.*

Residential-based socio-economic and demographic characteristics

To capture residence-based social, economic and demographic characteristics of towns we are building on the established ONS area classification methodology (ONS, 2015), which utilises residential-based 2011 Census information in combination with a k-means cluster approach. Residential-based Census information is gathered through surveys at the place of residence. For example, the data provides information on whether people are employed and where they live, but not where they are employed. From this data, we are using 59 variables across 5 variable domains, which include variables on demographic structure such as "the percentage of persons whose country of birth is the UK or Ireland", but also information on household composition, housing, socio-economic character, and employment. Variables are accessed through 2011 Census Key Statistics Quick Statistics table published by ONS for England and Wales at the geographic level of Output Areas (OA) and aggregated at BUA boundaries, where OA areas do not fall within or coincide with BUAs, we employ population-weighted centroids to aggregate values (see Figure 1).

Workplace-based socio-economic and demographic characteristics

To capture social, economic and demographic characteristics of the working population of a town we are utilising workplace-based 2011 Census information at the geographic level of Workplace Zones (WZ) and a classification methodology established by Cockings et al. (2015). Workplace Zones are the smallest available geography containing a consistent number of workers and hence allowing to capture statistics characteristics of the working population at a granular level. This Census information is gathered through surveys at the place of work and includes information on for example the age structure of the workers and workplaces across the UK. We aggregated 46 variables at the WZ level from three main variable domains to BUA boundaries, and converted figures to raw percentages. Where WZ areas do not fall within or coincide with BUAs, we employ population-weighted centroids to match and aggregate values (see Figure 1).

Data preparation

We have combined all variables into a single dataset. For the Census statistics, as well as street morphology indices used, figures were converted to raw percentages where these were not already provided as such. The base for these conversions was the residential population, workplace population and the total number of streets respectively. Furthermore, we controlled for multicollinearity and removed those variables with a pair-wise correlation higher or equal to R^2 of 0.85, as well as highly skewed variables. Finally, we accounted for potential outliers and differences in distribution by transforming the dataset using the Box-Cox method (ONS, 2015), and addressed issues of different scales or magnitudes between variables by standardising each variable using the range standardisation method (ONS, 2015).

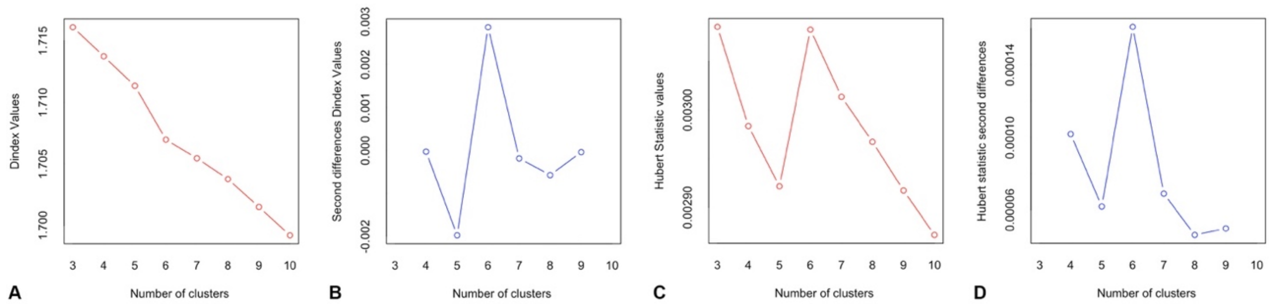

Figure 3. *A: D Index plot. B: D index second difference plot. C: Hubert Index plot. D: Hubert second difference plot. The D index and the Hubert index are graphical methods of determining the number of clusters.*

Results and Discussion

Unsupervised k-means cluster approaches can lead to multiple acceptable solutions and require the researcher to specify the parameter *k*, i.e., the number of clusters. Due to these multiple acceptable solutions, results are not strictly scientific in this sense and results need to be interpreted with caution. While there has not been a universally reliable method to determine *k* (Banks and Fienberg, 2003), there are several available methods that can aid a researcher in the decision-making process. We produced 30 indices as well as a visual comparison such as comparing the Hubert and D index to evaluate the optimal number of clusters using the R package *NbClust* (Charrad *et al.*, 2014). Figure 3 shows graphical representations of the In D and

Hubert index. For both indices, we seek a significant knee that corresponds to a significant increase of the value of the measure i.e the significant peak in D and Hubert index second differences plot. Both measures indicate an optimal cluster number of 6. We further investigated the geographic distribution of cluster solutions through mappings and visualisations which corresponds with the previous findings. The result is a two-tier hierarchy of 6 Supergroups (top-level) and 17 Groups (second level). The following section will elaborate on the top-level, i.e., the 6 Supergroups.

