Understanding the Dynamics and Context of New York Transportation Hubs

Yunzhe Liu[*1], Alex Singleton[†1], Daniel Arribas-Bel[‡1]

[1] Geographic Data Science Lab, Geography and Planning University of Liverpool, Liverpool, L69 7ZT

January 18, 2019

KEYWORDS: Transit-Oriented Development, Public Transit Planning, Self-Organising Map, TOD Topology

1. Introduction

Urban planning towards transit-oriented development (TOD) has become a focal point of a more comprehensive solution to typical transport challenges, such as traffic congestion and parking difficulties, air and noise pollution, excessive greenhouse gas emission, public health issues, and wellbeing-related social exclusion problems. Since Calthorpe (1993) codified the 'Three Ds' concept of TOD, namely, high density in development, diversity in land use and good urban design, such development pattern has been widely recognised and accepted as a leading planning strategy by most planning agencies around the world, aiming to deliver a favourable environment, highly compact, mixed-use, pedestrian- and cycling friendly, and transit supportive neighbourhood in vicinity of public transport nodes (e.g. transit stations). Seemingly, the macroscopic concept of TOD is consistent in its prescriptions for policy-making and planning, however, extensive studies have illustrated that the microscopic implementation of TOD is of necessity to be highly sensitive to practical-based conditions and customised accounted for local differences and complexities in order to deliver concrete policy and design for the targeted station area (Kamruzzaman et al., 2014; Higgins and Kanaroglou, 2016). In response, it is strongly suggested that to build context-based TOD typologies, or TOD classifications, for purpose of differentiating heterogeneous station catchment areas (usually within the walkable radius, e.g. 400-800m) into more homogeneous classifications based on the similarities of their multidimensional characteristics that potentially drive public transport use (Higgins and Kanaroglou, 2016; Lyu, Bertolini and Pfeffer, 2016; Papa et al., 2018). When done correctly, urban planners can utilise these TOD typologies as a first step to evaluate the performance of existing condition against TOD expectations; policymakers can introduce a context-sensitive policy that integrates land-use and transport to promote TOD and achieve even broader goals: e.g. accelerating the progress towards smart urbanism.

Within this context, this paper continues the tradition by constructing a TOD topology aiming at effectively monitoring the current condition of the subway catchment areas in the designated case study area, i.e. the New York City (NYC). However, this research improves upon previous research and makes contributions regarding the following aspects:

- Firstly, this paper explores the potentiality of importing open data into the field of context-based TOD typology. Most of the traditionally used inputs to build a TOD topology are supplemented /upgraded by the timely and open data.

- Secondly, in addition to the inputs reflecting the static characteristics, human's mobility data which are passively collected by the multi-source sensors, such as subway turnstile data, are also employed in this study to fill up the blankness in the field of TOD topology by emphasising the dynamic perspective.

- And finally, given the situation where the input variables form a hyperspace with large dimension, commonly used unsupervised classifier, e.g. k-means, is replaced by a more robust

*psyliu7@liverpool.ac.uk
†Alex.Singleton@liverpool.ac.uk
‡ D.Arribas-Bel@liverpool.ac.uk

unsupervised classifier, i.e. Self-Organizing Map (SOM), in order to differentiate the latent characteristics of subway catchment areas (within 800m buffer) in the NYC.

2. Data Description

Variable selection for this paper is proportionally based on the 'Five Ds' model (i.e. Density, Diversity, Design, Destination Accessibility, and Distance to Transit) introduced by Ewing and Cervero (2010), which is an extension of the traditional 'Three Ds' TOD principles conceptualised by Cervero and Kockelman (1997), and on the empirical findings refined from a Systematic Literature Review of more than 20 recently published studies, accompanying with the consideration of quality and practical availability of the data. The selected variables are categorised into four domains, i.e. Land Use and Built Environment, Transit-related, Location and Accessibility, and Socioeconomic and Demographic, which mainly come from the following open data sources: American Community Survey (ACS), National Walkability Index (NWI), Smart Location Database (SLD), NYC Open Data (NYCOD), NYC Planning (NYCP), and Metropolitan Transportation Authority (MTA). **Table 1** shows the example of selected variables for two typical subway stations located in NYC. And diagrams in **Figure 1** are the heatmaps illustrating the 'weekday profile' of the subway entries and exits (i.e. turnstile data) at the example station (111th Street Station).

