A Data-driven Perspective for Exploring the Urban Functional Identities in Hong Kong

Zidong Yu[*1] and Xintao Liu[†1]

[1] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University

January 5, 2021

Summary

Revealing urban functions in cities can be an essential means to describe the characteristics of human activities and places. Through the lens of points of interest and tourist online reviews collected in downtown Hong Kong, this study develops a semantic-based framework to detect and compare urban functional identities. The findings depict a unique image of city identity for the downtown area in Hong Kong that contains miscellaneous urban vibrancies. Ultimately, a profound insight is yielded in our study for the understanding of local activities, which can provide human-centric implications for evidence-based urban planning and establish real-world cognitions to the public.

KEYWORDS: Human Activity, Cognition, Urban Space, Semantics, Point of Interest

1. Introduction

City is as diverse as humans reside in it, in which people have multifarious and complex daily lives. Economies, cultures, demography, and social resources can shape the uniqueness of local experience for residents and tourists. Understanding urban functions that are closely associated with human activities can contribute to the knowledge of urban identities.

Several methodologies have been proposed to explore urban functions and human activities. For example, some early investigations have extracted urban functions using remote sensing images, social sensing data, and questionnaires (Zhou & Zhang, 2016). However, land-use details like commercial or residential zones cannot be well recognized by using satellite images, while the questionnaires are mostly time-consuming and costly. Therefore, a more comprehensive methodology is requisite to depict functional identities in a highly dense urban environment.

This study introduces a data-driven semantic perspective to explore the functional identity in a highly complex space. Using points of interest (POIs) collected in the downtown area of Hong Kong, we aim to address three research questions (RQ): What are the thematic place topics that can be extracted from a complex and liveable urban space? What are the differences in the spatial distributions of these topics? Given selected landmarks, how do these thematic functions constitute place identities in comparison to cognitions from tourist online reviews?

2. Study Area and Dataset

As shown in **Figure 1**, the downtown districts in Hong Kong are chosen as the study area. The high urban density and diverse human activities can provide an appropriate sample for exploring functional identities. The POIs provided by AutoNavi contain more than 93,000 records. As suggested by Lansley and Longley (2016) and prior knowledge of our study area, POIs are calculated across a 100 m grid space.

---

[*] zidong.yu@connect.polyu.hk
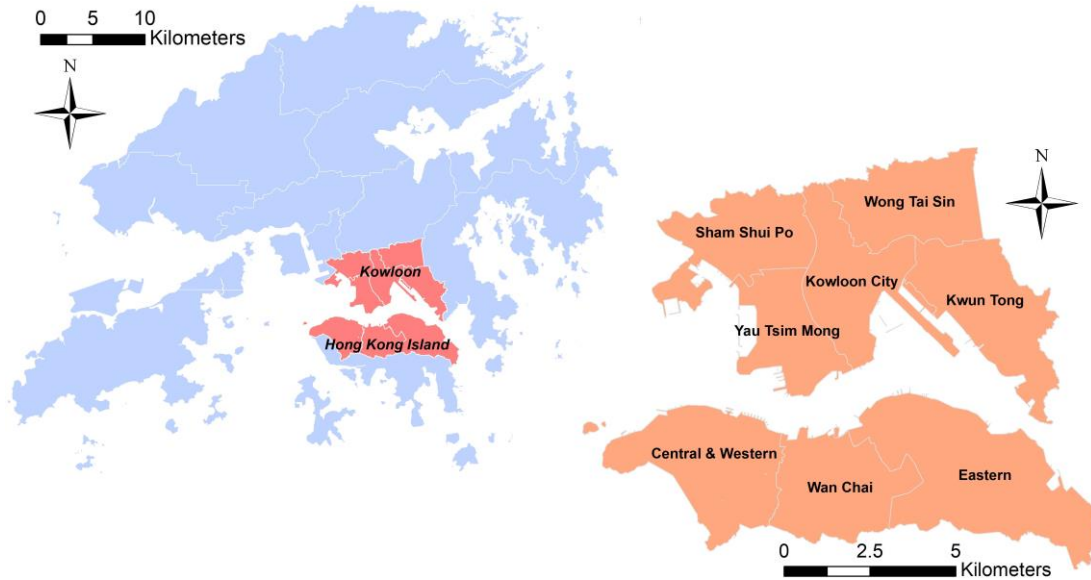[†] xintao.liu@polyu.edu.hk (Corresponding Author)

Figure 1 The downtown districts in Hong Kong.