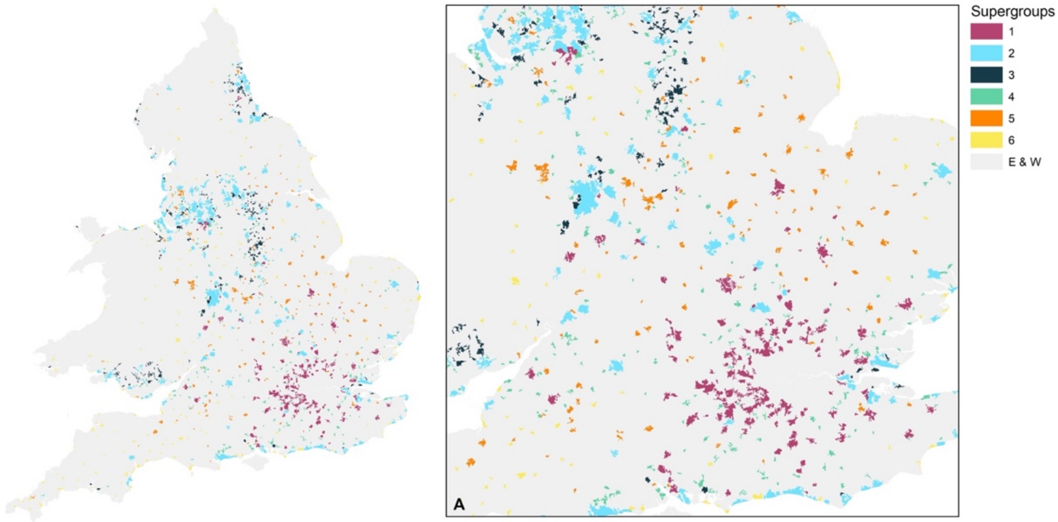

Figure 4. *Visualisation of Supergroups. 6-cluster solution, towns with social, economic, demographic and morphological similarities are grouped and coloured accordingly. A) shows a detailed zoom in to central England.*

Supergroups

The geographic distribution of the Supergroups is shown in Figure 4. The visualisation provides interesting insights into spatial similarities between different towns. For example, cluster 1 (purple) agglomerates predominantly in London's west around Slough but also includes commuting cities linked to London such as Oxford, Cambridge and Milton Keynes. Dense urban areas are captured by cluster 2 (light blue), but also towns in proximity to centres of larger cities (e.g., Liverpool, Manchester, Birmingham, Nottingham, among others). Cluster 3 (dark blue) highlights towns in former coal-mining areas around Newport and Newcastle, as well as between Nottingham and Leeds. Cluster 4 captures small towns in proximity to dense urban areas (i.e., clusters 1 and 5). Mid-sized towns predominantly located between major urban agglomerations are captured by cluster 5 (orange). Finally, smaller towns in Cornwall, Wales, as well as Seaside towns are captured by cluster 6 (yellow). Towns within the same Supergroup share similar characteristics to other Supergroup members and, hence, allow direct comparison.