Station Name	***Median Income ($)***	***Street Tree Density (per km^2)***	***POI & LU: Residential (%)***	***POI & LU: Commercial (%)***	***EA: Master or Doctorate (%)***	***D4d (per mile^2)***	**…**
5th Ave - 59th St	171603.20	96.38	14.47	48.68	44.26	11538.29	…
111th St	54656.67	105.14	82.61	3.26	0	2609.65	…

Table 1 Example of Selected Variables for Constructing Context-based TOD Typology. Totally, there are about 118 variables derived from the aforementioned data sources, comprised by 27 Socioeconomic & Demographic; 23 Land Use & Built Environment; 7 Location & Accessibility; 61 Transit-related (including 60 variables measuring the subway turnstile pattern, i.e. Entries and Exits).

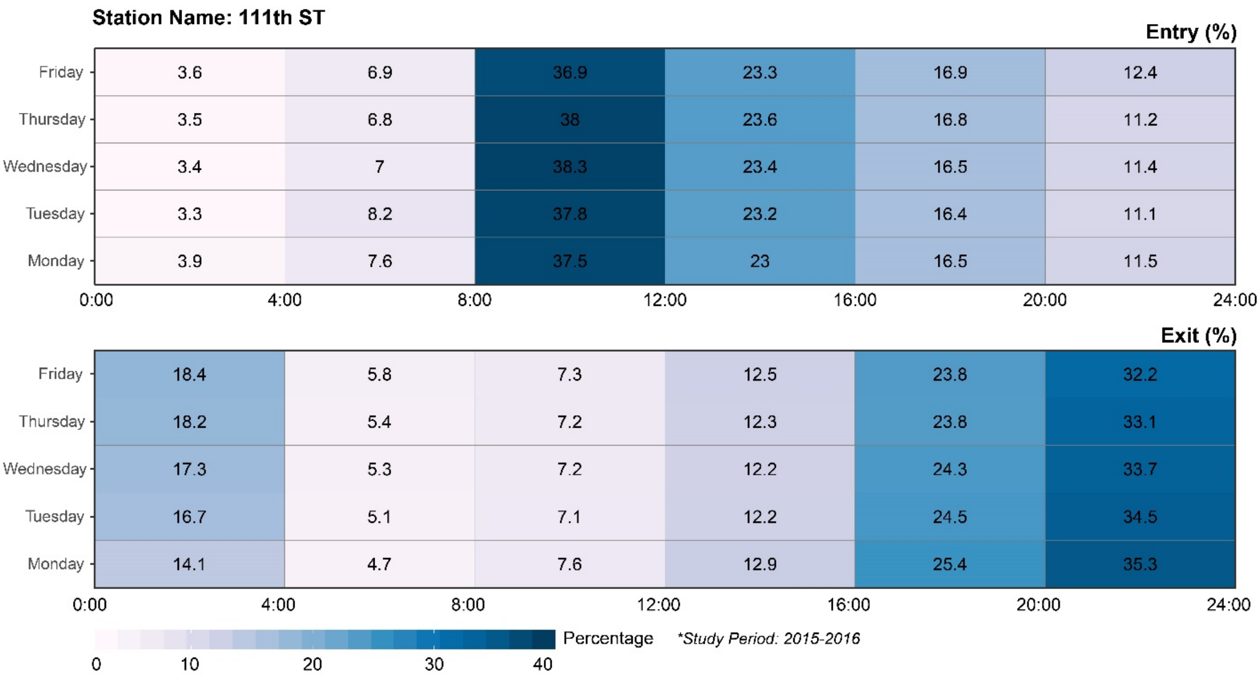

Figure 1 Travel 'Weekly Profiles' (Entry and Exit) of 111th Street Station, Queens, New York City. Note that the temporal resolution is 4-hour interval, meaning that a day consists of 12 variables describing entries and exits (i.e. 30 variables per week, weekends are excluded from the study).