3. Methodology

Latent Dirichlet Allocation (LDA) is implemented to extract thematic urban activities, which has been one of the most prevalent topic modeling approaches (Blei et al., 2003). Compared to other methods for geographic text classification like convolutional neural networks, LDA is an unsupervised generative probabilistic model with less computation, by taking a word-based numeric representation (Adams & McKenzie, 2018). Analogously, the extraction of human activities in urban space can be considered as the LDA process. POI categories are taken as words, for instance, primary school, financial company, and nightclub. A grid cell containing POIs is regarded as a document, while the entire study area is the corpus. As shown in **Figure 2**, two output matrices from LDA are illustrated for representing the functional identities of urban space.
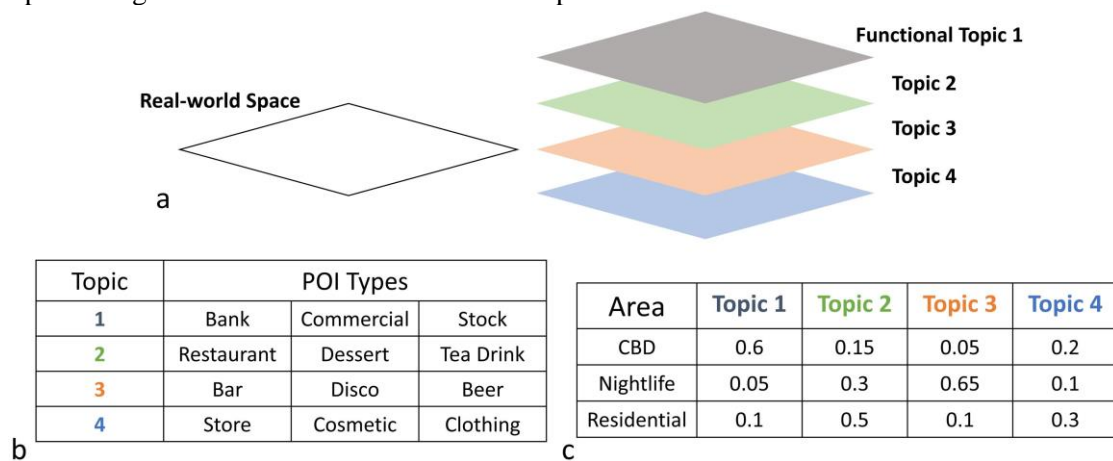

Figure 2 Illustration of identifying thematic functions using POIs: (a) Decomposing a composite function, (b) topic-to-word matrix, and (c) document-to-topic matrix.

4. Results and Discussion

4.1. Thematic Topics from LDA

In addressing RQ 1, we investigate and discuss the nine functional topics generated from the LDA process, which can be used to describe thematic human activities in urban space. **Figure 3** displays the comparison word cloud in terms of the most prevalent functions in each topic. Topic 1 is described as a business topic related to financial and commercial activities, implying an urban function of working. Topic 3 also suggests a similar function of working. However, topic 3 underlines more about public

services that are provided by government agencies and social organizations. Interestingly, tourist attractions seem to have a close spatial relationship with public institutes. Topics 4 and 7 are implied as shopping activities with distinct emphasis on fashion and household decoration, respectively. The remainder topics are also associated with thematic functions, such as cultural activities for topic 2 and health and beauty for topic 5.

Figure 3 Comparison of word cloud of latent topics.