Supergroup classification

We investigate similarities and differences of each Supergroup's characteristic by comparing median values of a representative sets of selected workplace- and residential-based employment and demographic variables, as well as median street cluster values. This provides insights into the dominant industrial activity of a town's economically active population, the demographic composition, as well as the dominant spatial characteristic. Supergroups can be divided into *international metropole towns*, *dense urban towns*, *industrial towns*, *knowledge-based towns*, *non-urban towns*, and *agricultural and tourism-based towns*. These groupings are not only reflected in the Census-based variables but produce distinctive geographic patterns across England and Wales (see Figure 4). We can observe that Supergroups have a relationship with the size of the area of a town, which is unexpected, as no size-dependent variable has been included in the analysis. This points to a general trend that larger towns tend to be more heterogeneous, and smaller towns tend to be more homogeneous and as such share similar socio, economic and spatial characteristics. Supergroup 1, captures spatially, economically as well as demographically the Greater London metropole area, but also towns in proximity to other major cities in England and Wales, such as Manchester or Birmingham. Supergroup 3, can be seen as the opposite pole to Supergroup 1 and is predominantly located in former coal-mining areas and post-industrial regions. Supergroup 6 describes the characteristic of seaside towns, as well as agricultural towns. A detailed breakdown of the differences and characteristics of each Supergroup can be found in Appendix 2 Table 1. Furthermore, we have compared the towns which have been selected to participate in the Governments Town Fund policy against the distribution of each Supergroup. This comparison has shown that the selected towns are unequally distributed among Supergroups. 58% of all selected towns are from Supergroup 2,15% from Supergroup 3, and 10% from 1,5 and 6, while no town of Supergroup 4 has been selected, albeit being the largest group.

Conclusion

This study has provided a new taxonomy of towns, based on three core social, economic and morphological domains. In order to do this, a series of assumptions have been made. Amongst these is first, that economic, social and spatial variables are able to capture the character of a town, and secondly, that k-means is an appropriate method for the classification of these. In addition, we like to highlight some limitations of this approach. Specifically, we have relied on the lowest publicly available geographic level of Census information which has been gathered in 2011. This could be seen as outdated and might, hence, not reflect the current condition of towns effectively. However, the resulting town classification should not be seen as finite to begin with, towns are like any other urban environment place of constant change and evolution, and transformations are likely in the future. Instead, the contribution of this work is the presentation of a method to establish a taxonomy of town, as well as an initial classification which can be used to group towns in a two-tier hierarchy and allow their direct comparison. As such it provides a valuable source of information to aid policy decision-making processes and inform efforts to level-up the UK.

References

Aurousseau, M. (1921) 'The Distribution of Population: A Constructive Problem', *Geographical Review*, 11(4). doi: 10.2307/208251.

Baker, C. (2018) 'City & Town Classification of Constituencies & Local Authorities'. London: House of Commons Library. Available at: https://commonslibrary.parliament.uk/research-briefings/cbp-8322/.

Banks, D. L. and Fienberg, S. E. (2003) 'Data Mining, Statistics', in *Encyclopedia of Physical Science and Technology*. doi: 10.1016/b0-12-227410-5/00164-2.

Beresford, M. W. (1985) *Time and Place: Collected Essays*. A&C Black.

Charrad, M. *et al.* (2014) 'Nbclust: An R package for determining the relevant number of clusters in a data set', *Journal of Statistical Software*, 61(6). doi: 10.18637/jss.v061.i06.

Cockings, S., Martin, D. and Harfoot, A. (2015) *A Classification of Workplace Zones for England and Wales (COWZ-EW) User Guide*. Available at: http://cowz.geodata.soton.ac.uk/download/.

Conzen, M. P. (2009) 'How cities internalize their former urban fringes: A cross-cultural comparison.', *Urban Morphology*, 13(1), pp. 29–54.

Gale, C. G. *et al.* (2016) 'Creating the 2011 area classification for output areas (2011 OAC)', *Journal of Spatial Information Science*, 12(2016). doi: 10.5311/JOSIS.2016.12.232.

Hanna, S. (2012) 'Comparative Analysis of Neighbourhoods Using', in *8th International Space Syntax Symposium*.

Harris, C. D. (1943) 'A Functional Classification of Cities in the United States', *Geographical Review*, 33(1). doi: 10.2307/210620.

Hartigan, J. A. and Wong, M. A. (1979) 'Algorithm AS 136: A K-Means Clustering Algorithm', *Applied Statistics*, 28(1). doi: 10.2307/2346830.

Healy, S. M. E., Duncan, O. D. and Reiss, A. J. (1957) 'Social Characteristics of Urban and Rural Communities, 1950', *The American Catholic Sociological Review*, 18(3). doi: 10.2307/3708727.

Hillier, B. and Hanson, J. (1984) *The Social Logic of Space*, *The Social Logic of Space*. doi: 10.4324/9780429450174-9.