3. Methodology

The general aim of this paper is to apply clustering analysis to the subway station areas within the NYC to classify them into heterogenous TOD typologies based on the (dis-)similar pattern underneath their own multidimensional characteristics. However, when such considerable variables are put together, it inevitably results in forming a discrete matrix with very high dimensionality, which sometimes can be quite difficult to be handled by many traditional clustering techniques, referring to the ‘curse of dimensionality’ phenomena. In this case, we adopt the Self-Organising Map (SOM), also referred to the Kohonen Map, to effectively reduce the complexity caused by the high-dimensional space. One of the distinctive advantages offered by SOM is the ability to visualise and simultaneously abstract complex, nonlinear statistical relationships between high-dimensional data. This algorithm has been implemented to various disciplines related to hyperspace visualisation, generation of feature maps, pattern recognition and classification: for instance, text mining in bioinformation, education, (satellite) image classification, psycholinguistic study, and finance and the stock market. Particularly, this technique has also been applied to urban study aiming to address the problem of describing the urban population (see Spielman and Folch, 2015; Arribas-Bel and Schmidt, 2013; Sohn, 2013).

After testing several combinations of parameters required by SOM, we identified that the combination of *rectangular* topology and the *Bubble* neighbourhood function with a *linear* decline in learning rate (ranging from 1.0 to 0.01) results in the smallest average quantisation error and average distortion measure, indicating a relatively good clustering result. The neighbourhood distance is calculated and visualised in **Figure 2** (the U-matrix), showing the distance between each of the neurons in the SOM topography, which can also be used as the foundation of posterior clustering analysis. Although the U-matrix from the SOM clearly shows some general patterns, in order to further simplify the complexity, here, we applied Hierarchical Clustering with connectivity constraints to the trained SOM network. The clustering results are presented by a circular dendrogram in **Figure 3**, depicting five unique clusters (namely, TOD typologies) identified in NYC.