4.2. Spatial Patterns of Human Activities

We next focus on RQ 2 and find out that spatial distribution of thematic POIs reveals distinct geographic patterns. **Figure 4** visualizes the spatial distribution of four of the identified functional topics. Financial and banking activities reach the highest level in *Central*, where most of the global financial institutions and banks are located. For the government and tourism topic, two clustering areas have appeared in the midst of *Admiralty* and *Wan Chai*, in what places a considerable number of government offices are located. Interestingly, we find that topic 4 and topic 6 indicate a similar spatial distribution pattern. A possible explanation is that shopping and eating always have a close association and both are main subjects in everyday leisure activities. In Hong Kong Island, *Causeway Bay* is shown with the highest density of related POIs, in which this district is one of the busiest commercial neighbourhoods jam-packed with various shops and restaurants.
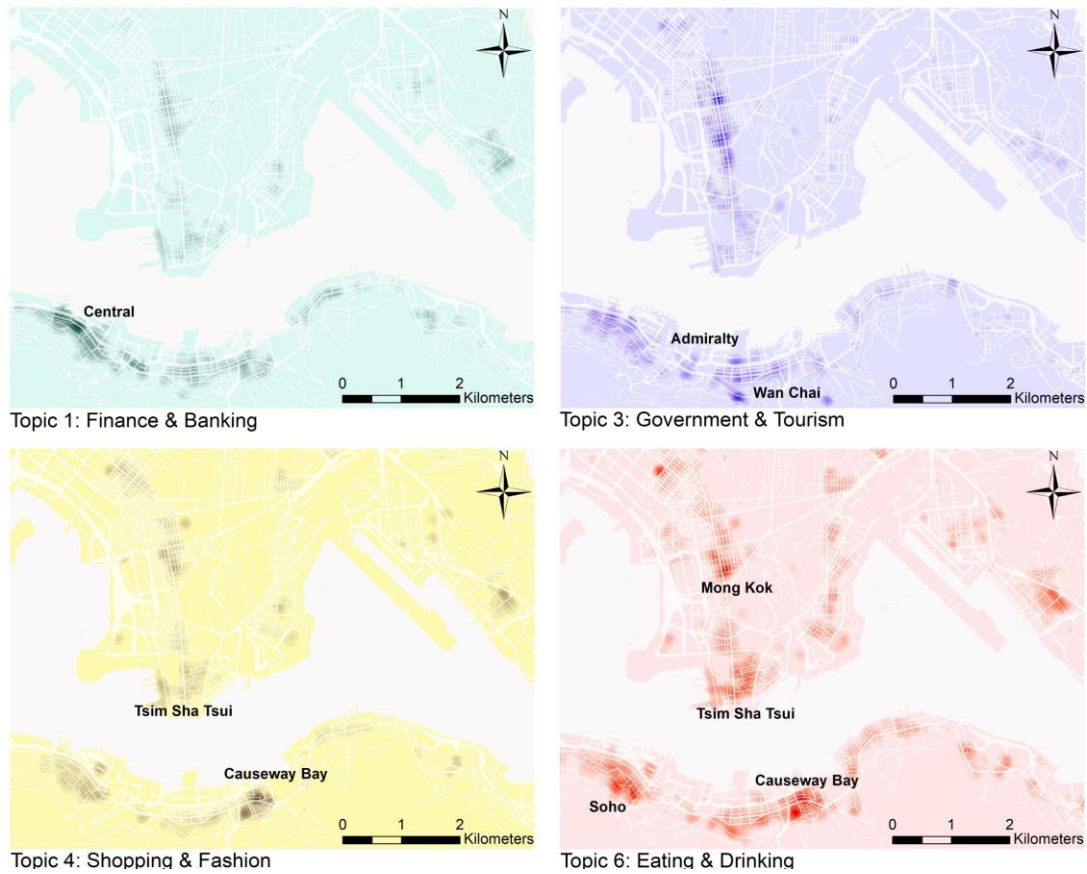
Figure 4 Spatial distribution of POIs from 4 selected topics.

4.3. Functional and Cognitive Identities of Local Landmarks

A street-level perspective is furthermore introduced to respond to RQ 3 by exploring the various activity patterns across selected landmarks. As shown in **Figure 5**, location quotient is used to quantify the association of each thematic topic in 6 selected landmarks (Lansley & Longley, 2016). This metric is calculated by dividing the percentages of corresponding areas for each topic within the area of landmarks by the overall percentages of the same topic in the entire study area. A strong relationship between a topic and a place can be reported if the corresponding value is closer to 1. From this heat map, *Causeway Bay* reveals a strong concentration effect on shopping and leisure activities. Additionally, we discover that POIs in the *Central* exhibit a higher degree in the topic of finance and banking. Place topics identified from *Immigration Tower* are overrepresented on government and public services. As we expected, *Soho* and *Lan Kwai Fong* unveil an obvious tendency towards eating and drinking activities. One unexpected outcome is that the shopping activities are underrepresented in Temple Street, where there is known for night market and flea market.
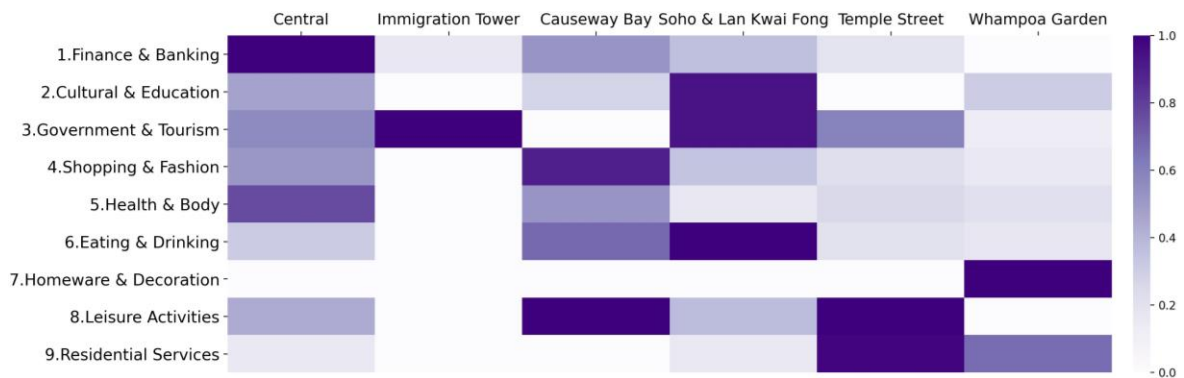