Luo, B., Wilson, R. C. and Hancock, E. R. (2003) 'Spectral embedding of graphs', *Pattern Recognition*, 36(10). doi: 10.1016/S0031-3203(03)00084-0.

Nelson, H. J. (1955) 'A Service Classification of American Cities', *Economic Geography*, 31(3). doi: 10.2307/142045.

ONS (2013) *2011 Built-up Areas - Methodology and Guidance Version*. Available at: https://geoportal.statistics.gov.uk/Docs/Built-up_areas_user_guidance.zip.

ONS (2015) *Methodology Note for the 2011 Area Classification for Output Areas*.

Robles-Kelly, A. and Hancock, E. R. (2003) 'Edit distance from graph spectra', in *Proceedings of the IEEE International Conference on Computer Vision*. doi: 10.1109/iccv.2003.1238347.

Varoudis, T. and Penn, A. (2021) 'Spectral Clustering and Integration: The Inner Dynamics of Computational Geometry and Spatial Morphology', in *Advances in Science, Technology and Innovation*. doi: 10.1007/978-3-030-57509-0_22.

APPENDIX 1

Table 1. Variables

Built environment characteristics		
be_cl_1_scaled	Scaled percentage share of street network cluster group 1	Street Network
be_cl_2_scaled	Scaled percentage share of street network cluster group 2	
be_cl_4_scaled	Scaled percentage share of street network cluster group 4	
be_cl_5_scaled	Scaled percentage share of street network cluster group 5	
be_cl_6_scaled	Scaled percentage share of street network cluster group 6	
be_cl_7_scaled	Scaled percentage share of street network cluster group 7	
be_cl_8_scaled	Scaled percentage share of street network cluster group 8	
Residential-based characteristics		
dem_rb_0_4	Percentage of persons aged 0 to 4 years	Demographic Structure
dem_rb_5_14	Percentage of persons aged 5 to 14 years	
dem_rb_45_64	Percentage of persons aged 45 to 64 years	
dem_rb_65_89	Percentage of persons aged 65 to 89 years	
dem_rb_90	Percentage of persons aged 90 years and over	
dem_rb_mar	Percentage of persons aged 16 years and over who are married or in a registered same-sex civil partnership	
dem_rb_div	Percentage of persons aged 16 years and over who are divorced or separated	
dem_rb_mixed	Percentage of persons who have mixed ethnicity or are from multiple ethnic groups	
dem_rb_ipb	Percentage of persons who are Asian/Asian British: Indian/Pakistani/Bangladeshi	
dem_rb_ch	Percentage of persons who are Asian/Asian British: Chinese or other	
dem_rb_acb	Percentage of persons who are Black/African/Caribbean/Black British	
dem_rb_a	Percentage of persons who are Arab or are from another ethnic group	
dem_rb_langnoeng	Percentage of persons whose main language is not English and cannot speak English well or at all1	
dem_rb_brith_uk	Percentage of persons whose country of birth is the UK or Ireland	
dem_rb_birth_eu_2001	Percentage of persons whose country of birth is in the new EU (post 2004 accession countries)	
hou_rb_nochild	Percentage of households with no children	Household Composition
hou_rb_nodepchild	Percentage of households with non-dependent children	
hou_rb_detached	Percentage of households who live in a detached house or bungalow	Housing
hou_rb_semi	Percentage of households who live in a semi-detached house or bungalow	
hou_rb_terrace	Percentage of households who live in a terrace or end-terrace house	
hou_rb_flat	Percentage of households who live in a flat	
hou_rb_temp	Percentage of households who live in a caravan or other mobile or temporary structure	
hou_rb_socrent	Percentage of households who are social renting	
hou_rb_privrent	Percentage of households who are private renting	
hou_rb_lessrooms	Percentage of households who have one fewer room or less rooms than required	
soc_rb_limited	Percentage of persons day-to-day activities limited a lot or a little (Standardised Illness Ratio)	