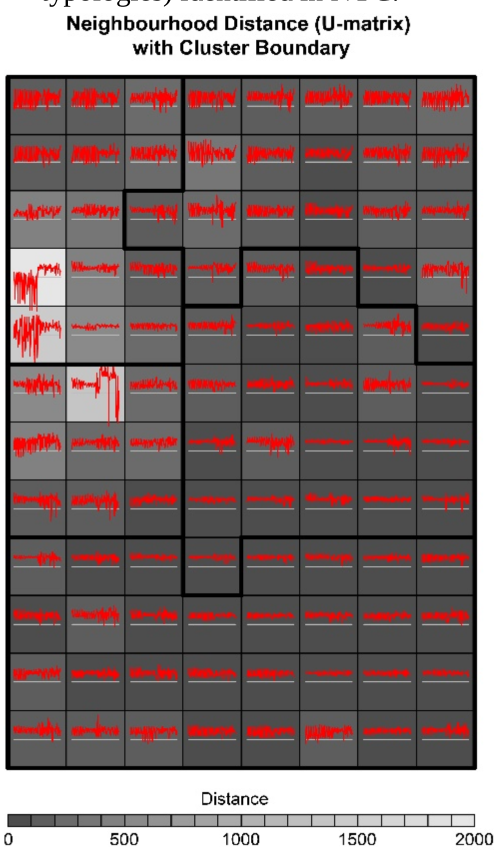

Figure 2 U-Matrix

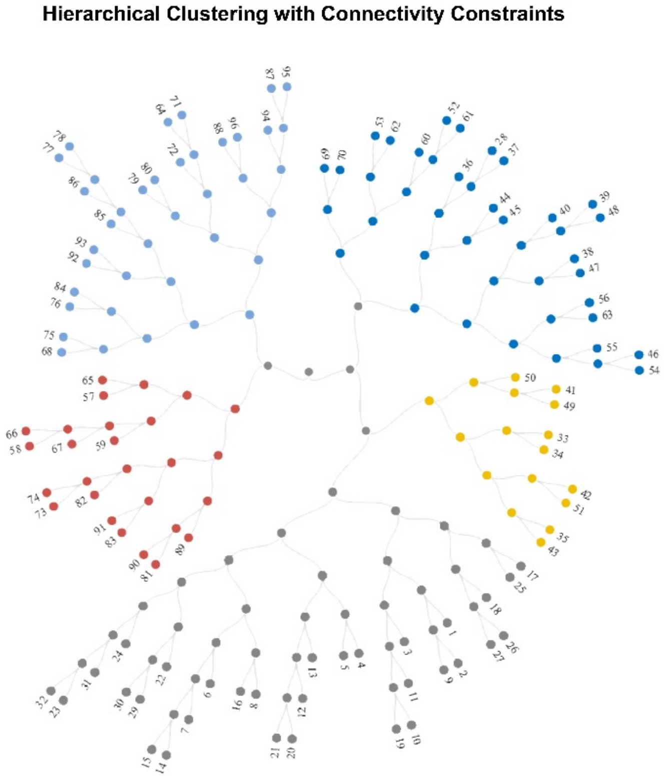

Figure 3 Dendrogram of Hierarchical Clustering Result. Each node indicates a neuron in the trained SOM network

4. Results and Conclusion

Figure 4 shows the within-cluster variable distribution for the five TOD typologies in NYC, which can be utilised to assist in interpreting and summarising the salient characteristics for each of the identified cluster. Generally, the dynamic variations (majorly indicated by the Transit-related variables) are successfully differentiated. Take Cluster 5 for instance, stations of this TOD typology are more likely to witness a high-demand of subway entry in the evening peak and exit in the morning peak, indicating a typical 'home to work' travel pattern; the land use within the catchment areas are mainly used for commercial and business. Moreover, according to the map shown in **Figure 5** these stations are primarily located in the downtown of Manhattan, a famous central business district of the NYC, which is in accord with the description above. More detailed interpretation will not be presented here in consideration of the length limit for this abstract.

In conclusion, this paper adds new forms of open data to the construction of conventional context-based TOD typology, aiming to enrich the diversity of aspects understanding the detailed TOD development in the study area. For instance, subway turnstile data and Point of Interests (POI) data are respectively utilised to describe the dynamic and social functional aspects of the TOD typology, which are overlooked in most of the previous studies. Moreover, the SOM provides an advantageous ability to handle such data with high dimensionality, as the preliminary results have shown, this neural-network based classifier can successfully differentiate the salient characteristics of the subway station catchments from the latent multidimensional information. However, there are also some limitations, for instance, the temporal resolution of the subway turnstile data (i.e. 4-hour interval) is relatively coarse, peaks and off-peaks are not easily separated, resulting in some small obstacles in cluster interpretation.