Figure 5 Heat map of location quotients for functional topics aggregated by landmarks.

In **Figure 6**, we explore three tourist-oriented landmarks by deriving high-frequency keywords from tourist reviews in *TripAdvisor,* which is an online travel platform with user-generated content. Interestingly, tourist impressions reflected by word clouds are mostly consistent with the findings from functional identities, which can be regarded as complemental knowledge. For example, the tourist's impression of *Soho & Lan Kwai Fong* is about nightlife containing keywords like *night* and *bar*, while *Causeway Bay* indicates a cognitive identity related to shopping activities. Surprisingly, reviews from *Temple Street* unveil an impression of night markets with cheap products and are not described in location quotients in Figure. 5. A reasonable explanation can be that outdoor events or activities cannot be well recorded in POI collection, such as open-air food stalls.

Figure 6 Semantic impressions of tourist landmarks extracted from tourist online reviews.

5. Conclusions

A thorough insight into human activities in a dense urban circumstance is presented in this case study. Particularly, various functional identities from the LDA approach demonstrate that complex and diverse human activities can be suitably segmented based on POIs semantic signatures. Commercial districts like *Central* or *Causeway Bay* perform highly composite urban functions regarding business, shopping, eating, and other daily activities. This is presumably due to the higher population density, more convenient surrounding infrastructures, and commerce-oriented land use schemes. Compared to the study by Bahrehdar et al. (2020) on *Flickr* tags that emphasizes perceptual user generated content, using POIs can reflect more objective and more sophisticated images regarding place functional properties and daily activities. Further, our analytical findings present a real-world impression to the public, in which location quotient and user-contributed reviews are taken together to describe a miscellaneous life of local famous landmarks. For example, keywords extracted from *Temple Street* depict an intelligible image of a bustling, open-air market with numerous stalls. By embedding tourist reviews into POIs-based place identities, we attempt to answer the questions raised by former literature of how perceptual and functional datasets can be integrated to explain city image (Bahrehdar et al., 2020). In sum, understanding subjective and objective aspects of urban life across highly dense urban spaces is a powerful footstep towards capturing diverse experiences from tourists or residents who are trying to thoroughly explore various activities in Hong Kong.

References

Adams, B. and McKenzie, G., 2018. Crowdsourcing the character of a place: Character-level convolutional networks for multilingual geographic text classification. *Transactions in GIS*, *22*(2), pp.394-408.

Bahrehdar, A.R., Adams, B. and Purves, R.S., 2020. Streets of London: Using Flickr and

OpenStreetMap to build an interactive image of the city. *Computers, Environment and Urban Systems*, *84*, p.101524.

Blei, D.M., Ng, A.Y. and Jordan, M.I., 2003. Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), pp.993-1022.

Lansley, G. and Longley, P.A., 2016. The geography of Twitter topics in London. *Computers, Environment and Urban Systems*, *58*, pp.85-96.

Lynch, K., 1960. *The image of the city* (Vol. 11). MIT press.

Zhou, X. and Zhang, L., 2016. Crowdsourcing functions of the living city from Twitter and Foursquare data. *Cartography and Geographic Information Science*, *43*(5), pp.393-404.

Biographies

Zidong Yu: Zidong is a second-year MSc student in the Department of Land Surveying and Geo-Informatics in Hong Kong Polytechnic University. He obtained his bachelor's Honour degree in Geomatics from the University of Waterloo, Canada. His research interests include Urban Informatics, and Geospatial Applications on Socioeconomics.

Xintao Liu: Dr. Liu is an Assistant Professor in the Department of Land Surveying and Geo-Informatics in Hong Kong Polytechnic University. His research interest focuses on GIScience, transportation geography, and complex network, with the aim of using state-of-art technologies to advance smart city for a better urban life.