soc_rb_unpaidcare	Percentage of persons providing unpaid care	Socio-Economic
soc_rb_quali_1_2	Percentage of persons aged 16 years and over whose highest level of qualification is Level 1, Level 2 or Apprenticeship	
soc_rb_quali_3	Percentage of persons aged 16 years and over whose highest level of qualification is Level 3 qualifications	
soc_rb_quali_4	Percentage of persons aged 16 years and over whose highest level of qualification is Level 4 qualifications and above	
soc_rb_schoolstudent	Percentage of persons aged 16 years and over who are schoolchildren or full-time students	
soc_rb_2cars	Percentage of households with 2 or more cars or vans	
soc_rb_pubtrans	Percentage of persons aged between 16 and 74 years who use public transport to get to work	
soc_rb_privtrans	Percentage of persons aged between 16 and 74 years who use private transport to get to work	
soc_rb_walkcycle	Percentage of persons aged between 16 and 74 years who walk, cycle or use an alternative method to get to work	
emp_rb_unemp	Percentage of persons aged between 16 and 74 years who are unemployed	Employment
emp_rb_parttime	Percentage of employed persons aged between 16 and 74 years who work part time	
emp_rb_fulltime	Percentage of employed persons aged between 16 and 74 years who work full-time	
emp_rb_sic_a	Percentage of employed persons aged between 16 and 74 years who work in the agriculture, forestry or fishing industries	
emp_rb_sic_b_f	Percentage of employed persons aged between 16 and 74 years who work in the mining, quarrying or construction industries	
emp_rb_sic_c	Percentage of employed persons aged between 16 and 74 years who work in the manufacturing industry	
emp_rb_sic_d_e	Percentage of employed persons aged between 16 and 74 years who work in the energy, water or air conditioning supply industries	
emp_rb_sic_g	Percentage of employed persons aged between 16 and 74 years who work in the wholesale and retail trade; repair of motor vehicles and motor cycles	
emp_rb_sic_h	Percentage of employed persons aged between 16 and 74 years who work in the transport or storage industries	
emp_rb_sic_i	Percentage of employed persons aged between 16 and 74 years who work in the accommodation or food service activities industries	
emp_rb_sic_k_l	Percentage of employed persons aged between 16 and 74 years who work in the financial, insurance or real estate activities	
emp_rb_sic_o	Percentage of employed persons aged between 16 and 74 years who work in the administrative or support service activities industries	
emp_rb_sic_pub_o	Percentage of employed persons aged between 16 and 74 years who work in the public administration or defence; compulsory social security industries	
emp_rb_sic_p	Percentage of employed persons aged between 16 and 74 years who work in the education sector	
emp_rb_sic_q	Percentage of employed persons aged between 16 and 74 years who work in the human health and social work activities industries	
Workzone-based characteristic		
dem_wp_f	Percentage of workplace population aged 16 to 74 years, females	Demographic Structure
dem_wp_f_25_39	Percentage of workplace population aged 25 to 39 years, females	
dem_wp_16_24	Percentage of workplace population aged 16 to 24 years	
dem_wp_60_74	Percentage of workplace population aged 60 to 74 years	
dem_wp_white	Percentage of workplace population of ethnic group: English, Welsh, Scottish, Northern Irish, British	
soc_wp_quali_4	Percentage of workplace population with highest level of qualification: Level 4 qualifications and above	