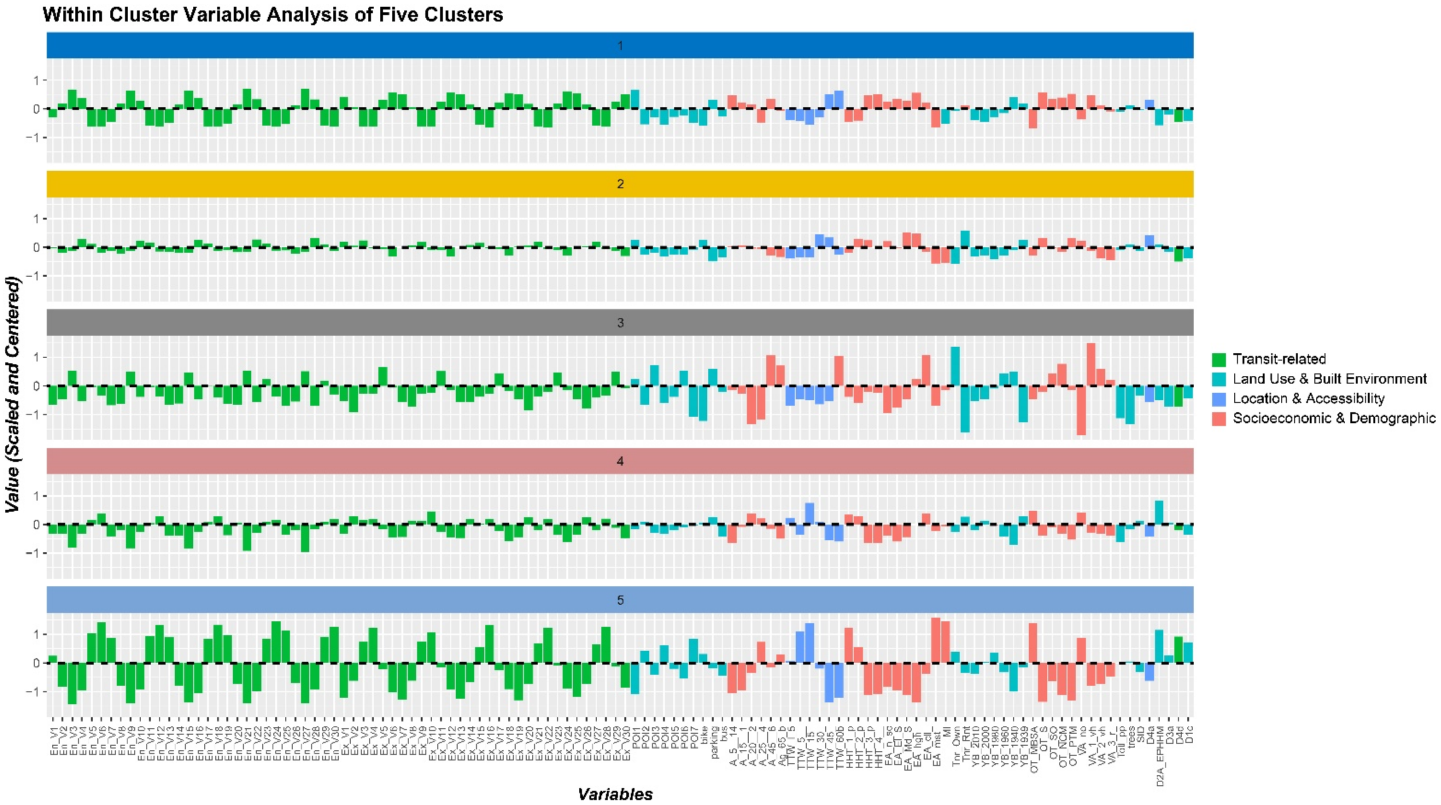

Figure 4 Within Cluster Variable Analysis of Five Clusters. Values are standardised (scaled and centred), i.e. the national mean is 0 (indicated by the hash line).