soc_wp_quali_3	Percentage of workplace population with highest level of qualification: Level 3 qualifications	
soc_wp_nssec_1	Percentage of workplace population with National Statistics Socio-Economic Classification: 1. Higher managerial, administrative and professional occupations	
soc_wp_nssec_2	Percentage of workplace population with National Statistics Socio-Economic Classification: 2. Lower managerial, administrative and professional occupations	
soc_wp_nssec_3	Percentage of workplace population with National Statistics Socio-Economic Classification: 3. Intermediate occupations	
soc_wp_nssec_5	Percentage of workplace population with National Statistics Socio-Economic Classification: 5. Lower supervisory and technical occupations	
soc_wp_nssec_6	Percentage of workplace population with National Statistics Socio-Economic Classification: 6. Semi-routine occupations	Socio-Economic
soc_wp_nssec_7	Percentage of workplace population with National Statistics Socio-Economic Classification: 7. Routine occupations	
soc_wp_distance_5	Percentage of workplace population who distance travelled to work is Less than five kilometres	
soc_wp_distance_20	Percentage of workplace population whose distance travelled to work is 20 kilometres and over	
soc_wp_distance_noplace	Percentage of workplace population with no fixed place	
soc_wp_pubtrans	Percentage of workplace population whose method of travel to work is: Underground, metro, light rail or tram; train; bus, minibus or coach	
soc_wp_walkbycle	Percentage of workplace population whose method of travel to work is: bicycle, on foot	
emp_wp_fullstudent	Percentage of workplace population who is a full-time student	
emp_wp_self_with_fullpart	Percentage of workplace population which is self-employed with employees: Full or part-time	
emp_wp_hours_f49	Percentage of workplace population which worked full-time for 49 or more hours	
emp_wp_hours_p16_30	Percentage of workplace population which worked part-time for 16 to 30 hours	
emp_wp_hours_p15	Percentage of workplace population which worked part-time for 15 hours or less	
emp_wp_sic_c	Percentage of workplace population employed in (SIC) C Manufacturing	
emp_wp_sic_d_e	Percentage of workplace population employed in (SIC) D Electricity, gas, steam and air conditioning supply / Standard Industrial Classification (SIC): E Water supply; sewerage, waste management and remediation activities	
emp_wp_sic_f	Percentage of workplace population employed in (SIC) F Construction	
emp_wp_sic_g	Percentage of workplace population employed in (SIC) G Wholesale and retail trade; repair of motor vehicles and motor cycles / L Real estate activities	
emp_wp_sic_h	Percentage of workplace population employed in (SIC) H Transport and storage	Employment
emp_wp_sic_i	Percentage of workplace population employed in (SIC) I Accommodation and food service activities	
emp_wp_sic_j	Percentage of workplace population employed in (SIC) J Information and communication	
emp_wp_sic_k	Percentage of workplace population employed in (SIC) K Financial and insurance activities	
emp_wp_sic_m	Percentage of workplace population employed in (SIC) M Professional, scientific and technical activities	
emp_wp_sic_o	Percentage of workplace population employed in (SIC) O Public administration and defence; compulsory social security	
emp_wp_sic_p	Percentage of workplace population employed in (SIC) P Education	
emp_wp_sic_q	Percentage of workplace population employed in (SIC) Q Human health and social work activities	
emp_wp_sic_r_s	Percentage of workplace population employed in (SIC) R,S Arts, entertainment and recreation; other service activities	

APPENDIX 2

Table 1. Table showing codes, names and characteristics for Town Supergroups.

Supergroup	Supergroup name	Employment characteristics	Demographic characteristics	Street network
1	International metropole towns	High share of professional, scientific, technical, information and communication, and financial, insurance, or real estate activities. Lowest share of manufacturing activities.	Youngest population. Highest ethnic diversity, high level of non-English speakers and lowest share of British and white population. Highest share of EU citizens.	Highest average share and high shares of clusters 3, 4,5 and 6.
2	Dense urban towns	Highest average share across industries, with high share of transport, public administration and defence activities. Low share of professional, scientific, communication and finance activities. Lowest share of agriculture, forestry or fishing activities.	Young population. high ethnic diversity, highest level of non-English speakers and average share of British population. Low shares of white population. Lowest level of married couples.	Highest share of network clusters 1 and 5.
3	Industrial towns	High share of electricity, gas, steam, and water supply, as well as construction and manufacturing activities. Lowest share of professional, scientific and technical activities.	Highest share of British and white population, lowest share of Arab or other ethnic groups.	Highest share of cluster 3.
4	Knowledge-based towns	Highest share of education activities. High shares of professional, scientific, technical, information and communication, financial and insurance activities	Older population (i.e., high shares of 45–90-year-olds). Highest share of married couples. Average ethnic diversity.	Highest share of cluster 4, and high shares of cluster 8.
5	Non-urban towns	Most average share across industries, with lowest share in human health and social work activities. High share of agriculture, forestry or fishing industries.	Average age and ethnic diversity.	Highest share of network clusters 6, 7 and 8.
6	Agriculture & Tourism-based towns	Highest share in accommodation, food service, arts, entertainment and recreation activities. Highest shares in agriculture, forestry or fishing industries.	Highest share of 65–90-year-olds. High share of British and white population, low share of Arab or other ethnic groups. Lowest share of Black, African or Caribbean population.	High shares of clusters 3 and 4.