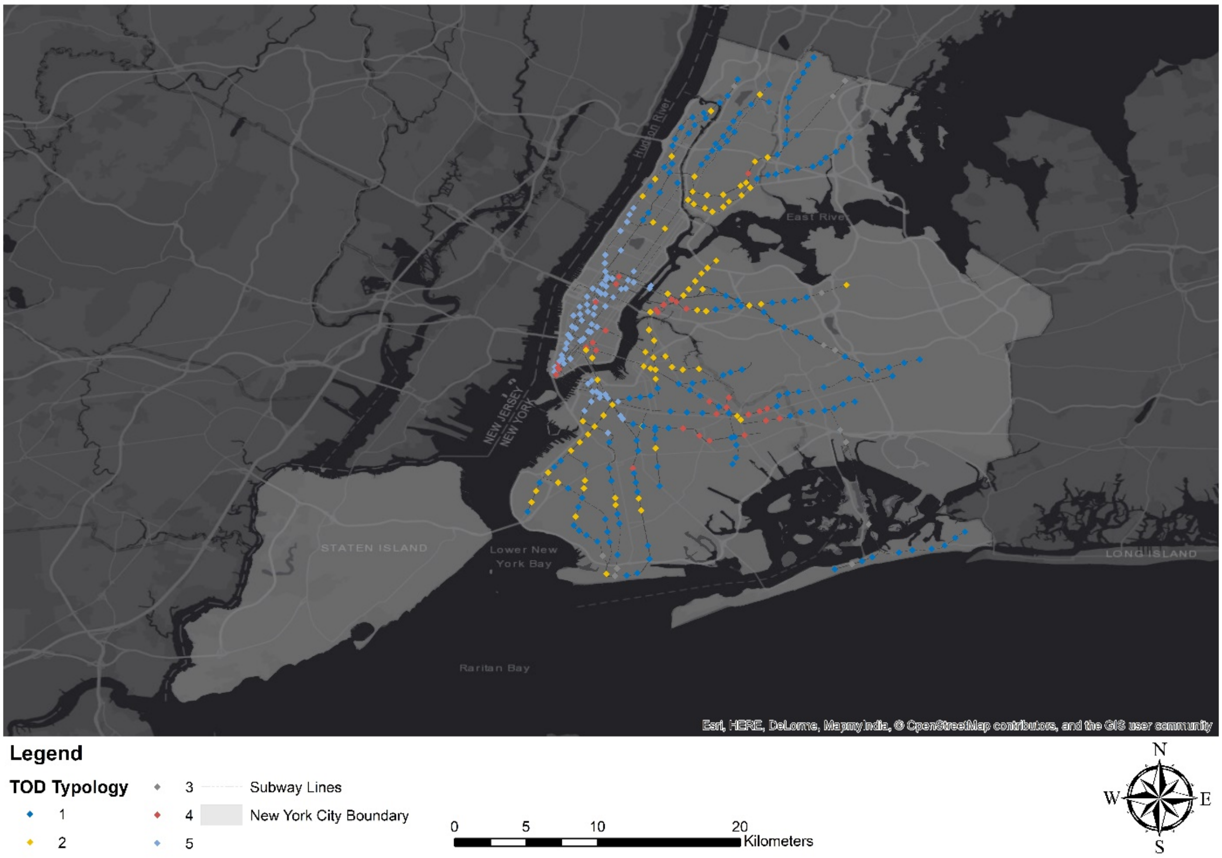

Figure 5 Geographic distribution of the five created TOD typologies in the New York City

5. References

Arribas-Bel, D. and Schmidt, C. (2013). Self-organizing maps and the US urban spatial structure. *Environment and Planning B: Planning and Design*. 40 (1), p362-371.

Calthorpe (1993). *The Next American Metropolis: Ecology, Community, and the American Dream*. New York: Princeton Architectural Press.

Cervero, R. and Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*. 2 (3), p199-219.

Ewing, R. and Cervero, R. (2010). Travel and the Built Environment. *Journal of the American Planning Association*. 76 (3), p265-294.

Higgins, C. D. and Kanaroglou, P. S. (2016). A latent class method for classifying and evaluating the performance of station area transit-oriented development in the Toronto region. *Journal of Transport Geography, 52*, 61–72.

Kamruzzaman, M., Baker, D., Washington, S., and Turrell, G. (2014). Advance transit-oriented development typology: Case study in Brisbane, Australia. *Journal of Transport Geography, 34*, 54–70.

Lyu, G. Bertolini, L. and Pfeffer, K. (2016). Developing a TOD typology for Beijing metro station areas. *Journal of Transport Geography*. 55, 40-50.

Papa E., Carpentieri G., Angiello G. (2018) A TOD Classification of Metro Stations: An Application in Naples. In: Papa R., Fistola R., Gargiulo C. (Eds.) *Smart Planning: Sustainability and Mobility in the Age of Change. Green Energy and Technology.* Springer, Cham, 285-300

Sohn, K. (2013). Feature Mapping the Seoul Metro Station Areas Based on Self-Organizing Map. *Journal of Urban Technology*. 20 (4), 23-42.

Spielman, S. and Folck, D. (2015). Social Area Analysis and Self-Organizing Maps. In: Brunsdon, C. and Singleton, A.*Geocomputation: A Practical Primer*. London: SAGE Publications Ltd. 152-168.